Biomedical Signal Processing and Control 77 (2022) 103825

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Overall optimization of CSP based on ensemble learning for motor imagery EEG decoding

Shaorong Zhang [a,b,c], Zhibin Zhu [d,*], Benxin Zhang [e], Bao Feng [c], Tianyou Yu [f], Zhi Li [c], Zhiguo Zhang [a,b], Gan Huang [a,b], Zhen Liang [a,b]

[a] School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China
[b] Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging, Shenzhen 518060, China
[c] School of Electronic Information and Automation, Guilin University of Aerospace Technology, Guilin 541004, China
[d] School of Mathematics and Computing Science, Guilin University of Electronic Technology, Guilin 541004, China
[e] School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China
[f] School of Automation Science and Engineering, South China University of Technology, Guangzhou 510000, China

ARTICLE INFO

keywords:
Brain-computer interface
EEG decoding
Motor imagery
Common spatial pattern
Ensemble learning
LASSO

ABSTRACT

The common spatial pattern (CSP) is an effective feature extraction method in motor imagery-based brain-computer interface (BCI) system. However, CSP also has many defects. Existing CSP improvement methods only make partial improvements, without considering the overall optimization of CSP. In this paper, a new ensemble learning algorithm framework is proposed to improve the decoding performance of CSP, in which the regularization, temporal-spatial-frequency joint optimization, and pair number of spatial filters for CSP are comprehensively considered. First, a new temporal-spatial-frequency feature extraction method based on Tikhonov regularization CSP (TRCSP) is proposed, multiple feature subsets with diversity are extracted by TRCSP with different time windows, regularization parameters, and pair numbers of spatial filters. Second, the least absolute shrinkage and selection operator (LASSO) as base classification model is used for feature selection and classification, in which multiple diversified base classification models are trained. Finally, the base classification models with diversity and higher accuracy are used for ensemble model construction using a new integration rule, during which most of the temporal-spatial-frequency information is fully excavated and utilized. The effectiveness of the proposed method is verified by five motor imagery data sets and the average classification accuracy of all data sets is 85.99%. Compared with the existing CSP methods, the proposed method achieved a better classification effect, and with a small amount of calculation, low model complexity, and high robustness.

1. Introduction

The brain-computer interface (BCI) system based on motor imagery has huge advantages in rehabilitation training and motor control [1]. However, the signal-to-noise ratio of the motor imagery electroencephalogram (EEG) signals is low, and the randomness is strong, especially the individual difference and non-stationarity are very obvious [2], resulting in the low accuracy and stability of the motor imagery EEG decoding, which directly affects the use effect of BCI system. Therefore, EEG decoding has always been the focus and difficulty of BCI system research.

A lot of studies have shown that the common spatial pattern (CSP) based EEG decoding method is effective for motor imagery signals [3,4]. However, CSP has two main defects. First, CSP is implemented based on sample covariance, which is sensitive to noise. When the sample size is relatively small, overfitting is easy to occur due to inaccurate covariance estimation [5]. Second, the effectiveness of CSP depends on the selection of frequency band and time window [6]. How to effectively select the optimal frequency band and time window is still a problem that has not been completely solved in EEG decoding.

In view of the two defects of CSP, a lot of CSP improvement methods have been proposed. The first class of methods uses the regularization technique to alleviate the noise sensitivity and overfitting problems of CSP. Lotte et al [7] proposed four new regularized CSP methods and

Abbreviations: EEG, electroencephalogram; LASSO, least absolute shrinkage and selection operator.

* Corresponding author.
E-mail address: optimization_zhu@163.com (Z. Zhu).

https://doi.org/10.1016/j.bspc.2022.103825
Received 15 January 2022; Received in revised form 9 April 2022; Accepted 16 May 2022
Available online 21 May 2022

compared 11 different regularized CSP methods. Literatures [8] and [9] respectively proposed two CSP improvement methods based on l_1 norm regularization. In order to address the overfitting issue of CSP, Onaran et al [10] proposed a smooth version of l_1 norm regularized CSP method. Li et al [11] proposed a sparse version of CSP-L1 [8,9]. Recently, Qi et al [12] used group sparsity to regularize the spatio-temporal filter, and Wang et al [13] reformulated CSP as a constrained minimization problem and established the equivalence of the reformulated and the original CSPs.

Compared with the regularization improvement work of CSP, more researches focus on the optimization and selection of frequency band and time window for CSP. The frequency band selection of CSP includes filter bank-based methods such as FBCSP [14], SFBCSP [15], and SBLFB [16], as well as intelligent optimization-based methods such as particle swarm optimization (PSO) [17]. Similarly, there is a lot of work for the selection of time window, such as literatures [18–22]. In recent years, the joint optimization of frequency band and time window has received more and more attention and research. There are many such types of researches, including the selection of optimal frequency bands and time windows based on methods such as PSO [23] and Fisher score [24], as well as temporal-spatial-frequency joint optimization methods [25–28].

In addition, much existing research work has demonstrated that the pair number of spatial filters has an impact on the performance of CSP and its improved method [29–31]; too many spatial filters will lead to information redundancy, and too few will lead to information loss. Therefore, the selection of the pair number of spatial filters is also very important.

To sum up, the regularization, time window, frequency band, and the pair number of spatial filters all have a significant impact on the performance of CSP. If only one of the influencing factors is considered, CSP will not achieve optimal performance. Although the existing CSP improvement methods have achieved better EEG decoding performance, most of the improvement methods only make partial improvements to CSP, without considering the overall optimization of CSP. In this paper, we will propose a new algorithm framework based on ensemble learning [32], in which the main influencing factors of CSP are comprehensively considered.

Inspired by literature [33], Tikhonov regularization CSP (TRCSP) can be integrated into the temporal-spatial-frequency joint optimization to simultaneously solve the noise sensitivity and overfitting problems of CSP, as well as the selection of time windows, frequency bands, and the pair number of spatial filters. Furthermore, the joint optimization of temporal-spatial-frequency can also be integrated into the ensemble learning framework to make full use of the temporal-spatial-frequency information of EEG [34]. In addition, ensemble learning can effectively alleviate the problem of the non-stationarity of EEG signals [35,36] to a certain extent, thereby improving the stability and robustness of EEG decoding methods.

Based on the above analysis, a new EEG decoding algorithm framework based on ensemble learning is proposed in this paper. First, signal preprocessing. The band-pass filtering with 8–30 Hz is performed on the original EEG signals, and then EEG data of three time windows with 0.5–2.5 s, 1–3 s, and 1.5–3.5 s are extracted. Second, feature extraction. For each time window, TRCSP with different regularization parameters and pair numbers of spatial filters is used for spatial filtering, and then the spatially filtered signals are decomposed into multiple sub-bands using a filter bank. For each sub-band, logarithmic variance is extracted as the feature. The diversified feature subsets can be obtained by combination of different time windows, regularization parameters, and pair numbers of spatial filters. Third, feature selection and classification. The least absolute shrinkage and selection operator (LASSO) model is trained to achieve diversified base classification models. In the test stage, the base models with the top 80% cross-validation accuracy in the training stage are selected for ensemble, the output values of the base models are integrated using a weight sum rule. The effectiveness of the proposed method is validated on four public data sets from the BCI competition and one self-collected data set from our laboratory.

The main contributions of this paper are summarized as follows:

1) A new algorithm framework based on ensemble learning is proposed, in which the regularization, temporal-spatial-frequency joint optimization, and the pair number of spatial filters for improving CSP are comprehensively considered. The proposed ensemble learning framework is deeply integrated with the feature extraction and feature selection process of motor imagery EEG, in which the temporal-spatial-frequency information is fully excavated and utilized. In addition, a new model integration rule is proposed.
2) A new temporal-spatial-frequency feature joint optimization method is proposed, which has small amount of calculation, low model complexity, and richer spatial information. In the feature extraction part, the TRCSP (or CSP) spatial projection is used to reduce the dimension of the original EEG signals, and then the filter bank is used for band-pass filtering, which significantly reduces the amount of calculation. In the feature ensemble selection part, a simple LASSO model is used as the base classifier, and with a small ensemble scale. In addition, the extracted features not only contain spatial information of different time–frequency segmentations, but also the spatial information generated by the combination of different regularization parameters and pair numbers of spatial filters.
3) A new method of generating feature diversity is proposed. TRCSP with different time windows, regularization parameters, and pair numbers of spatial filters is used to generate diversified feature subsets.
4) A large number of data sets are used to verify the effectiveness, universality, and robustness of the proposed method. In the existing literature, only 2 to 3 data sets are used, and the number of subjects is usually less than 50. Five data sets are used in this paper, including 98 subjects. Experimental results show that the proposed method achieves better classification results on each data set, proving the superiority of the proposed method.

The rest of this paper is organized as follows: Section II describes the data sets used in the experiment; Section III introduces the proposed method; Section IV presents the experimental results. The discussions and conclusions are contained in Section V and VI, respectively.

2. Materials

The relevant information of the five data sets is given in Table 1. For datasets 1 and 3, four classes of tasks are arranged and combined to form 6 groups of binary classification tasks. For specific processing methods, please refer to literatures [37,38]. For dataset 2, only the calibration data is used for classification [32]. For dataset 4, only the data of the third session are analyzed [16].

For dataset 5, since the sample size collected by each subject is different, we take percentages to divide the training set and test set. The scalp EEG signals are collected using a NuAmps amplifier from Neuroscan company. The experimental paradigm is similar to the BCI competition data, and the motor imagery time is 4 s. Two reference electrodes are placed under the left and right earlobes respectively.

3. Methods

The data processing flow of the new proposed algorithm framework is shown in Fig. 1, which consists of six steps. First, signal preprocessing. The original EEG signals are filtered by band-pass filtering with 8–30 Hz. Second, time window segmentation. Three time windows (i.e. tw1 = 0.5–2.5 s, tw2 = 1–3 s, and tw3 = 1.5–3.5 s) are used to extract single-trial data after the band-pass filtering. Third, spatial filtering. In this paper, CSP and TRCSP are used for spatial filtering. Fourth, feature extraction. The spatially filtered signals are filtered into multiple sub-bands using a filter bank with 8–12 Hz, 10–14 Hz, …, 26–30 Hz, and

Table 1
The relevant information of the five data sets.

Data sets	The number of subjects	The number of channels	The sampling rate	The number of tasks	The number of training and test sets for each task	Data sources
Dataset 1	3 (K3, K6, L1)	60	250 Hz	4 (left-hand, right-hand, foot, tongue)	Training set: 45 (K3), 30 (K6, L1); Test set: 45 (K3), 30 (K6, L1).	data set IIIa of BCI competition III (2005) [39]
Dataset 2	7 (a, b, …, g)	59	100 Hz	2 (left-hand, right-hand, and foot, two classes of which are selected)	Training set: 50; Test set: 50.	data set I of BCI competition IV (2008) [40]
Dataset 3	9 (A01, A02, …, A09)	22	250 Hz	4 (left-hand, right-hand, foot, tongue)	Training set: 72; Test set: 72.	data set IIa of BCI competition IV (2008) [41]
Dataset 4	9 (B01, B02, …, B09)	3	250 Hz	2 (left-hand, right-hand)	Training set: 40; Test set: 40.	data set IIb of BCI competition IV (2008) [42]
Dataset 5	10 (S01, S02, …, S10)	30	250 Hz	2 (left-hand, right-hand, foot, tongue)	Training set: 60%; Test set: 40%.	self-collected by our laboratory

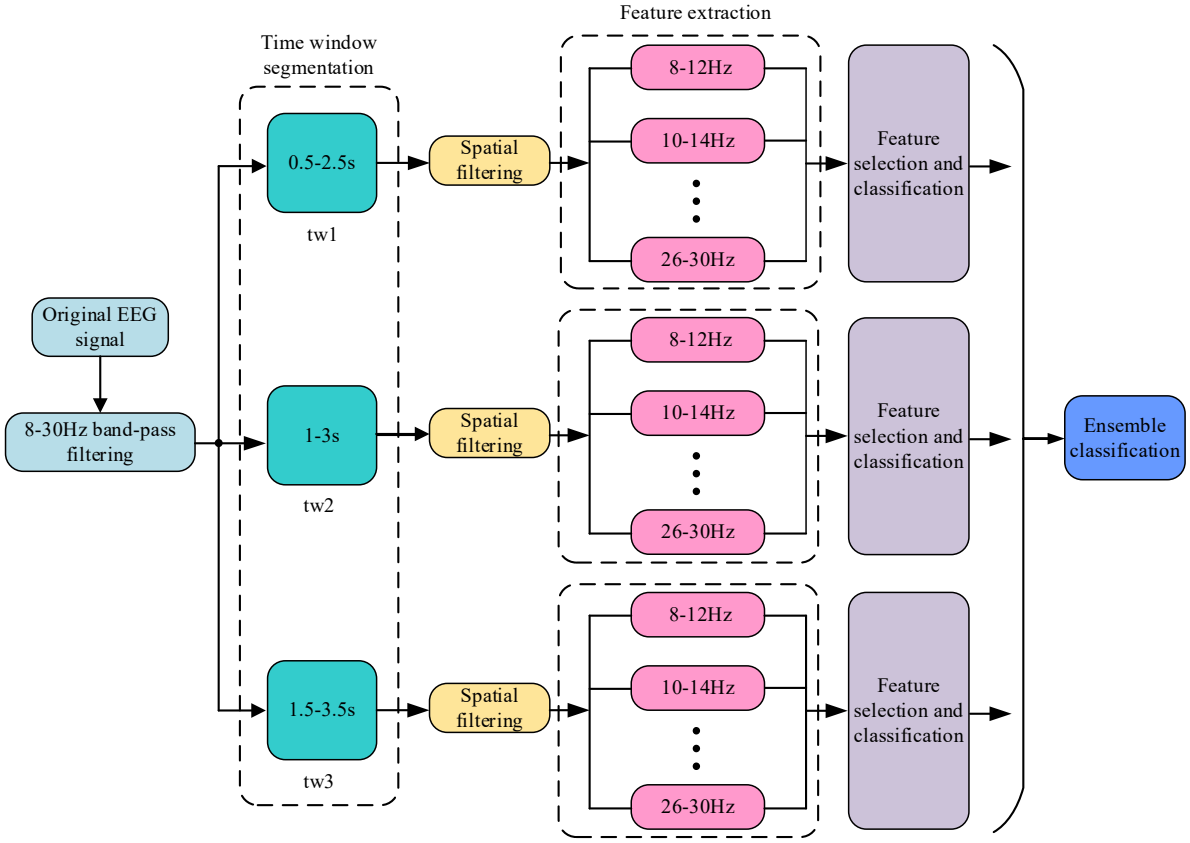

Fig. 1. The data processing flow of the proposed method.

then logarithmic variance features are extracted on each sub-band. Fifth, feature selection and classification. In order to reduce the data processing steps, we select a classification model that can simultaneously perform feature selection and classification. In this paper, the LASSO model is selected. Sixth, ensemble classification. In the training stage, the training set is used to train multiple base classification models; in the testing stage, the predicted values output by the multiple base classification models are integrated to obtain the final predicted value.

In the following, we will discuss the three core parts of the proposed method in more detail, namely spatial filtering, feature selection and classification, and ensemble classification.

3.1. Spatial filtering

3.1.1. CSP based on the sample covariance

The objective function of CSP is given as follows [7]:

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T \overline{\boldsymbol{C}}_1 \boldsymbol{w}}{\boldsymbol{w}^T \overline{\boldsymbol{C}}_2 \boldsymbol{w}} \tag{1}$$

where $\overline{\boldsymbol{C}}_k$ represents the average covariance matrix of k-th class of task. Since we only study the binary classification problem, hence $k \in \{1,2\}$. For the calculation of $\overline{\boldsymbol{C}}_k$, please refer to the literature [7].

The traditional CSP method is realized by diagonalizing the average covariance matrix of the two classes of tasks at the same time [7]. However, as the number of EEG channels increases, the sample covariance matrix has a square magnitude increase in its data dimension. When the sample size of the training set is small, the estimation of the sample covariance matrix is an ill-conditioned problem. In other words, the estimation is biased or inaccurate. If the signal has a low signal-to-noise ratio, the estimation deviation will be greater [5]. Therefore, the traditional CSP method is very sensitive to noise and prone to overfitting.

3.1.2. TRCSP

To solve the noise sensitivity problem of CSP and the overfitting problem caused by covariance estimation, researchers proposed a series of regularized CSP methods, of which TRCSP is one of the regularization methods. TRCSP [7] added Tikhonov regularization to the objective function of CSP, details as follows:

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T \overline{\boldsymbol{C}}_1 \boldsymbol{w}}{\boldsymbol{w}^T \overline{\boldsymbol{C}}_2 \boldsymbol{w} + \alpha P(\boldsymbol{w})} = \frac{\boldsymbol{w}^T \overline{\boldsymbol{C}}_1 \boldsymbol{w}}{\boldsymbol{w}^T \overline{\boldsymbol{C}}_2 \boldsymbol{w} + \alpha \boldsymbol{w}^T \boldsymbol{I} \boldsymbol{w}} = \frac{\boldsymbol{w}^T \overline{\boldsymbol{C}}_1 \boldsymbol{w}}{\boldsymbol{w}^T \overline{\boldsymbol{C}}_2 \boldsymbol{w} + \alpha \|\boldsymbol{w}\|_2^2} \tag{2}$$

where $\boldsymbol{I}$ is the identity matrix. It is essentially a quadratic l_2 norm constraint on the spatial filter vector $\boldsymbol{w}$ in the denominator. The advantage of TRCSP is that the new objective function is still a generalized eigenvalue problem, which can be solved by solving the eigenvalue problem. However, since the regular term is added, the numerator and denominator of the objective function have no symmetry property, and two eigenvalue solutions are needed to get the final spatial filter. Specifically, we need to solve the following two objective functions:

$$J_1(\boldsymbol{w}) = \frac{\boldsymbol{w}^T \overline{\boldsymbol{C}}_1 \boldsymbol{w}}{\boldsymbol{w}^T \overline{\boldsymbol{C}}_2 \boldsymbol{w} + \alpha \|\boldsymbol{w}\|_2^2} \tag{3}$$

$$J_2(\boldsymbol{w}) = \frac{\boldsymbol{w}^T \overline{\boldsymbol{C}}_2 \boldsymbol{w}}{\boldsymbol{w}^T \overline{\boldsymbol{C}}_1 \boldsymbol{w} + \alpha \|\boldsymbol{w}\|_2^2} \tag{4}$$

Maximizing $J_1(\boldsymbol{w})$ means maximizing the variance of the first class of tasks while minimizing the variance of the second class of tasks. Maximizing $J_2(\boldsymbol{w})$ means maximizing the variance of the second class of tasks while minimizing the variance of the first class of tasks. The first eigenvector matrix $\boldsymbol{M}_1 = (\overline{\boldsymbol{C}}_2 + \alpha \boldsymbol{I})^{-1} \overline{\boldsymbol{C}}_1$ can be obtained by maximizing $J_1(\boldsymbol{w})$, and the second eigenvector matrix $\boldsymbol{M}_2 = (\overline{\boldsymbol{C}}_1 + \alpha \boldsymbol{I})^{-1} \overline{\boldsymbol{C}}_2$ can be obtained by maximizing $J_2(\boldsymbol{w})$. The final spatial filter $\boldsymbol{W}$ is formed by taking the eigenvectors corresponding to the first m largest eigenvalues of $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$. The subsequent feature extraction method is the same as the CSP method [7].

3.1.3. Parameter setting of CSP and TRCSP

A new selection method for the pair number of spatial filters is proposed in this paper, instead of a fixed value, the value of m can be selected in the candidate subset. For CSP, the pair number of spatial filters is set as: $m \in \{1, \cdots, 10\}$ for dataset 1 and dataset 2, $m \in \{1, \cdots, 5\}$ for datasets 3 and 5, $m = 1$ for dataset 4. For details of the data sets, please refer to the data description section. For TRCSP, the pair number of spatial filters is set the same as that of the CSP method. The alternative set of regularization parameters is: $\alpha \in \{10^{-10}, 10^{-9}, \cdots, 10^{-1}\}$ [7].

Multiple feature subsets with diversity are generated by the combination of different time windows, regularization parameters, and pair numbers of spatial filters. These feature subsets will be used to train diversified base classification models. Fig. 2 shows the base classification model composition of the TRCSP method in the training stage. The total number of base models is the product of the number of time windows, the number of regularization parameters, and the pair number of spatial filters. For example, assuming that the regularization parameters and the pair number of spatial filters are selected from the set of $\alpha \in \{10^{-10}, 10^{-9}, \cdots, 10^{-1}\}$ and $m \in \{1, \cdots, 10\}$ respectively, then the number of regularization parameters is 10 and the pair number of spatial filters is 10. In addition, from the preprocessing section, we can know that the number of time windows is 3. Finally, the number of base models is $10 \times 10 \times 3 = 300$.

3.2. Feature selection and classification

After feature extraction, the feature sample matrix of the training set is denoted as $\boldsymbol{X}_{train} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)^T$, where $\boldsymbol{X}_{train} \in R^{N \times P}$, N is the total number of samples in the training set, P is the feature dimension of a sample, and $\boldsymbol{x}_i \in R^P$, $i \in (1, 2, \ldots, N)$ represents the i-th sample (feature vector). The sample label corresponding to the training set is denote as $\boldsymbol{y}_{train} = (y_1, y_2, \ldots, y_N)^T$, and $y_i \in \{-1, 1\}$. In order to select effective temporal-spatial-frequency features, the LASSO regression model is used to select and classify different feature subsets. The feature samples and labels of the training set are used to train the LASSO model as follows:

$$\min_{\boldsymbol{w}} \frac{1}{2} \|\boldsymbol{y}_{train} - \boldsymbol{X}_{train} \boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \tag{5}$$

where $\lambda > 0$ is the regularization parameter, $\boldsymbol{\beta}$ is the feature weight, $\|\boldsymbol{\beta}\|_1 = \sum_{i=1}^{P} |\boldsymbol{\beta}_i|$, and $|\boldsymbol{\beta}_i|$ is the absolute value of the i-th element of $\boldsymbol{\beta}$.

After solving formula (5), we get an optimal feature weight. For a new test sample $\boldsymbol{x}$, its predicted value is as follows:

$$y_{predict} = \boldsymbol{\beta}^T \boldsymbol{x} \tag{6}$$

Since we use the LASSO regression model, $y_{predict}$ is a regression value.

3.3. Ensemble classification

In the training stage, multiple base classification models are obtained. In the test stage, we select the base models with the top 80% of the cross-validation accuracy for integration. Assuming the number of model integrations is M, we use the summation rule to integrate the predicted values of the base models. The details are as follows:

$$y_{test} = \text{sign}\left(\sum_{i=1}^{M} y_{predict}(i)\right) \tag{7}$$

where $y_{predict}(i)$ represents the predicted value of the i-th model, sign$(\bullet)$ represents the symbolic function. The predicted values of all base models are summed first, and then the symbolic function is used to get the final predicted value. $y_{test} = 1$ indicates that this sample belongs to the first class of motor imagery; $y_{test} = -1$, indicates that this sample

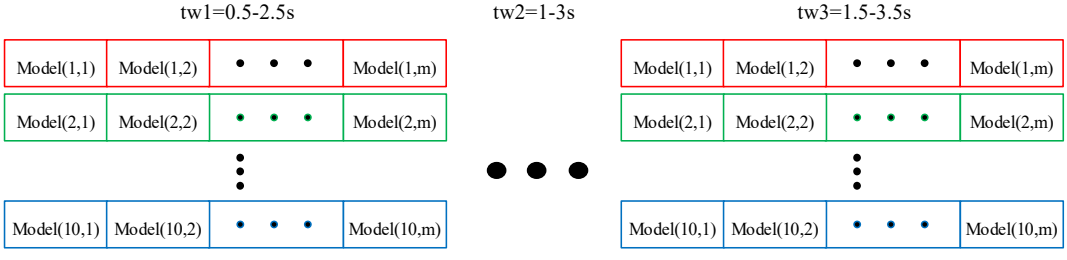

Fig. 2. The base classification model composition of TRCSP in the training stage.

belongs to the second class of motor imagery. We believe that the magnitude of the absolute value of $y_{predict}$ reflects the separability of features to some extent, and the classification results are relatively reliable. Therefore, the essence of formula (7) is to integrate the base classification models using the rule of weight sum.

The feature extraction method in this paper consists of three processes: time window (TW) segmentation, spatial filtering with CSP and TRCSP, and band-pass filtering with a filter bank (FB), from which we get two new feature extraction methods, we called TW-CSP-FB and TW-TRCSP-FB. In the feature classification stage, the LASSO model is used for feature selection and classification. Therefore, two EEG decoding methods are proposed by the combination of feature extraction and classification methods, that is TW-CSP-FB + LASSO and TW-TRCSP-FB + LASSO.

4. Experimental study

4.1. Comparison methods and model parameter selection

There are six methods to participate in the comparison, the first four methods are the existing methods, and the last two methods are the proposed methods. The existing methods are divided into 4 categories. First, the traditional CSP method. Second, the regularization improvement of CSP, including TRCSP. Third, the time–frequency optimization method of CSP, including STFSCSP [25]. Fourth, FB-TRCSP + RF [43], which combines regularization technique and frequency band optimization in the process of feature extraction, and random forest (RF) ensemble classifier is used for classification.

For all comparison methods, if SVM is used as a classifier, SVM is implemented by the LIBSVM toolbox [44]. In addition, the model parameters of SVM adopt the default settings of the toolbox, and the linear kernel is selected. For all methods, a 6-order Butterworth filter is used for band-pass filtering. For the pair number of spatial filter, unless otherwise specified, $m = 1$ for dataset 4 (Dataset 4 has only 3 channels, so m can only take the value of 1), $m = 3$ for other datasets. Unless otherwise stated, the time window for dataset 2 is 0.5–3.5 s and for other datasets is 0.5–2.5 s.

CSP: The 8–30 Hz band-pass filtering is performed on the original EEG signals, and then the single-trial data of the corresponding time window is extracted. CSP is used to extract features [7], and SVM is used for classification.

TRCSP: The 8–30 Hz band-pass filtering is performed on the original EEG signals, and then the single-trial data of the corresponding time window is extracted. TRCSP is used to extract features. Combined with SVM, the optimal regularization parameter α is selected by using 10-fold cross-validation. The candidate set is $\alpha \in \{10^{-10}, 10^{-9}, \cdots, 10^{-1}\}$, which is consistent with that of the literature [7].

FB-TRCSP + RF: The original EEG signals are filtered into 9 sub-bands using a filter bank with 4–8 Hz, 8–12 Hz, …, 36–40 Hz, and then the single-trial data of the corresponding time window is extracted. Then TRCSP feature extraction is performed on each sub-band, and finally, RF is used for classification [43].

STFSCSP: The original EEG signals are filtered into 17 sub-bands using a filter bank with 4–8 Hz, 6–10 Hz, …, 36–40 Hz. For each sub-band, multiple time windows are selected to extract the single-trial data. The time windows for dataset 1 are 0.5–2.5 s, 1–3 s, and 1.5–3.5 s, and the time windows for other datasets are 0.5–2.5 s, 1–3 s, 1.5–3.5 s, and 2–4 s. CSP feature extraction is performed on each time–frequency segmentation, and two optimal spatial filter vectors are selected as spatial filters using the Fisher score. LASSO and Fisher's linear discriminant analysis are combined to select the optimal temporal-spatial-frequency features. Finally, weighted naive Bayes is used for classification [25].

TW-CSP-FB + LASSO: TW-CSP-FB is used to extract features. Feature selection and classification are performed simultaneously using LASSO.

TW-TRCSP-FB + LASSO: TW-TRCSP-FB is used to extract features. Feature selection and classification are performed simultaneously using LASSO.

Because the name of the proposed method is too long, in the following experimental results and discussion sections, TW-CSP-FB + LASSO and TW-TRCSP-FB + LASSO are referred to as TW-CSP-FB and TW-TRCSP-FB respectively. LASSO is implemented by the SLEP toolbox [45]. The 10-fold cross-validation and grid search methods are used to select the optimal regularization parameters of the LASSO model. The alternative hyperparameter for the LASSO model is $\lambda \in [2^{-5}, 2^{-4.8}, \ldots, 2^{4.8}, 2^{5}]$.

4.2. Experimental results

Table 2 lists the average classification accuracy of the five data sets, the highest classification accuracy is shown in bold. The last row of Table 2 is the average classification accuracy and standard deviation of all data sets. As can be seen from Table 2, TW-CSP-FB and TW-TRCSP-FB are significantly superior to other methods. TW-TRCSP-FB achieves the highest average classification accuracy for all data sets. Except for datasets 3 and 5, the classification results of TW-CSP-FB are better than other existing methods. The experimental results of a large amount of data show that the proposed method has good universality and robustness.

Except for FB-TRCSP + RF, other methods are better than CSP. FB-TRCSP + RF is a combination of FBCSP and TRCSP, which has the advantages of both, but the experimental results are not ideal. The reason may be that the model parameters of TRCSP and RF in FB-TRCSP + RF are selected separately and the same training set is used multiple times, which may lead to overfitting. Among the existing methods, STFSCSP achieves the best classification effect, indicating that the temporal-spatial-frequency joint optimization is very effective and important.

Fig. 3 shows the classification accuracy of all subjects in the five data sets. The value of the ordinate represents the classification accuracy of the proposed method, and the value of the abscissa represents the classification accuracy of the compared method. The points above the diagonal indicate that the proposed method is superior to the compared methods. As can be seen from Fig. 3, most of the points are located above the diagonal, which demonstrates the superiority of the proposed method.

In order to more comprehensively compare the performance of the proposed methods, the classification results of datasets 1 to 4 in recent existing literatures are presented in Table 3. It should be noted that, for dataset 1 and dataset 2, only the results of the binary classification task of left-hand versus right-hand (L vs R) are compared. Since it is impossible to replicate the algorithms of all references, we directly compare the average classification accuracy obtained from these references. As can be seen from Table 3, TW-TRCSP-FB achieved the best classification results in all data sets. Although the classification result of TW-CSP-FB

Table 2
The average classification accuracy of various methods (all data sets).

Datasets	CSP	TRCSP	FB-TRCSP +RF	STFSCSP	TW-CSP-FB	TW-TRCSP -FB
Dataset 1	88.46	89.63	86.48	83.92	92.56	**92.87**
Dataset 2	82.86	84.57	79.29	76.71	88	**90.57**
Dataset 3	80.31	80.54	81.84	84.4	84.36	**85.02**
Dataset 4	78.19	78.33	74.86	80.56	81.53	**81.81**
Dataset 5	67.27	65.41	66.2	78.65	76.6	**79.4**
Mean ± std	80.46 ±14.32	80.75 ±14.17	80.27 ±12.70	82.82 ±13.32	85.08 ±12.51	**85.99 ±12.07**

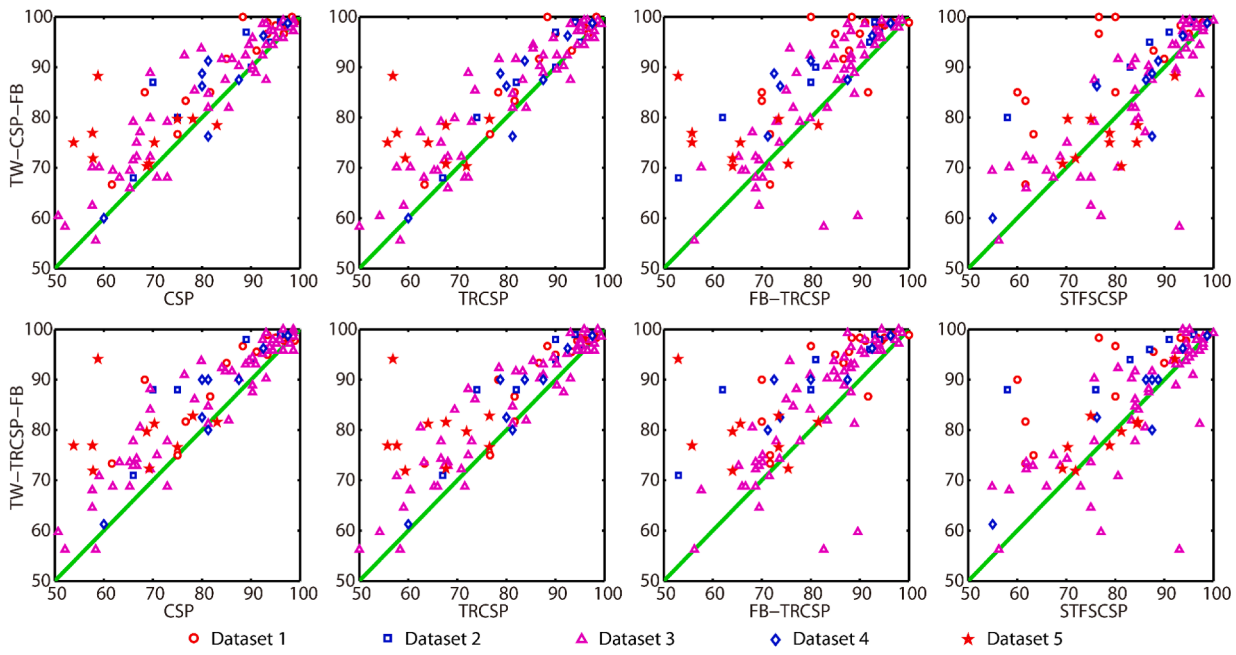

Fig. 3. The classification accuracy of the proposed methods and the compared methods (all data sets).

Table 3
Comparison of the classification accuracy between the proposed methods and other existing methods (BCI competition data sets).

Methods	Dataset 1 (L vs R)	Methods	Dataset 2	Methods	Dataset 3 (L vs R)	Methods	Dataset 4
WOLA-CSP [46] (2018)	84.25± 16.08	PELM [49] (2018)	70.00 ± 10.33	GRU-RNN [54] (2018)	73.56 ± 4.38	DBN [58] (2018)	70.72± 2.65
DASS [47] (2019)	85± 16.72	LRFCSP [50] (2019)	84.70 ± 11.22	ERDSA [46] (2018)	78.86 ± 15.07	CapsNet [59] (2019)	78.44± 14.44
CSP\AM-BA-SVM [48] (2018)	86.57± 9.03	OPTICAL [51] (2019)	82.53 ± 8.17	IST-TSVM [55] (2019)	70.22 ± 19.74	SGRM[60] (2019)	78.0± 2.3
L1-CSP [13] (2020)	89.82± 10.49	BF [52] (2020)	89.50 ± 2.12	ICA + PSR + CSP [56] (2020)	74.39 ± 15.18	NCFS [61] (2020)	81.52± 13.72
TSGSP [26] (2018)	87.63± 14.49	SCSP-RDA [53] (2020)	87.21 ± 11.26	p-LTCSP [57] (2020)	69.56 ± 11.10	SBLCSP [16] (2017)	81.7± 15.1
TW-CSP-FB	88.52± 15.46	TW-CSP-FB	88 ± 10.14	TW-CSP-FB	78.32 ± 14.97	TW-CSP-FB	81.53± 15.93
TW-TRCSP-FB	**90.18± 11.92**	TW-TRCSP-FB	**90.57 ± 8.97**	TW-TRCSP-FB	**79.48 ± 15.39**	TW-TRCSP-FB	**81.81± 15.99**

has no obvious advantages, it is better than most existing methods.

Furthermore, we also investigate the influence of some key parameters on the proposed method. First, the influence of different frequency band divisions or different number of frequency bands on the proposed method is investigated. On the one hand, we subdivide the frequency bandwidth (FBW) into 2 Hz, specifically: 8–10 Hz, 9–11 Hz, …, 28–30 Hz (FBW = 2 Hz). On the other hand, we expand the frequency bandwidth to 6 Hz, specifically: 8–14 Hz, 12–18 Hz, …, 24–30 Hz (FBW = 6 Hz). Second, the influence of the regularization parameters of LASSO model on the proposed method is investigated. The exponential interval (EI) between the values of the current regularization parameter set ($\lambda \in [2^{-5}, 2^{-4.8}, \ldots, 2^{4.8}, 2^{5}]$) is relatively small, the amount of parameters is relatively large, and it is time-consuming to traverse all parameters to determine the optimal LASSO model. We set the exponent interval to be 0.5 and 1, specifically: $\lambda \in [2^{-5}, 2^{-4.5}, \ldots, 2^{4.5}, 2^{5}]$ (EI = 0.5) and $\lambda \in [2^{-5}, 2^{-4}, \ldots, 2^{4}, 2^{5}]$ (EI = 1). Experimental validation is performed using dataset 3 (Data set IIa of BCI competition IV, left-hand vs right-hand tasks), and the experimental results are given in Table 4. In Table 4, the parameters in the sixth column are that used in the previous results, and we use the results of these parameters as a comparison benchmark. From the classification results in Table 4, it can be seen that the subdivision or expansion of the frequency band will lead to a decrease in the average classification accuracy, but the difference is not very large. In addition, a smaller regularization parameter interval can find more accurate model parameters and thus train a better LASSO model, so the classification results are better.

Table 4
The classification accuracy of dataset 3 (left-hand vs right-hand tasks) with some key parameters of the proposed method.

Subjects	FBW = 2, EI = 0.2	FBW = 6, EI = 0.2	FBW = 4, EI = 0.5	FBW = 4, EI = 1.0	FBW = 4, EI = 0.2
A01	89.58	**92.36**	91.67	90.97	91.67
A02	**61.11**	56.94	54.86	55.56	56.25
A03	**95.83**	95.14	**95.83**	95.14	**95.83**
A04	**69.44**	68.75	68.75	68.75	68.75
A05	**60.42**	57.64	56.25	56.25	56.25
A06	68.75	75	74.31	73.61	**75**
A07	79.86	79.17	**81.25**	**81.25**	**81.25**
A08	96.53	95.83	96.53	**97.22**	**97.22**
A09	91.67	91.67	**93.06**	92.36	**93.06**
Mean ± std	79.24 ±13.84	79.17 ±14.71	79.17 ±15.57	79.01 ±15.35	**79.48** ±15.39

Finally, we analyze the classification performance of four methods (CSP, TRCSP, TW-CSP-FB, and TW-TRCSP-FB) under different numbers of training samples. Dataset 3 (Data set IIa of BCI competition IV, left-hand vs right-hand tasks) is still used for experimental validation. In Fig. 4, the abscissa represents the number of training samples per class, and the ordinate represents the average classification accuracy of dataset 3 on different numbers of training sets. Only the number of samples in the training set is changed, and the test set remains the same. It can be seen from Fig. 4 that the classification effect of TW-TRCSP-FB is better than other methods whether in the case of small samples or in the process of increasing samples.

5. Discussion

By comparing the classification results of TW-TRCSP-FB and TW-CSP-FB, we can conclude that the regularization technique combined with temporal-spatial-frequency joint optimization can effectively improve the decoding performance of CSP.

Moreover, the better classification results achieved by the proposed method depend to a certain extent on the construction of the ensemble model. To demonstrate this conclusion, Table 5 presents the classification results of the cross-validation model and the ensemble model. In Table 5, the TW-CSP-FB_CV (TW-TRCSP-FB_CV) method consists of two parts, including feature extraction using TW-CSP-FB (TW-TRCSP-FB) and the selection of the optimal LASSO model using 10-fold cross-validation. As can be seen from Table 5, the ensemble model significantly outperforms the cross-validation model. Compared with the optimal model selected by cross-validation, the ensemble model can make full use of the temporal-spatial-frequency information of EEG and avoid the loss of useful information, which may be the reason why the proposed method achieves better classification results.

In order to illustrate the importance of making full use of the temporal-spatial-frequency information, we compare the classification results that integrate information from different time windows. Figs. 5 and 6 respectively show the average classification accuracy of TW-TRCSP-FB and TW-CSP-FB in a single time window and in all time windows. The classification results of TW-TRCSP-FB and TW-CSP-FB methods are consistent, that is, the single time window obtaining the optimal average classification accuracy in each data set is different, but the classification effect integrated information from all time window is superior to that of single time window. These experimental results show that making full use of information in different time windows can improve classification performance, and it may lose useful information if only a single time window is used. Similarly, different time windows,

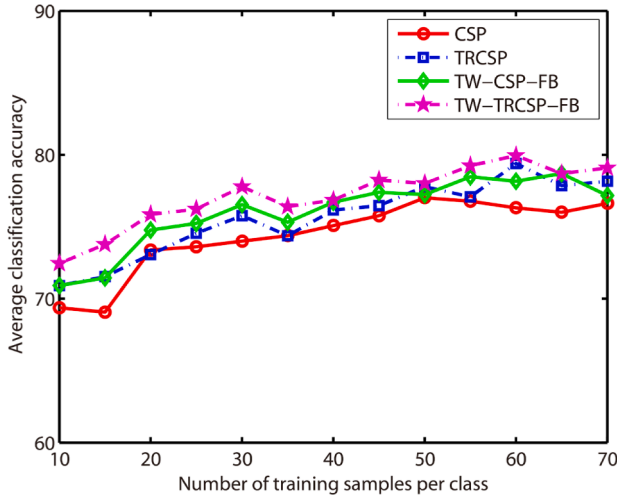

Fig. 4. The average classification accuracy of dataset 3 (left-hand vs right-hand tasks) with different numbers of training samples.

Table 5
Comparison of the classification accuracy between the cross-validation model and the ensemble model.

Datasets	TW-CSP-FB_CV	TW-CSP-FB	TW-TRCSP-FB_CV	TW-TRCSP-FB
Dataset 1	88.58	92.56	88.12	**92.87**
Dataset 2	78.71	88.00	77.00	**90.57**
Dataset 3	81.12	84.36	80.75	**85.02**
Dataset 4	79.03	81.53	81.11	**81.81**
Dataset 5	73.35	76.60	73.38	**79.4**
All data	81.33 ±13.88	85.08 ±12.51	81.12 ±13.58	**85.99 ±12.07**

frequency bands, spatial filters, and other information have been integrated under the proposed algorithm framework. Therefore, the proposed method achieves better classification results.

From the above analysis, we can draw the conclusion that incorporating the CSP regularization technique and temporal-spatial-frequency joint optimization into the ensemble learning framework can improve the decoding performance of CSP. In the following, we further analyze the amount of calculation, model complexity, and robustness of the proposed method.

Compared with the existing temporal-spatial-frequency joint optimization methods, the proposed method has less amount of calculation and lower model complexity. In terms of the amount of calculation, the feature extraction in STFSCSP [25] and TSGSP [26] methods need to filter the original EEG signals with all channels into multiple time–frequency segmentations. Especially TSGSP, with 6 time windows, each with 17 frequency bands, a total of 112 time–frequency segmentations are needed to process, and the amount of calculation is very large. However, in the proposed temporal-spatial-frequency feature extraction method, the TRCSP (or CSP) spatial projection is used to reduce the dimension of the original EEG signals, and then the filter bank is used for band-pass filtering. After spatial projection, the EEG channels are greatly reduced, only 20 at most, which significantly reduces the amount of calculation of band-pass filtering. Furthermore, the proposed method only needs to process 3 time windows, each with 10 frequency bands. In terms of model complexity, the feature selection method in TSGSP is a combination model of sparse group LASSO and fused LASSO, three model parameters need to be determined. Therefore, the TSGSP model has high complexity and long model training time. However, the simple LASSO model is used in the proposed method, which has low model complexity and short model training time.

Furthermore, compared with the existing ensemble learning methods, the ensemble model proposed in this paper has a small ensemble scale. For example, FB-TRCSP + RF [43] uses RF as the classifier, and the number of base classifiers is selected as 1000. The maximum number of base classifier (LASSO model) in our method is only 300, and only 80% of them are selected.

We discuss the robustness of the proposed method from three aspects. First, at the feature extraction level. In TRCSP method, the Tikhonov regularization is used to constrain the spatial filter $\boldsymbol{w}$, which penalize large weights of the spatial filter $\boldsymbol{w}$ in CSP, so that the value of the spatial filter $\boldsymbol{w}$ becomes smooth [7]. Excessive weights in the spatial filter $\boldsymbol{w}$ may be caused by amplified noise, and Tikhonov regularization penalizes it to reduce the impact of noisy data. Thus, compared to CSP, TRCSP has better robustness to the artifacts and outliers of EEG data. In addition, Tikhonov Regularization can reduce the complexity of the model and shrink the solution space [62], so that TRCSP can alleviate the overfitting problem to a certain extent. Second, at the feature selection level. The proposed method performs spatial-frequency feature selection for each time window, and the dimension of spatial-frequency features is significantly less than existing methods (such as STFSCSP and TSGSP). In addition, a simple LASSO model is used for feature selection. The small dimension of the sample feature and low model complexity

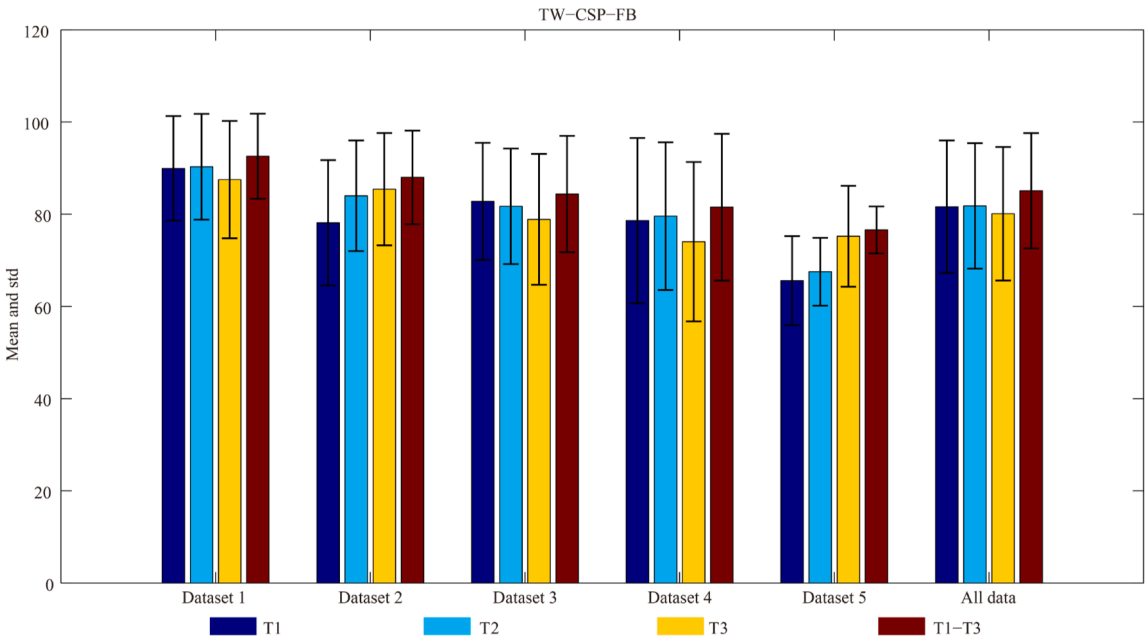

Fig. 5. The average classification accuracy of TW-CSP-FB integrated with different time windows.

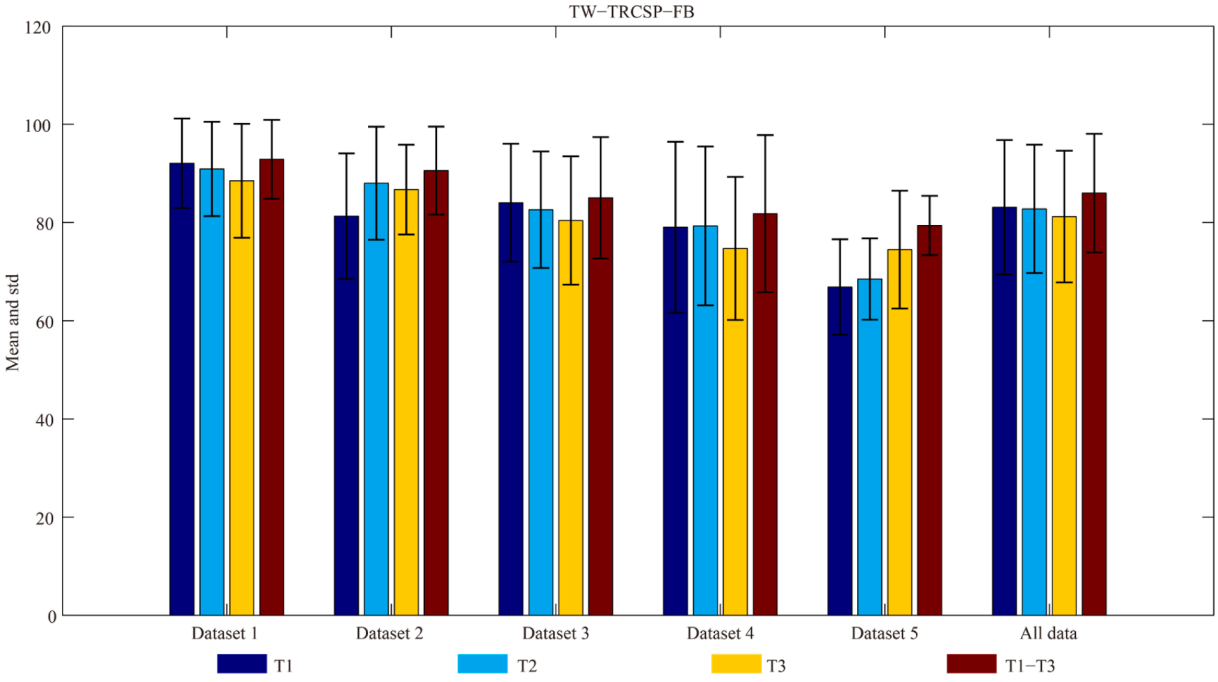

Fig. 6. The average classification accuracy of TW-TRCSP-FB integrated with different time windows.

make the overfitting risk of the proposed method significantly reduced under the same number of samples. Finally, at the model integration level. As mentioned above, we choose the base classification model with higher classification accuracy for ensemble, so the stability of the ensemble model will be better.

Although the proposed algorithm framework has achieved better classification performance, it also has certain limitations. First, the proposed ensemble model only considers the diversity of features and does not consider the diversity of base classifiers. In future research, we will select different types of classifiers for integration, so as to further improve the performance of EEG decoding. Second, we did not consider the phase information in the overall optimization of CSP, and the phase information also has a relatively large impact on the performance of CSP [63]. In addition to phase information optimization, the multi-class extension of the proposed method is also an important part of our future work. In addition, the parallel processor such as field programmable gate array (FPGA) is considered to process the training and integration of multiple base classification models in parallel, so as to further improve the timeliness of the algorithm.

6. Conclusion

In this paper, a new algorithm framework based on ensemble learning is proposed to improve the decoding performance of CSP, in

which the overall optimization of CSP is considered. TRCSP with different time windows, regularization parameters, and pair numbers of spatial filters is used to solve the noise sensitivity and overfitting problems of CSP and extract richer temporal-spatial-frequency feature information. LASSO as the base classification model is used to select and optimize temporal-spatial-frequency features. Under the new ensemble learning framework, feature extraction and selection are effectively integrated, in which the diversity and high accuracy of the base classification model are considered at the same time. In addition, a new simple and effective integration rule is proposed for ensemble model construction. The temporal-spatial-frequency information of the motor imagery EEG is fully excavated and utilized in the ensemble model, which effectively improves the decoding performance of CSP. The experimental results show that the proposed method not only has a better classification effect but also has less amount of calculation, low model complexity, and high robustness.

CRediT authorship contribution statement

Shaorong Zhang: Conceptualization, Methodology, Software, Writing – original draft. **Zhibin Zhu:** Conceptualization, Methodology, Software, Writing – original draft. **Benxin Zhang:** Methodology, Software. **Bao Feng:** Data curation, Writing – original draft. **Tianyou Yu:** Visualization, Investigation. **Zhi Li:** Visualization, Investigation. **Zhiguo Zhang:** Supervision, Project administration. **Gan Huang:** Writing – review & editing. **Zhen Liang:** 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.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Nos. 61967004, 11901137, and 81960324), Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (Nos. YQ20113 and YQ22209).

References

[1] N. Padfield, J. Zabalza, H. Zhao, V. Masero, J. Ren, EEG-based brain-computer interfaces using motor-imagery: techniques and challenges, Sensors 19 (6) (2019) 1423.
[2] 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.
[3] A. Jiang, J. Shang, X. Liu, Y. Tang, H.K. Kwan, Y. Zhu, Efficient CSP algorithm with spatio-temporal filtering for motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (4) (2020) 1006–1016.
[4] Y. Miao, J. Jin, I. Daly, C. Zuo, X. Wang, A. Cichocki, T.-P. Jung, Learning common time-frequency-spatial patterns for motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 699–707.
[5] Haiping Lu, How-Lung Eng, Cuntai Guan, K.N. Plataniotis, A.N. Venetsanopoulos, Regularized common spatial pattern with aggregation for EEG classification in small-sample setting, IEEE Trans. Biomed. Eng. 57 (12) (2010) 2936–2946.
[6] N. Singh Malan, S. Sharma, Time window and frequency band optimization using regularized neighbourhood component analysis for Multi-View Motor Imagery EEG classification, Biomed. Signal Process. Control 67 (2021) 102550.
[7] F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms, IEEE Trans. Biomed. Eng. 58 (2) (2010) 355–362.
[8] H. Wang, Q. Tang, W. Zheng, L1-norm-based common spatial patterns, IEEE Trans. Biomed. Eng. 59 (3) (2011) 653–662.
[9] P. Li, P. Xu, R. Zhang, L. Guo, D. Yao, L1 norm based common spatial patterns decomposition for scalp EEG BCI, Biomed. Eng. Online 12 (1) (2013).
[10] I. Onaran, N.F. Ince, A.E. Cetin, Sparse spatial filter via a novel objective function minimization with smooth ℓ1 regularization, Biomed. Signal Process. Control 8 (3) (2013) 282–288.
[11] X. Li, X. Lu, H. Wang, Robust common spatial patterns with sparsity, Biomed. Signal Process. Control 26 (2016) 52–57.
[12] F. Qi, W. Wu, Z.L. Yu, Z. Gu, Z. Wen, T. Yu, Y. Li, Spatiotemporal-filtering-based channel selection for single-trial EEG classification, IEEE Trans. Cybern. 51 (2) (2021) 558–567.
[13] B. Wang, C.M. Wong, Z. Kang, F. Liu, C. Shui, F. Wan, C.L.P. Chen, Common spatial pattern reformulated for regularizations in brain–computer interfaces, IEEE Trans. Cybern. 51 (10) (2021) 5008–5020.
[14] K.K. Ang, Z.Y. Chin, C. Wang, C. Guan, H. Zhang, Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b, Front. Neurosci. 6 (2012).
[15] Y.u. Zhang, G. Zhou, J. Jin, X. Wang, A. Cichocki, Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface, J. Neurosci. Methods 255 (2015) 85–91.
[16] Y.u. Zhang, Y.u. Wang, J. Jin, X. Wang, Sparse Bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification, Int. J. Neural Syst. 27 (02) (2017) 1650032.
[17] S. Kumar, A. Sharma, A new parameter tuning approach for enhanced motor imagery EEG signal classification, Med. Biol. Eng. Comput. 56 (10) (2018) 1861–1874.
[18] H. Ghaheri, A.R. Ahmadyfard, Extracting common spatial patterns from EEG time segments for classifying motor imagery classes in a Brain Computer Interface (BCI), Sci Iran 20 (2013) 2061–2072.
[19] O. Aydemir, Common spatial pattern-based feature extraction from the best time segment of BCI data, Turk. J. Electr. Eng. Comput. Sci. 24 (2016) 3976–3986.
[20] J. Wang, Z. Feng, N.a. Lu, J. Luo, Toward optimal feature and time segment selection by divergence method for EEG signals classification, Comput. Biol. Med. 97 (2018) 161–170.
[21] J. Feng, E. Yin, J. Jin, R. Saab, I. Daly, X. Wang, D. Hu, A. Cichocki, Towards correlation-based time window selection method for motor imagery BCIs, Neural Networks 102 (2018) 87–95.
[22] J. Wang, Z. Feng, X. Ren, N.a. Lu, J. Luo, L. Sun, Feature subset and time segment selection for the classification of EEG data based motor imagery, Biomed. Signal Process. Control 61 (2020) 102026.
[23] P. Xu, T. Liu, R. Zhang, Y. Zhang, D. Yao, Using particle swarm to select frequency band and time interval for feature extraction of EEG based BCI, Biomed. Signal Process. Control 10 (2014) 289–295.
[24] Y. Yang, S. Chevallier, J. Wiart, I. Bloch, Subject-specific time-frequency selection for multi-class motor imagery-based BCIs using few Laplacian EEG channels, Biomed. Signal Process. Control 38 (2017) 302–311.
[25] M. Miao, H. Zeng, A. Wang, C. Zhao, F. Liu, Discriminative spatial-frequency-temporal feature extraction and classification of motor imagery EEG: An sparse regression and Weighted Naïve Bayesian Classifier-based approach, J. Neurosci. Methods 278 (2017) 13–24.
[26] Y.u. Zhang, C.S. Nam, G. Zhou, J. Jin, X. Wang, A. Cichocki, Temporally constrained sparse group spatial patterns for motor imagery BCI, IEEE Trans. Cybern. 49 (9) (2019) 3322–3332.
[27] L.i. Wang, W. Huang, Z. Yang, C. Zhang, Temporal-spatial-frequency depth extraction of brain-computer interface based on mental tasks, Biomed. Signal Process. Control 58 (2020) 101845.
[28] 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 (1) (2019) 016019.
[29] V. Mishuhina, X. Jiang, Feature weighting and regularization of common spatial patterns in EEG-based motor imagery BCI, IEEE Signal Process. Lett. 25 (6) (2018) 783–787.
[30] Caramia N, Lotte F, Ramat S. Optimizing spatial filter pairs for EEG classification based on phase-synchronization[C]. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014: 2049-2053.
[31] Yang Y, Chevallier S, Wiart J, et al. Automatic selection of the number of spatial filters for motor-imagery BCI[C]. The proceeding of 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2012).
[32] C. Zuo, Y. Miao, X. Wang, L. Wu, J. Jin, Temporal frequency joint sparse optimization and fuzzy fusion for motor imagery-based brain-computer interfaces, J. Neurosci. Methods 340 (2020) 108725.
[33] S.H. Park, S.G. Lee, Small sample setting and frequency band selection problem solving using sub-band regularized common spatial pattern, IEEE Sens. J. 17 (10) (2017) 2977–2983.
[34] S.H. Park, D. Lee, S.G. Lee, Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (2) (2018) 498–505.
[35] S.-L. Wu, Y.-T. Liu, T.-Y. Hsieh, Y.-Y. Lin, C.-Y. Chen, C.-H. Chuang, C.-T. Lin, Fuzzy integral with particle swarm optimization for a motor-imagery-based brain–computer interface, IEEE Trans. Fuzzy Syst. 25 (1) (2017) 21–28.
[36] H. Raza, D. Rathee, S.-M. Zhou, H. Cecotti, G. Prasad, Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related EEG-based brain-computer interface, Neurocomputing 343 (2019) 154–166.
[37] F. Qi, Y. Li, W. Wu, RSTFC: a novel algorithm for spatio-temporal filtering and classification of single-trial EEG, IEEE Trans. Neural Networks Learn. Syst. 26 (12) (2015) 3070–3082.
[38] A.K. Das, S. Suresh, N. Sundararajan, A discriminative subject-specific spatio-spectral filter selection approach for EEG based motor-imagery task classification, Expert Syst. Appl. 64 (2016) 375–384.
[39] H. Wang, C. Tang, T. Xu, T. Li, L. Xu, H. Yue, P. Chen, J. Li, A. Bezerianos, An approach of one-vs-rest filter bank common spatial pattern and spiking neural

networks for multiple motor imagery decoding, IEEE Access 8 (2020) 86850–86861.
[40] J. Jin, C. Liu, I. Daly, Y. Miao, S. Li, X. Wang, A. Cichocki, Bispectrum-based channel selection for motor imagery based brain-computer interfacing, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (10) (2020) 2153–2163.
[41] C. Zuo, J. Jin, R. Xu, L. Wu, C. Liu, Y. Miao, X. Wang, Cluster decomposing and multi-objective optimization based-ensemble learning framework for motor imagery-based brain–computer interfaces, J. Neural Eng. 18 (2) (2021) 026018.
[42] J. Luo, X. Gao, X. Zhu, B. Wang, N.a. Lu, J. Wang, Motor imagery EEG classification based on ensemble support vector learning, Comput. Methods Programs Biomed. 193 (2020) 105464.
[43] R. Zhang, X. Xiao, Z. Liu, W. Jiang, J. Li, Y. Cao, J. Ren, D. Jiang, L. Cui, A new motor imagery EEG classification method FB-TRCSP+RF based on CSP and random forest, IEEE Access. 6 (2018) 44944–44950.
[44] C.C. Chang, C.J. Lin, LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol. (TIST) 2 (3) (2011) 1–27.
[45] Liu J, Ji S, Ye J. SLEP: Sparse learning with efficient projections[J]. Arizona State University, 2009, 6(491): 7.
[46] K. Belwafi, O. Romain, S. Gannouni, F. Ghaffari, R. Djemal, B. Ouni, An embedded implementation based on adaptive filter bank for brain–computer interface systems, J. Neurosci. Methods 305 (2018) 1–16.
[47] K. Belwafi, S. Gannouni, H. Aboalsamh, H. Mathkour, A. Belghith, A dynamic and self-adaptive classification algorithm for motor imagery EEG signals, J. Neurosci. Methods 327 (2019) 108346.
[48] S. Selim, M.M. Tantawi, H.A. Shedeed, A. Badr, A CSP\AM-BA-SVM approach for motor imagery BCI system, IEEE Access 6 (2018) 49192–49208.
[49] Y. Dai, X. Zhang, Z. Chen, X. Xu, Classification of electroencephalogram signals using wavelet-CSP and projection extreme learning machine, Rev. Sci. Instrum. 89 (7) (2018) 074302.
[50] Y. Park, W. Chung, Frequency-optimized local region common spatial pattern approach for motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (7) (2019) 1378–1388.
[51] S. Kumar, A. Sharma, T. Tsunoda, Brain wave classification using long short-term memory network based OPTICAL predictor, Sci. Rep. 9 (1) (2019) 1–13.
[52] Z. Luo, X. Lu, X. Xi, EEG feature extraction based on a bilevel network: minimum spanning tree and regional network, Electronics 9 (2) (2020) 203.
[53] R. Fu, M. Han, Y. Tian, P. Shi, Improvement motor imagery EEG classification based on sparse common spatial pattern and regularized discriminant analysis, J. Neurosci. Methods 343 (2020) 108833.
[54] E.C. Djamal, R.D. Putra, Brain-computer interface of focus and motor imagery using wavelet and recurrent neural networks, TELKOMNIKA Telecommun. Comput. Electron. Control 18 (4) (2020) 2748–2756.
[55] Y. Xu, J. Hua, H. Zhang, R. Hu, X. Huang, J. Liu, F. Guo, Improved transductive support vector machine for a small labelled set in motor imagery-based brain-computer interface, Comput. Intell. Neurosci. 2019 (2019) 1–16.
[56] E. Dong, K. Zhou, J. Tong, S. Du, A novel hybrid kernel function relevance vector machine for multi-task motor imagery EEG classification, Biomed. Signal Process. Control 60 (2020) 101991.
[57] Z. Yu, T. Ma, N.a. Fang, H. Wang, Z. Li, H. Fan, Local temporal common spatial patterns modulated with phase locking value, Biomed. Signal Process. Control 59 (2020) 101882.
[58] Y. Chu, X. Zhao, Y. Zou, W. Xu, J. Han, Y. Zhao, A decoding scheme for incomplete motor imagery EEG with deep belief network, Front. Neurosci. 12 (2018).
[59] K.W. Ha, J.W. Jeong, Motor imagery EEG classification using capsule networks, Sensors 19 (13) (2019) 2854.
[60] Y. Jiao, Y.u. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, A. Cichocki, Sparse group representation model for motor imagery EEG classification, IEEE J. Biomed. Health. Inf. 23 (2) (2019) 631–641.
[61] M.K.I. Molla, A.A. Shiam, M.R. Islam, T. Tanaka, Discriminative feature selection-based motor imagery classification using EEG signal, IEEE Access. 8 (2020) 98255–98265.
[62] D. Calvetti, E. Somersalo, Inverse problems: from regularization to Bayesian inference, Wiley Interdiscip. Rev. Comput. Stat. 10 (3) (2018) e1427.
[63] Kumar S, Reddy T, Behera L. EEG based motor imagery classification using instantaneous phase difference sequence[C]. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2018: 499–504.