



A multi-classification algorithm based on multi-domain information fusion for motor imagery BCI

Jiaqi Wang, Wanzhong Chen, Mingyang Li*

College of Communication Engineering, Jilin University, Changchun, China



ARTICLE INFO

Keywords:

Brain Computer Interface (BCI)
EEG signal
Motor imagery
Channel selection
Multi-domain information fusion
Optimized common spatial pattern

ABSTRACT

The current problem of motor imagery Electroencephalogram(EEG) signal classification is low classification accuracy and fixed EEG channel selection. We proposed a novel classification algorithm for motor imagery EEG signals, which overcomes the contradiction between the number of channels and the representational ability of features. Higher classification accuracy is achieved using less number of channels. The algorithm makes a combination of time windows, filter banks, and an optimal sorting of the projection space to reveal multi-domain information. Experiments based on the two datasets of BCI Competition have proved that the channel selection strategy used in this paper can adapt to the subject's neural information and select the optimal channel combination. The feature extraction algorithm proposed can achieve excellent classification accuracy (77.7 %) and kappa value (0.70). The results are improved by 26.2 % compared to the One Versus One-Common Spatial Pattern (OVO-CSP) method and by 8.2 % compared to the One Versus One-Filter bank common spatial pattern (OVO-FBCSP) method. Additionally, the proposed method has outperformed to the other state-of-the-art methods using the same data set in terms of the performance. The proposed methodology can be employed as a promising tool for a motor imagery BCI device.

1. Introduction

Brain-computer interface (BCI) aims to establish a system that does not rely on the brain's normal output pathways of peripheral nerves and muscles but uses computers or other output devices to directly build up a special channel to communicate with the brain [1,2]. Electroencephalogram (EEG) is a non-invasive and low-cost technique of acquiring brain signals, which is widely used in BCI systems for research. Motor imagery (MI) has become a hot issue in the field of BCIs, as a phenomenon that the power of motor-relevant cortex EEG signals is decreased or increased when people imagine limb movements; the changes are known as event-related desynchronization (ERD) or event-related synchronization (ERS) [3,4].

The CSP is a particularly popular and effective signal processing technique for EEG-based BCIs, which can achieve a powerful performance using some tricks of the trade [5]. But this method has shortcomings such as noise sensitivity, overfitting, and only binary-class availability, which can lead to various extensions of the CSP. The filter bank CSP (FBCSP) is used to filter the EEG signal into multiple frequency bands to extract CSP feature, which can improve the classification

results [6]. Shrinkage Regularized Filter Bank CSP (SR- FBCSP) is a regularization approach based on shrinkage estimation, which can handle small sample problem and retain subject-specific discriminative features [7]. Regularized CSP (RCSP) has been proposed to reduce the overfitting of CSP. Four RCSP terms, including two proposed regularization terms for optimizing the objective function, are suggested in [8]. Analytic CSP (ACSP) can provide a more comprehensive picture of the underlying activity by explicitly considering the amplitude and phase information in the EEG [9]. A class discrepancy-guided sub-band filter-based CSP (CDFCSP) algorithm is proposed to automatically recognize and augment the discriminative frequency bands for CSP algorithms in [10]. A framework of information theoretic feature extraction (ITFE) is proposed to address the question of optimality of CSP in terms of the minimal achievable classification error and extensions to multiclass paradigms [11]. Alexandre et al. use spatial covariance matrices obtained from CSP as EEG signal descriptors and rely on Riemannian geometry to directly classify these matrices using the topology of the manifold of symmetric and positive definite (SPD) matrices [12]. A maximum mutual information linear transformation (MMI-LinT), and a nonlinear transformation(MMI-NonLinT) framework are proposed to

* Corresponding author.

E-mail address: limingyang@jlu.edu.cn (M. Li).

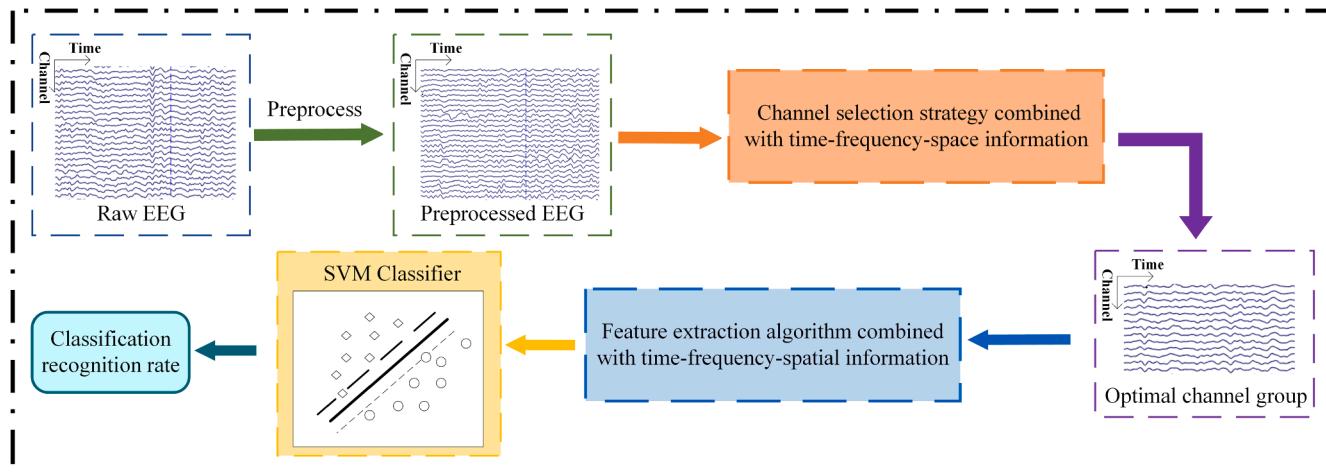


Fig. 1. General framework of motor imagery EEG classification algorithm based on multi-domain information fusion.

select the feature vectors obtained from the FBCSP, and a graphical model based hierarchical decoding framework is proposed to solve multi-class problems, which achieves excellent results [13]. An optimal channel selection method is proposed to improve CSP features by selecting the channels in terms of correlation coefficient values, which compute the Fisher score of the feature output based on FBCSP to apply the channel group to solve the binary classification problem [14].

Different from the perspective of the above algorithms, we do not only employ CSP as a feature extractor but also as a channel selector. Inspired by the characteristic that the projection space obtained from CSP contains spatial information of each channel, we design a criteria to search the optimal ordering of projection space. Based on this optimal ordering, we propose a channel selection strategy and an optimized feature extraction method. Moreover, the fusion of multi-domains may achieve better results, so we have taken advantage of the time windows and filter banks. Finally, we construct a classification framework based on time-frequency-space fusion, which effectively discards the number of channels irrelevant to the motion imagery and improves the classification accuracy significantly.

Experiment results show that the algorithms we proposed can achieve a higher classification accuracy and kappa value in the multi-classification tasks of motor imagery. The rest of this paper is organized as follows. In section 2, we briefly describe the experimental data we used and the pre-processing operation of the raw EEG signals. In section 3, all steps of the proposed algorithm are described in detail. Accordingly, the algorithm obtained results are illustrated in section 4, followed by a summary of our study in section 5.

2. Experimental data and preprocessing

In this paper, the EEG data based on four tasks (the left hand, right hand, both feet, and tongue) of motor imagery are derived from the BCI Competition III Dataset IIIa and the BCI Competition IV Dataset 2a.

Dataset IIIa contains the experimental data of 3 subjects (k3b, k6b, and l1b). The raw EEG data of a total of 64 channels were collected in the experiment. The experimental data of subject k3b is divided into 180 training samples and 180 testing samples, and the EEG data of subjects k6b and l1b are divided into 120 training samples and 120 testing samples. The four types of motor imagery tasks in the training samples and testing samples are the same, and details can be found in [15]. Dataset 2a contains the experimental data of 9 subjects (A01 ~ A09). The raw EEG data of a total of 22 channels were collected in the experiment. All EEG data of each subject is divided into 288 training samples and 288 testing samples, and the four types of motor imagery tasks in the training samples and testing samples have the same number of trials, 72 times in each class, and details can be found in [16].

To improve the signal-to-noise ratios (SNR) of the EEG signals, we perform a simple preprocessing of the EEG signals of each dataset. First, the blank point data of "NaN" in the EEG data is set to 0 [17], and a 5-order Butterworth bandpass filter is used to filter the EEG signals with 4–32hz. To further reduce the noise between each channel, the common average reference method (CAR) commonly used in the spatial filtering of the EEG signals is processed [18], and finally, the EOG channel and artifacts are removed manually.

3. Classification algorithm based on multi-domain information fusion

The overall pipeline of the algorithm proposed in this paper is shown in Fig. 1. The details are as follows. Firstly, the raw EEG signal is pre-processed, and then the pre-processed EEG signals are selected using the channel selection strategy combined with time-frequency-space information, as shown in Fig. 3. Secondly, the feature extraction algorithm combined with time-frequency-space domain information is used to extract features from the optimal channel group, and the specific pipeline is shown in Fig. 5. Finally, the obtained fused features are fed into the multi-class support vector machine (SVM) classifier to obtain the classification results.

3.1. OVO-CSP algorithm

The CSP algorithm designs spatial filters based on two types of signals, and by simultaneously diagonalizing two covariance matrices that the EEG signal variance between different classes can be maximized to distinguish [19]. We briefly describe the CSP algorithm. In section 3.2.1 and section 3.3.1 we specify the improvements we have made to the CSP algorithm.

We consider the variable $X \in \mathbb{R}^N$ to represent the EEG data, where N is the number of recording channels, then the covariance matrix of a class of EEG signals can be expressed as

$$C_i = \frac{XX^T}{\text{trace}(XX^T)} \quad (1)$$

The composite spatial covariance matrix can be obtained as $C_c = \overline{C_a} + \overline{C_b}$, $\overline{C_a}$ and $\overline{C_b}$ respectively represent the average covariance matrix of the two classes of motor imagery tasks. And C_c can be decomposed into $C_c = U_c \lambda_c U_c^T$, where U_c is the eigenvector of the matrix and λ_c is the eigenvalue. The whitening transformation matrix P can be obtained by U_c and λ_c . Then the $\overline{C_a}$ and $\overline{C_b}$ can be transformed into $Z_a = P\overline{C_a}P^T$ and $Z_b = P\overline{C_b}P^T$. It can be obtained by proof that the whitened Z_a and Z_b have the same eigenvector B , such that

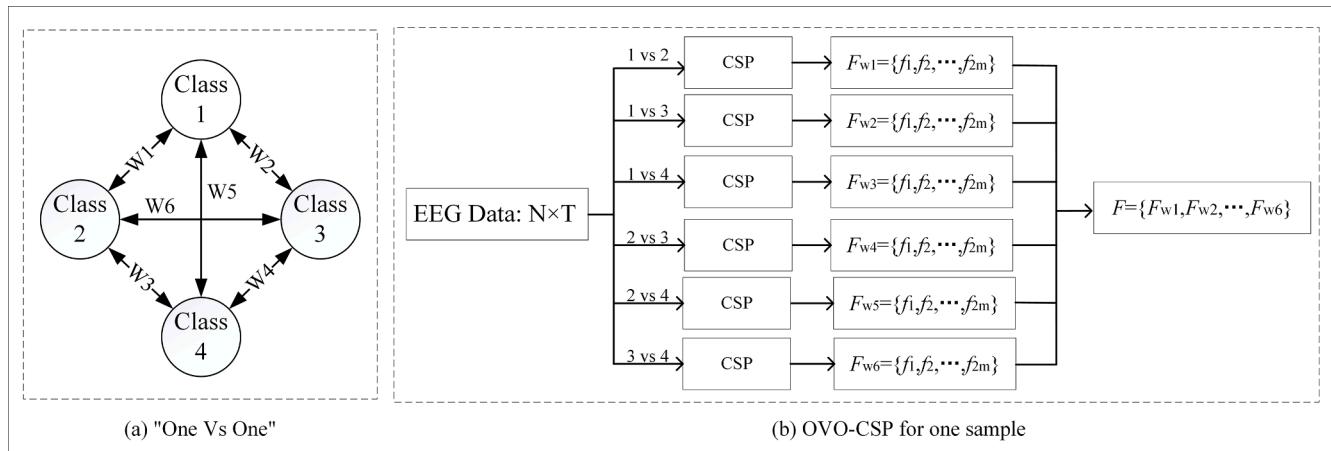


Fig. 2. One Versus One classification logic(a) and OVO-CSP feature extraction process for one sample(b). In (b), N is the number of channels and T is the value points in the sample.

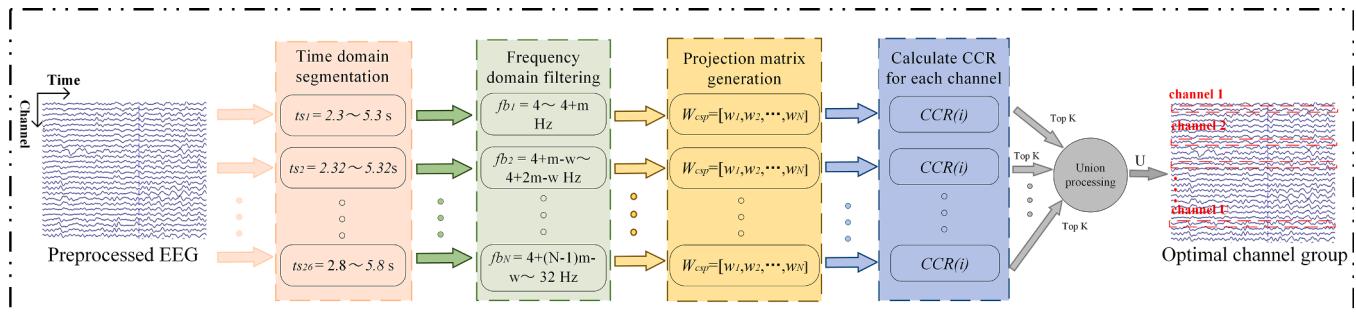


Fig. 3. Flowchart of channel selection strategy combined with time-frequency-space information. The ts_i represents the i-th time window, fb_i represents the i-th frequency band and W_{CSP} represents the generated projection matrix.

$$Z_a = B\lambda_a B^T \text{ and } Z_b = B\lambda_b B^T (\lambda_a + \lambda_b = I)$$

The projection matrix W can be calculated by the eigenvector B and the whitened EEG signal as.

$$W = B^T P \quad (3)$$

Select the front m and back m rows of the projection matrix W to form a spatial filter to obtain the feature, then the EEG data of the two types of task experiments E can be transformed into $S = W_{2m \times N}E$.

The projected signal $S_p (p = 1, \dots, 2m)$ is changed as the characteristic value as follows:

$$f_p = \lg \left(\frac{\text{var}(S_p)}{\sum_{i=1}^{2m} \text{var}(S_i)} \right) \quad (4)$$

where $\text{var}(S_p)$ denotes the variance of the p -th row component in S . Thus, the feature vector f_p is obtained by formula (4).

To solve the multi-classification tasks of motor imagery, we use the idea of the “one versus one” extension to extract features from the EEG data. The “one versus one” method aims to transform an N classification problem into $N \times (N-1)/2$ binary classification problems [20], which is shown in Fig. 2(a), the process of single-sample OVO-CSP feature extraction is shown in Fig. 2(b).

3.2. Channel selection strategy

As the number of channels selected increases, problems such as channel information redundancy, high computation and time overhead will appear. Due to the characteristic that the CSP can only achieve a significant effect when more electrode channels are input, how to eliminate useless channel information and increase the retention rate of

effective channel information is one of the directions that the CSP algorithm can optimize.

3.2.1. Channel screening criteria based on the optimal ordering of OVO-CSP projection space

Taking advantage of the characteristic that the spatial information of each channel is recorded in the projection space W , which generated during the process of the CSP, an optimization method combining 2 norm and *Frobenius* norm selection criteria is proposed. The specific steps of the method are as follows:

Step 1: Let $W_{CSP} \in \mathbb{R}^{N \times N}$ be the projection matrix calculated by formula (3). N is the number of channels. We denote the i -th column vector in W_{CSP} as w_i . The w_i represents the weight of each channel signal $x_i (i = 1, 2, \dots, N)$ in the projection space [21]. Based on the characteristics of projection space described above, a method for calculating the channel contribution rate (CCR) through the vector 2 norm and the matrix *Frobenius* norm is proposed:

$$CCR(i) = \frac{\|w_i\|_2}{\|W_{CSP}\|_F} \quad (6)$$

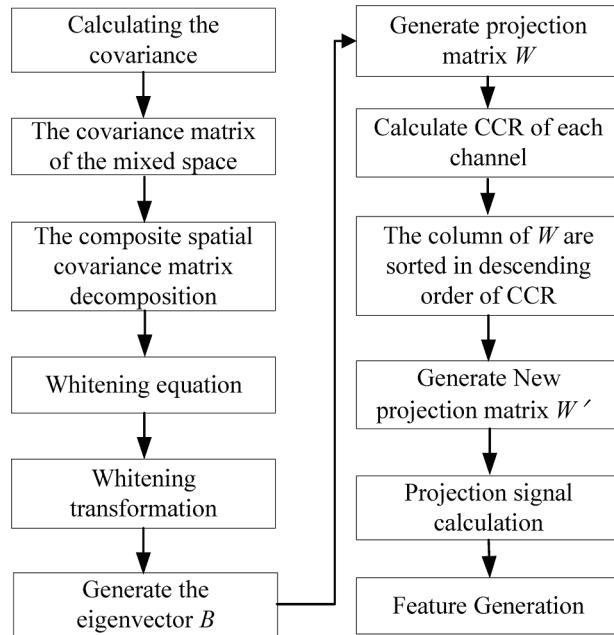
Step 2: Calculate the CCR of each channel and rank w_i in W_{CSP} in descending order according to their CCR. Then select the top K channels with the CCR as the optimized channel group. We denote the EEG data of optimal channel group as $E \in \mathbb{R}^K$, and the signal S' after spatial filtering is obtained by.

$$S' = W'_{CSP} E' = \begin{bmatrix} w_{11} & \dots & w_{1k} \\ \vdots & \ddots & \vdots \\ w_{k1} & \dots & w_{kk} \end{bmatrix} \begin{bmatrix} E_1 \\ \vdots \\ E_K \end{bmatrix} \quad (7)$$

Table 1

The parameter settings for the Dataset IIIa and Dataset 2a.

Parameter setting	m (Hz)	w (Hz)	N	Frequency sub-bands
Dataset IIIa	4	2	14	4–8 Hz, 6–10 Hz ... 26–30 Hz, 28–32 Hz
Dataset 2a	7	4	7	4–11 Hz, 7–14 Hz ... 22–29 Hz, 25–32 Hz

**Fig. 4.** Flowchart of OVO-CSP feature extraction optimization method based on the optimal ranking of projection space (OVO-CSSP).

3.2.2. Channel selection strategy combined with time-frequency-space information

In this paper, a channel selection strategy combined with time-frequency-space domain information is proposed, which is shown in Fig. 3. The strategy first performs time-domain segmentation and frequency-domain filtering on the preprocessed EEG signals to obtain different time-frequency domain sub-bands, and selects the optimal channel group by combining the channel screening criteria based on the optimal ordering of OVO-CSP projection space. The steps of this strategy include the following:

Step 1: The pre-processed EEG signals were intercepted for analysis using a set of time windows with fixed length 3 s, the time interval of each time window is 0.02 s, so the EEG data from 2.3 s ~ 5.8 s could be divided into 26 time windows.

Step 2: The filter banks filtering is performed for each time window. we use a set of second-order IIR filters with a frequency bandwidth of m Hz and an overlap width of w Hz [22]. We define that this operation generates N filters covering 4–32 Hz. Depending on the number of raw EEG channels, we provide two parameter settings, Table 1 gives an intuitive setting of the two datasets we used.

Step 3: The projection matrix generation operation is performed on the frequency sub-bands in each time window, and the CCR of each channel is calculated. The top K channels with the highest CCR are selected. However, there is a possibility that the top K channels selected in each frequency band are not the same, so we use the method of taking the union set to record all the channels that appear. Finally, the selected

U different channels are used as the optimized channel group, where U is the number of channels after doing the union processing. The parameter K is selected by traversing a range of values to obtain the optimal value, we choose the range of values from 7 to 20.

3.3. Feature extraction algorithm

We propose a feature extraction algorithm generating the feature vectors containing multi-domain information, which fully increase the characterization capability and richness of features [23]. The overall pipeline is shown in Fig. 5.

3.3.1. Feature extraction method based on one versus one-common spatial sorting patterns

Combining the OVO-CSP feature extraction method with the channel screening criteria based on the optimal ordering of the projection space, a feature extraction method based on the optimal ordering of the projection space in OVO-CSP can be summarized, which is named as OVO-CSSP.

The implementation process of this method is shown in Fig. 4. This method first performs the OVO-CSP process to obtain the projection matrix W as shown in formula (3), secondly, the CCR of each channel in the projection space is calculated according to formula (6), and then arranges the projection matrix column vectors in descending order according to the CCR to obtain the new projection matrix W' , then performs the projection signal calculation, finally, the feature vectors are obtained.

3.3.2. Feature extraction method based on Mu and Beta rhythm window energy

Mu (8–13 Hz) and Beta (14–30 Hz) rhythms are strongly related to the ERD and ERS phenomena in the motor cortex, so we use a light-weight computational but effective approach to extract the rhythmic window energy in these two relevant frequency bands [24].

The details of the approach are as follows: All value points in the time window are filtered to the frequency band corresponding to the Mu and Beta rhythm bands using a 5th order Butterworth filter, respectively. The window energy of the rhythm band is expressed by first calculating the sum of squares of all sample points in the window and then taking the logarithm [25]. Therefore, the window energy of channel k on Mu or Beta rhythm is written as follows:

$$f_{ch_k} = \ln \left(\frac{1}{M} \sum_{i=1}^M v_i^2 \right) \quad (8)$$

Where M is the number of value points in the time window, v_i^2 represents the square of the i -th value point.

3.3.3. Feature extraction algorithm combined with time-frequency-space information

The steps of the feature extraction algorithm combining time-frequency-space information are as follows:

Step 1: A set of time windows is used to intercept the optimal channel group of EEG singals for analysis, the method and parameter setting of time windows generated are the same way as step 1 in section 3.2.2. Then Mu and Beta rhythm window energy based on the optimal channel group are extracted, which we define the feature vectors as F_1 ;

Step 2: Perform filter banks frequency domain filtering on the EEG signals in the time window. The filtering method is consistent with the processing method of step 2 in section 3.2.2. The OVO-CSSP feature extraction is performed for each generated frequency domain sub-band in the time window, and the feature vectors obtained from all frequency bands need to concatenate, which we define the feature vectors as F_2 ;

Step 3: We fuse the feature vectors generated by Step1 and Step2 in

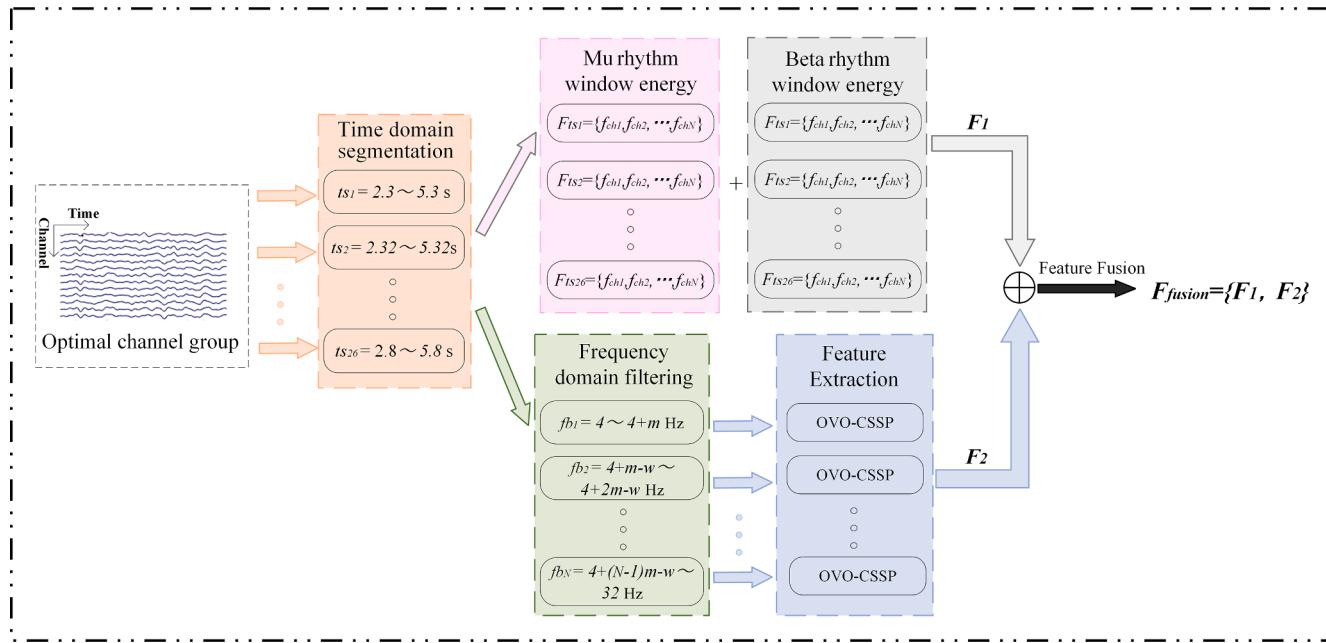


Fig. 5. Flowchart of feature extraction algorithm combined with time-frequency-space information.

the form $F_{\text{fusion}} = \{F_1, F_2\}$, and we define F_{fusion} to be the final feature vectors.

4. Experimental results

4.1. Classifier and evaluation indicators

SVM is one of the most commonly used classifiers in the research of multi-class MI classification tasks [26], not only it can achieve good

Table 2

The recognition rate, Kappa value, and the number of selected channels of three subjects from the Dataset IIIa.

Subject	k3b	k6b	l1b	Mean
Recognition rate/%	97.987	75.904	83.133	85.675
Kappa value	0.973	0.679	0.775	0.809
Number of channels	17	15	15	—

Table 3

The recognition rate, Kappa value, and the number of selected channels of nine subjects from the Dataset 2a.

Subject	A01	A02	A03	A04	A05	A06	A07	A08	A09	Mean
Recognition rate/%	90.036	59.364	86.081	71.053	58.333	54.419	91.336	85.239	79.167	75.00
Kappa value	0.867	0.458	0.814	0.614	0.444	0.392	0.884	0.803	0.722	0.667
Number of channels	17	13	22	18	21	19	17	22	10	—

Table 4

Comparison of classification accuracy of Dataset2a subjects using different feature extraction methods.

Methods	Subjects and their classification accuracy(%)									
	A01	A02	A03	A04	A05	A06	A07	A08	A09	Mean
F1(m = 1)	48.26	34.38	67.36	37.15	25.35	25.35	42.36	68.40	63.54	45.79
F1(m = 2)	56.60	36.81	68.40	40.97	25.34	29.17	46.53	72.57	68.75	49.46
F1(m = 3)	60.07	42.71	69.10	43.40	25	32.64	48.96	67.01	70.14	51.00
F2(m = 1)	81.94	57.98	75.69	64.24	62.85	45.14	81.94	77.08	69.44	68.48
F2(m = 2)	81.25	55.90	75.69	59.72	59.72	43.40	82.64	78.13	63.19	66.63
F2(m = 3)	75.69	55.56	76.04	61.11	59.38	38.19	80.56	76.04	64.24	65.20
F3	87.19	58.66	85.45	68.42	58.06	52.09	89.89	84.24	75.76	73.31
F4	90.04	59.36	86.08	71.05	58.33	54.42	91.34	85.24	79.17	75.00

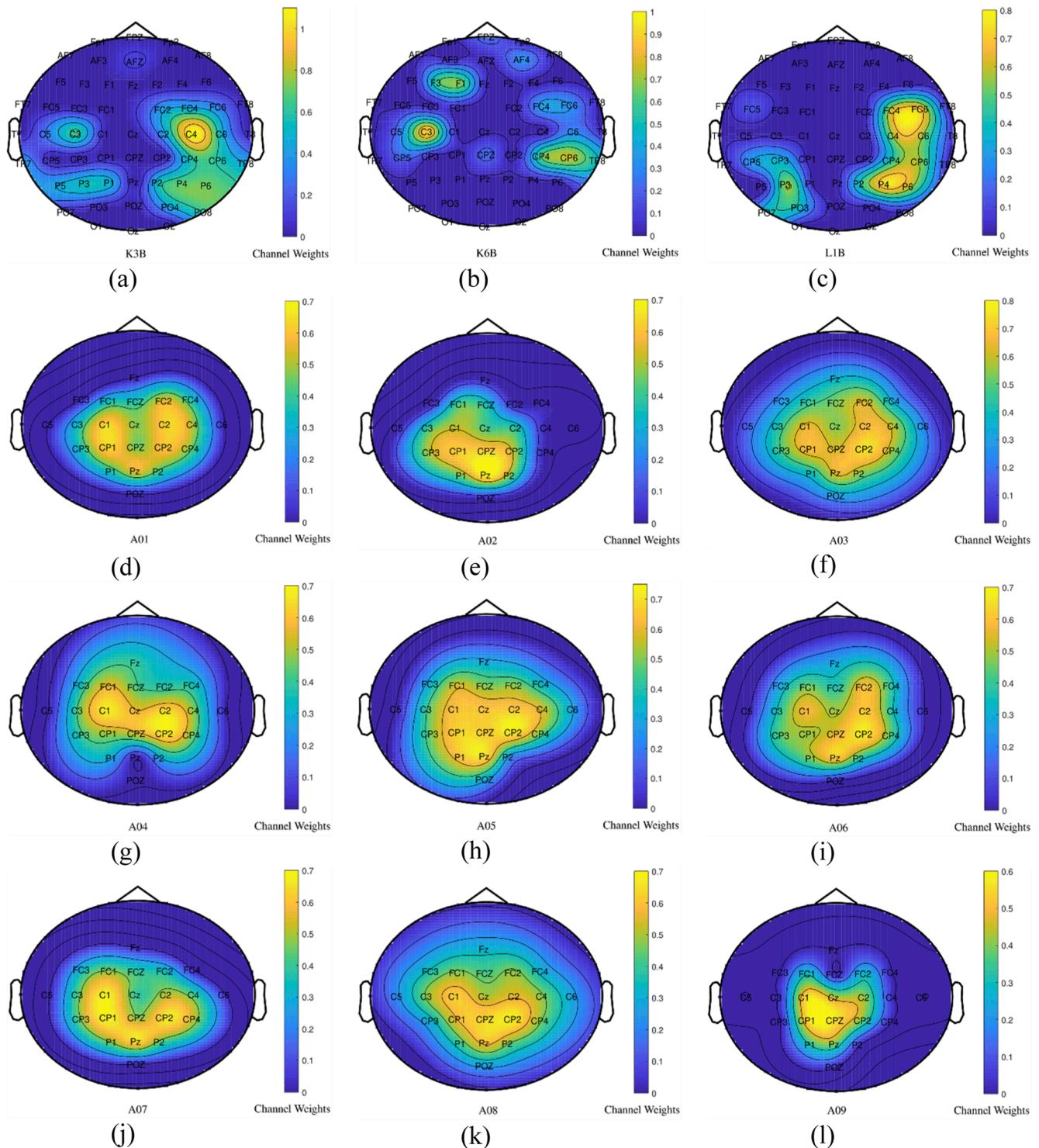


Fig. 6. Brain topology map of optimal channel group weight in every subject.

It can be seen from Table 2 that three subjects, k3b, k6b, and l1b have achieved the average classification accuracy of 85.68 % and the average kappa value of 0.81. The Dataset IIIa uses 60 channels (excluding the ocular electrical channels) to collect EEG data. Subjects k6b and l1b select only 15 channels, which accounts for the number of original channels 25 %. In Table 3, A01 ~ A09 achieve the average classification accuracy of 75.00 %, and the average kappa value of 0.67. The Dataset 2a only uses 22 channels (excluding the ocular electrical channels) to collect EEG data, 2/9 of the subjects have a full selection of channels but the remaining subjects have a significant reduction in the number of

channels selected, the best selection effect is 10 channels, accounting for 45 % of the original number of channels.

To verify the effectiveness of the algorithm, this paper takes nine subjects of the Dataset 2a as an example and conducts comparative experiments using different feature extraction methods. F1, F2 is used to denote the conventional OVO-CSP algorithm, OVO-FBCSP algorithm [28], respectively. F3 represents the algorithm we proposed but does not include the feature extraction method based on Mu and Beta rhythm window energy, and F4 denotes the complete algorithm we proposed. Table 4 shows the results of the highest classification accuracy obtained

Table 5

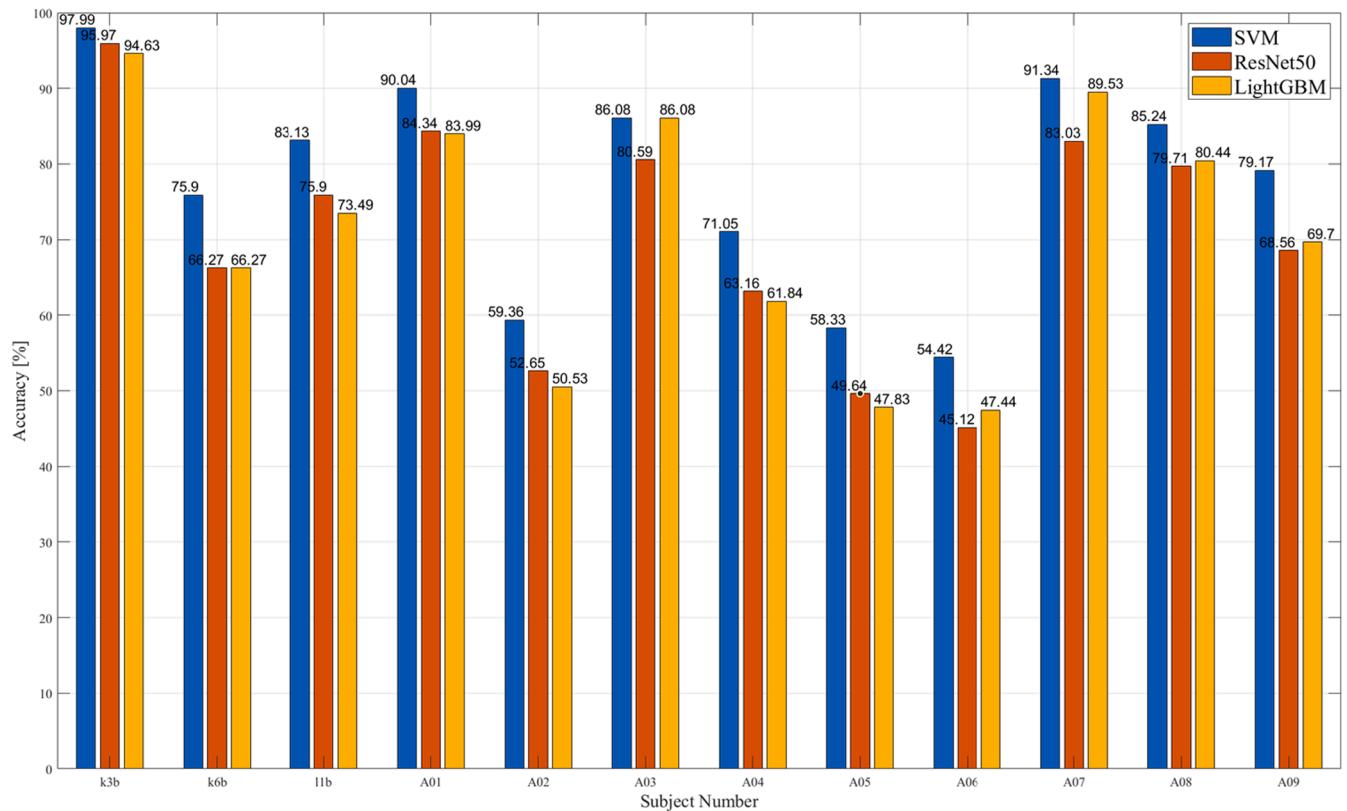
Classification recognition rate (%) based on different channel combination.

Subject	C3, C4, Cz	Combination of 11 channels	All Channel	The strategy of this paper
k3b	82.55	93.29	97.32	97.99
k6b	44.58	68.67	69.88	75.90
l1b	71.08	82.34	81.93	83.13
A01	65.84	83.99	88.61	90.04
A02	49.12	53.00	55.48	59.36
A03	67.40	79.12	86.08	86.08
A04	43.86	65.35	69.30	71.05
A05	44.20	55.43	57.97	58.33
A06	42.33	55.35	54.42	54.42
A07	70.04	84.84	88.44	91.34
A08	62.73	81.18	85.24	85.24
A09	63.26	76.52	74.24	79.17

from F1,F2,F3,F4, where m represents the number of groups of selected spatial filters in the OVO-CSP algorithm and the OVO-FBCSP algorithm.

It can be found from Table 4 that for all subjects, the proposed method in this paper have a significant improvement in recognition rate compared with F1, F2 feature extraction algorithms. Specifically, the accuracy rate of the proposed algorithm has improved by about 26.2 % compared with the OVO-CSP algorithm and about 8.2 % compared with the OVO-FBCSP. By comparing with the F3 feature extraction method, it is found that the fused features after adding Mu and Beta rhythm window energy can effectively improve the classification performance.

A large number of studies([29,30,31]) have proved that people will induce EEG signals in the brain area near the cerebral motor cortex according to performing different motor imagery tasks. To verify whether the channel selection strategy combining time-frequency-space information is consistent with neurophysiological characteristics, we use brain topology maps to confirm whether the selected optimal

**Fig. 7.** Comparison of classification accuracy between different classifiers for all subjects.**Table 6**

Comparison of kappa value between the method in this paper and other paper methods.

Subjects	Method of this paper	CCSP-SVM [33]	Siamese NN [34]	OVO-FBCSP CNN [35]	BSC CSP-SVM [36]	BCI III1st [37]	BCI IV1st [38]
k3b	0.973	0.94	—	—	0.800	0.822	—
k6b	0.679	0.63	—	—	0.533	0.756	—
l1b	0.775	0.70	—	—	0.777	0.800	—
A01	0.867	0.72	0.819	0.758	—	—	0.68
A02	0.458	0.40	0.340	0.440	—	—	0.42
A03	0.814	0.70	0.788	0.676	—	—	0.75
A04	0.614	0.55	0.392	0.523	—	—	0.48
A05	0.444	0.20	0.340	0.504	—	—	0.40
A06	0.392	0.35	0.389	0.268	—	—	0.27
A07	0.884	0.66	0.434	0.759	—	—	0.77
A08	0.803	0.78	0.705	0.695	—	—	0.75
A09	0.722	0.77	0.778	0.592	—	—	0.61
Average(k3b-l1b)	0.809	0.793	—	—	0.704	0.793	—
Average(A01-A09)	0.667	0.57	0.554	0.579	—	—	0.57
Average	0.702	0.617	—	—	—	—	—

Table 7

Time consumption statistics of the algorithms used in this paper.

Time consumption(s)		Time window length	Select the optimal channel group	Feature extraction	Classification
Training		3	0.011	0.01	24
Testing		3	—	0.01	0.065

channel group distribution and the weight of each channel in the motor imagery tasks are consistent with the neurophysiological knowledge.

Fig. 6 shows the brain topology map of the optimal channel group distribution selected by the 12 subjects according to the algorithm we proposed, the weight value of each channel in the optimal channel group represents the CCR in the projection space, and the weight value of the unselected channel is 0. As shown in the figure, the optimal channel groups of the 12 subjects screened by the algorithm we proposed are mainly located in the cerebral motor cortex area (near the C3 and C4 electrodes), which is consistent with the fact that motor imagery evokes relevant EEG signals in the motor cortex region of the brain. The above discussion indicates that our channel selection strategy is viable.

Based on the distribution of channels, it can be concluded that although the distribution of the optimal channel group in different subjects is concentrated in the motor areas of the cerebral cortex, the most suitable channels are various for different subjects. For example, in Fig. 6, the most important channels of subjects k3b and k6b are located at C4 and C3 electrodes, subjects A01 and A09 are located at C1 and CP1 electrodes, while subjects A03 and A08 are located at CP2 electrode, etc.

Three different channel combination methods are considered to compare with the optimal channel groups selected by the method we proposed, including the most common combination of the C3, C4, Cz channels, the channel group formed by manually selecting 11 channels concentrated in the motor cortex of the brain (FC1, FCZ, FC2, C3, C1, CZ, C2, C4, CP1, CPZ, CP2), and the all channels. Results of using the same feature extraction method proposed in section 3.3.3 are shown in Table 5.

Among the four channel selection methods, the best results were obtained using the channel selection strategy based on multi-domain information fusion. This shows that our proposed channel selection strategy selects the channel combination that matches the neural information of the subjects, which verifies the effectiveness and practicability of the strategy. The results of Table 5 also verify that using a suitable channel selection strategy is more scientific and accurate than the conventional way of channel selection, such as selecting a fixed combination of channels or selecting all channels.

Furthermore, we also used Deep Learning Network such as ResNet50 [32] and the machine learning classifier algorithms based on an integrated idea such as LightGBM [33] to classify the feature vectors extracted in this paper, however, the results were slightly less effective than SVM on all subjects, as shown in Fig. 7.

Table 6 compares the results of the kappa value achieved by our method with other methods using the same dataset in the last three years and with the first place winners of the two competitions. Ghanbar et al. [34] uses an algorithm based on temporal correlation CSP regularization algorithm combined with SVM for classification. Shahtalebi et al. [35] uses a Siamese Neural Networks algorithm that combines OVR and OVO ideas for feature extraction and classification. Holm et al. [36] uses an OVO-FBCSP combined with a CNN algorithm for classification. Jin et al. [37] proposed a channel selection method combining logarithmic amplitude with first-order spectral moment feature (BCS) and combined with CSP algorithm and SVM for classification. The first place in the BCI Competition III Dataset IIIa [38] uses the CSP algorithm to calculate the Fisher ratio of the time-frequency domain channel for channel selection and the SVM algorithm for classification. The first place in the BCI Competition IV Dataset 2a [39] uses the OVR method to extend the improved filter bank CSP to multiple classes, and the classifier used is the Naive Bayes Parzen Window classifier. It is proved through experiments that the kappa values obtained by the algorithm proposed in this

paper are all higher than the above-mentioned methods.

The satisfied results of this proposed algorithm can be attributed to the following reasons. On the one hand, the channel selection strategy proposed in this paper can better select the channel combination suitable for the current subject, reduce the information interference of useless channels, and lay a good foundation for subsequent feature extraction. On the other hand, the feature extraction algorithm proposed in this paper can generate feature vectors containing time-frequency-space information, so that the features describing EEG signals are no longer single, more sufficient description information can be obtained [40].

Further discuss the generalization ability of the algorithm we proposed. From the results, our proposed algorithm is adaptive, which indeed has generalization ability. The reasons are as follows, from the perspective of time domain, we found that in the interval of 2.3 s-5.8 s, the subjects' motor imagery is the most active and the most suitable for analysis; from the perspective of frequency domain, two parameter settings are proposed to cope with more and less original channels (the number of channels of DatasetIIIa is nearly three times that of Dataset2a). However, when conducting cross-subject experiments, the prerequisite is to use the same number of channels to record raw EEG, so the parameters in frequency domain is consistent for each subject; further analysis of the channel selection strategy proposed, the optimal channel group can be obtained by the union processing of top K channels in each time-frequency subband. The selection of parameter K is done by determining a range of values and obtaining the best effect by traversal. Such a uniform process and parameter setting achieve excellent and effective results on the 12 subjects, where there is no manual intervention. So it is reasonable to infer that this algorithm can achieve excellent results on other subjects as well.

Further discussing the effect of classes growing on the computational complexity of the algorithm in this paper, the effort of some operations are independent of the class number, such as time windows generation in the time domain and filter banks filtering in the frequency domain, but both the CCR calculation and the OVO-CSSP feature extraction method are based on the idea of the "one versus one", which convert N classification problems into $N \times (N-1)/2$ binary classification problems, as the number of classes boosting, the computation of these parts will increase. However, we conduct time-consuming statistics on each part of the proposed algorithm, as shown in Table 7. The results show that the time consumption of the proposed algorithm can well meet the requirements of the online BCI system.

5. Conclusion

In this paper, we propose a novel classification algorithm based on multi-domain information fusion. The algorithm solves some drawbacks of the traditional CSP including: the contradiction between powerful performance and numerous input channels, the lack of time-frequency domain information, the poor feature extraction performance on multi-classes MI EEG signals. The major contribution of this work is to use less number of channels, but achieve a promising classification results. The proposed channel selection strategy can adapt to the neural information contained in the EEG signals of subjects, which can select optimal channel groups located at the motor areas of the cerebral cortex, further excluding the interference of channels not related to the MI tasks. Moreover, The excellent classification results achieved on the two datasets we used show that the feature vectors extracted by our method are characterized by strong representation ability and rich information.

The follow-up work of this paper is to continue to study and optimize the proposed algorithm, and to design an online system of MI-BCI based on the algorithm of this paper using the existing conditions in our laboratory.

CRediT authorship contribution statement

Jiaqi Wang: Conceptualization, Software, Writing – original draft.
Wanzhong Chen: Formal analysis, Supervision. **Mingyang Li:** Methodology, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the Science and Technology Development Plan in Jilin Province of China (Grant No. 20190302034GX), and National Natural Science Foundation of China (Grant No. 62203183)

References

- [1] J.R. Wolpaw, N. Birbaumer, W.J. Heetderks, et al., Brain-computer interface technology: a review of the first international meeting, *IEEE Trans. Rehabil. Eng.* 8 (2000) 164–173, <https://doi.org/10.1109/TRE.2000.847807>.
- [2] J.J. Shih, D.J. Krusinski, J.R. Wolpaw, Brain-computer interfaces in medicine, *Mayo Clinic Proc.* 87 (2012) 268–279. doi: 10.1016/j.mayocp.2011.12.008.
- [3] N. Padfield, J. Zabalza, H. Zhao, et al., EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges, *Sensors.* 19 (2019) 1423, <https://doi.org/10.3390/s19061423>.
- [4] G. Pfurtscheller, F.H. Lopes Da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles, *Clin. Neurophysiol.* 110 (1999) 1842–1857, [https://doi.org/10.1016/S1388-2457\(99\)00141-8](https://doi.org/10.1016/S1388-2457(99)00141-8).
- [5] B. Blankertz, et al., Optimizing spatial filters for robust EEG single-trial analysis, *IEEE Signal Process Mag.* 25 (1) (2007) 41–56, <https://doi.org/10.1109/MSP.2008.4408441>.
- [6] Ang, Kai Keng, et al., Filter bank common spatial pattern (FBCSP) in brain-computer interface, in: IEEE international joint conference on neural networks, 2008, pp. 2390–2397. doi:10.1109/IJCNN.2008.4634130.
- [7] H.V. Shenoy, A.P. Vinod, C. Guan, Shrinkage estimator based regularization for EEG motor imagery classification, in: 2015 10th International Conference on Information, Communications and Signal Processing (ICICS), 2015, pp. 1–5, <https://doi.org/10.1109/ICICS.2015.7459836>.
- [8] F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms, *IEEE Trans. Biomed. Eng.* 58 (2010) 355–362, <https://doi.org/10.1109/TBME.2010.2082539>.
- [9] O. Falzon, K.P. Camilleri, J. Muscat, The analytic common spatial patterns method for EEG-based BCI data, *J. Neural Eng.* 9 (2012) 045009, <https://doi.org/10.1088/1741-2560/9/4/045009>.
- [10] J. Luo, J. Wang, R. Xu, et al., Class discrepancy-guided sub-band filter-based common spatial pattern for motor imagery classification, *J. Neurosci. Methods* 323 (2019) 98–107, <https://doi.org/10.1016/j.jneumeth.2019.05.011>.
- [11] M. Grosse-Wentrup, M. Buss, Multiclass common spatial patterns and information theoretic feature extraction, *IEEE Trans. Biomed. Eng.* 55 (8) (2008) 1991–2000, <https://doi.org/10.1109/TBME.2008.921154>.
- [12] A. Barachant, S. Bonnet, et al., Multiclass brain-computer interface classification by Riemannian geometry, *IEEE Trans. Biomed. Eng.* 59 (4) (2011) 920–928, <https://doi.org/10.1109/tbme.2011.2172210>.
- [13] O. Ozdenizci, D. Erdogmus, Information theoretic feature transformation learning for brain interfaces, *IEEE Trans. Biomed. Eng.* 67 (1) (2019) 69–78, <https://doi.org/10.1109/TBME.2019.2908099>.
- [14] Y. Park, W. Chung, Optimal channel selection using correlation coefficient for CSP based EEG classification, *IEEE Access* 8 (2020) 111514–111521, <https://doi.org/10.1109/ACCESS.2020.3003056>.
- [15] B. Blankertz, K.R. Muller, D.J. Krusinski, et al., The BCI competition III: Validating alternative approaches to actual BCI problems, *IEEE Trans. Neural Syst. Rehabil. Eng.* 14 (2006) 153–159, <https://doi.org/10.1109/TNSRE.2006.875642>.
- [16] M. Tangermann, K.R. Müller, A. Aertsen, et al., Review of the BCI competition IV, *Front. Neurosci.* 6 (2012) 55, <https://doi.org/10.3389/fnins.2012.00055>.
- [17] S. Qu, J. Liu, W. Chen, et al., Pattern recognition of motor imagery EEG signal in noninvasive brain-computer interface, in: 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2018, pp. 1814–1819. doi: 10.1109/ICIEA.2018.8398003.
- [18] D. Wu, J.T. King, C.H. Chuang, et al., Spatial filtering for EEG-based regression problems in brain-computer interface (BCI), *IEEE Trans. Fuzzy Syst.* 26 (2017) 771–781, <https://doi.org/10.1109/TFUZZ.2017.2688423>.
- [19] L. He, Y. Hu, Y. Li, et al., Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG, *Neurocomputing.* 121 (2013) 423–433, <https://doi.org/10.1016/j.neucom.2013.05.005>.
- [20] E. Dong, C. Li, L. Li, et al., Classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain–computer interfaces, *Medical Biol. Eng. Comput.* 55 (2017) 1809–1818, <https://doi.org/10.1007/s11517-017-1611-4>.
- [21] C. Zich, M. De Vos, C. Kranczioch, et al., Wireless EEG with individualized channel layout enables efficient motor imagery training, *Clin. Neurophysiol.* 126 (2015) 698–710, <https://doi.org/10.1016/j.clinph.2014.07.007>.
- [22] S. Chan, Carson, K. Pun, et al., On the design and implementation of FIR and IIR digital filters with variable frequency characteristics, in: *Circuits & Systems II Analog & Digital Signal Processing IEEE Transactions on.* 49(11, 2002, pp. 689–703. doi: 10.1109/TCSII.2002.807574.
- [23] C. Xu, C. Sun, G. Jiang, et al., Two-level multi-domain feature extraction on sparse representation for motor imagery classification, *Biomed. Signal Process. Control* 62 (2020) 102160, <https://doi.org/10.1016/j.bspc.2020.102160>.
- [24] C. Neuper, G.R. Müller-Putz, R. Scherer, et al., Motor imagery and EEG-based control of spelling devices and neuroprostheses, *Prog. Brain Res.* 159 (2006) 393–409, [https://doi.org/10.1016/S0079-6123\(06\)59025-9](https://doi.org/10.1016/S0079-6123(06)59025-9).
- [25] M.S. Bascil, A.Y. Tesneli, F. Temurtas, Multi-channel EEG signal feature extraction and pattern recognition on horizontal mental imagination task of 1-D cursor movement for brain computer interface, *Australas. Phys. Eng. Sci. Med.* 38 (2015) 229–239, <https://doi.org/10.1007/s13246-015-0345-6>.
- [26] A. Mahmood, R. Zainab, R.B. Ahmad, et al., Classification of multi-class motor imagery EEG using four band common spatial pattern, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017, pp. 1034–1037, <https://doi.org/10.1109/EMBC.2017.8037003>.
- [27] F. Xu, W. Zheng, D. Shan, et al., Decoding spectro-temporal representation for motor imagery recognition using ECoG-based brain-computer interfaces, *J. Integ. Neurosci.* 19 (2020) 259–272, <https://doi.org/10.31083/j.jin.2020.02.1269>.
- [28] Z.Y. Chin, K.K. Ang, C. Wang, et al., Multi-class filter bank common spatial pattern for four-class motor imagery BCI, in: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 571–574. doi: 10.1109/EMBS.2009.5332383.
- [29] F.P. de Lange, K. Roelofs, I. Toni, Motor imagery: a window into the mechanisms and alterations of the motor system, *Cortex.* 44 (2008) 494–506, <https://doi.org/10.1016/j.cortex.2007.09.002>.
- [30] J. Munzert, B. Lorey, K. Zentgraf, Cognitive motor processes: the role of motor imagery in the study of motor representations, *Brain Res. Rev.* 60 (2009) 306–326, <https://doi.org/10.1016/j.brainresrev.2008.12.024>.
- [31] S.C. Wriessnegger, C. Brunner, G.R. Müller-Putz, Frequency specific cortical dynamics during motor imagery are influenced by prior physical activity, *Front. Psychol.* 9 (2018) 1976, <https://doi.org/10.3389/fpsyg.2018.01976>.
- [32] K. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition,* 2016, pp. 770–778, <https://doi.org/10.1109/CVPR.2016.90>.
- [33] G. Ke, Q. Meng, T. Finley, et al., Lightgbm: A highly efficient gradient boosting decision tree, *Adv. Neural Inform. Process. Syst.* 30 (2017) 3146–3154.
- [34] K.D. Ghanbar, T.Y. Rezaei, A. Farzamnia, et al., Correlation-based common spatial pattern (CCSP): A novel extension of CSP for classification of motor imagery signal, *PLoS ONE* 16 (2021) 1–18, <https://doi.org/10.1371/journal.pone.0248511>.
- [35] S. Shahtalebi, A. Asif, A. Mohammadi, Siamese neural networks for EEG-based brain-computer interfaces, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, <https://doi.org/10.1109/EMBC44109.2020.9176001>.
- [36] N.S. Holm, S. Puthusserapady, An improved five class MI based BCI scheme for drone control using filter bank CSP, in: 2019 7th International Winter Conference on Brain-Computer Interface, 2019, p. (BCI).1-6, <https://doi.org/10.1109/IWW-BCI.2019.8737263>.
- [37] J. Jin, C. Liu, I. Daly, et al., Bispectrum-based channel selection for motor imagery based brain-computer interfacing, *IEEE Trans. Neural Syst. Rehabil. Eng.* 28 (2020) 2153–2163, <https://doi.org/10.1109/TNSRE.2020.3020975>.
- [38] Berlin Brain-Computer Interface. BCI competition III -final results [Online], available: <http://www.bbci.de/competition/iii/results/index.html>, July21, 2022.
- [39] Berlin Brain-Computer Interface. BCI competition IV -final results [Online], available: <http://www.bbci.de/competition/iv/results/index.html>, July21, 2022.
- [40] D. Gajic, Z. Djurovic, J. Gligorijevic, et al., Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis, *Front. Comput. Neurosci.* 9 (2015) 38, <https://doi.org/10.3389/fncom.2015.00038>.