Feature Extraction Method Based on Filter Banks and Riemannian Tangent Space in Motor-Imagery BCI

Hua Fang, Jing Jin, Ian Daly and Xingyu Wang

***Abstract*—Optimal feature extraction for multi-category motor imagery brain-computer interfaces (MI-BCIs) is a research hotspot. The common spatial pattern (CSP) algorithm is one of the most widely used methods in MI-BCIs. However, its performance is adversely affected by variance in the operational frequency band and noise interference. Furthermore, the performance of CSP is not satisfactory when addressing multi-category classification problems. In this work, we propose a fusion method combining Filter Banks and Riemannian Tangent Space (FBRTS) in multiple time windows. FBRTS uses multiple filter banks to overcome the problem of variance in the operational frequency band. It also applies the Riemannian method to the covariance matrix extracted by the spatial filter to obtain more robust features in order to overcome the problem of noise interference. In addition, we use a One-Versus-Rest support vector machine (OVR-SVM) model to classify multi-category features. We evaluate our FBRTS method using BCI competition IV dataset 2a and 2b. The experimental results show that the average classification accuracy of our FBRTS method is 77.7% and 86.9% in datasets 2a and 2b respectively. By analyzing the influence of the different numbers of filter banks and time windows on the performance of our FBRTS method, we can identify the optimal number of filter banks and time windows. Additionally, our FBRTS method can obtain more distinctive features than the filter banks common spatial pattern (FBCSP) method in two-dimensional embedding space. These results show that our proposed method can improve the performance of MI-BCIs.**

***Index Terms*—Motor imagery, brain-computer interfaces (BCI), filter banks, Riemannian Tangent Space.**

I. Introduction

BRAIN computer interfaces (BCIs) can generate control commands through the recognition of brain activity to help people with disabilities of movement control external devices such as wheelchairs and exoskeletons [1]–[3]. The electroencephalogram (EEG) is widely used because of its high time resolution, convenience, and low cost [4]. Current state-of-the-art BCI systems, exploit a few key properties of the EEG including event-related potentials [5], steady-state visual evoked potentials [6], and motor imagery [7]–[9]. Motor imagery has attracted attention because it is intuitive and can be performed without reliance on external cues [10], [11]. When people perform motor imagery, there is a localized suppression of synchronous EEG activity over the motor cortex [12], [13]. Specifically, the energy of EEG rhythm in the contralateral motor-sensory area of the cerebral cortex is significantly reduced, while the energy of the EEG in the ipsilateral motor-sensory area is increased [14]. This phenomenon is called event-related desynchronization (ERD)/ event-related synchronization (ERS) [15]. According to this rule, different control commands can be generated by classifying the EEG signals from the different parts of the brain that control individual limbs [16], [17].

The nonstationary, low amplitude and low signal-to-noise ratio of the EEG bring great challenges to the stability of MI-BCI systems [18]. To overcome these challenges, it is very important to extract effective features for use in EEG recognition. The common spatial patterns (CSP) algorithm is one of the most frequently used feature extraction methods that has been proposed to address this problem. It extracts effective features by constructing an optimal spatial filter that differentiates two conditions. However, the performance of the traditional CSP method depends largely on the specific operational frequency band; the frequency band that contains the majority of the useful information in the EEG during the BCI control task. Inter-trial or inter-category variance in this operational frequency band can reduce the effectiveness of CSP [18], [19]. Furthermore, the covariance matrix constructed by the CSP method is vulnerable to

Manuscript received April 20, 2021; revised November 23, 2021; accepted January 22, 2022. Date of publication January 27, 2022; date of current version June 6, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62176090, in part by the Program of Introducing Talents of Discipline to Universities through the 111 Project under Grant B17017, in part by the ShuGuang Project supported by the Shanghai Municipal Education Commission and the Shanghai Education Development Foundation under Grant 19SG25, in part by the Ministry of Education and Science of the Russian Federation under Grant 14.756.31.0001, and in part by the Polish National Science Center under Grant UMO-2016/20/W/NZ4/00354. *(Corresponding author: Jing Jin.)*

Hua Fang is with the Key Laboratory of Smart Manufacturing in Energy Chemical Process Ministry of Education, East China University of Science and Technology, Shanghai 200237, China (e-mail: fanghua_1217@126.com).

Jing Jin is with the Key Laboratory of Smart Manufacturing in Energy Chemical Process Ministry of Education, East China University of Science and Technology, Shanghai 200237, China, and also with the Shenzhen Research Institute, East China University of Technology, Shen Zhen, Guangdong 518063, China (e-mail: jinjingat@gmail.com).

Ian Daly and Xingyu Wang are with the Brain-Computer Interfacing and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, U.K. (e-mail: i.daly@essex.ac.uk; xywang@ecust.edu.cn).

Digital Object Identifier 10.1109/JBHI.2022.3146274

noise interference, resulting in poor generalization performance of CSP.

To solve the problem of variance in the operational frequency band, the Sub-band Common Spatial Pattern (SBCSP) method [20] was proposed and achieved good results on public datasets. SBCSP uses Gabor filter banks to decompose EEG measurements into multiple sub-bands, it then uses discriminant analysis to extract SBCSP features. Then, the SBCSP features are input into linear discriminant analyzers (LDA) to identify the scores reflecting the classification ability of each band. The sub-band scores are then fused by recursive band elimination or via a classification algorithm [20].

An alternative method that has been proposed to solve the problem of operational frequency band variance is called Filter Banks Common Spatial Pattern (FBCSP) [19]. The FBCSP algorithm calculates the corresponding features from the participant-specific spatiotemporal filters to classify the EEG on a single trial basis [21]. The FBCSP method first filters the EEG into multiple frequency bands and then extracts CSP features from each frequency band. It then uses a feature selection algorithm to automatically select the corresponding CSP features. Finally, it uses a classification algorithm to classify the CSP features.

To solve the problem of noise interference, A. Barachant *et al.* [22], [23] proposed the use of Riemannian methods to improve the robustness of the CSP algorithm and to reduce noise interference. The first method they proposed was to directly process the spatial covariance matrix of the EEG in the original space by making use of the robustness of Riemannian distance and Riemannian mean [22]. Because the spatial information about the EEG is embedded in the spatial covariance matrix, spatial filtering is no longer needed. Another method is to map the covariance matrix to the Riemannian tangent space [23]. In Riemannian tangent space, the matrices can be vectorized and treated as Euclidean objects. Similarly, the concepts of Riemannian distance and mean value are also used in this method. The results show that the first method has comparable performance to the traditional CSP method, while the second method has better performance than the CSP method.

Although the above method has achieved satisfactory results to a certain extent, it does not solve the two problems of variance in the operational frequency and interference from noise at the same time. In this work, we propose a fusion method combining multiple filter banks and Riemannian tangent space (FBRTS) to solve the two aforementioned problems at the same time. First, the original EEG is band-pass filtered into multiple frequency bands, and then the covariance matrix of the EEG is calculated to identify spatial filters in each frequency band. Because the Riemannian distance and mean value are robust to noise in the Riemannian tangent space, the covariance matrix is vectorized and treated as a set of Euclidean features. In addition, to account for inter-participant differences, we divide the motor imagery process into multiple time windows and use the fusion method to extract the covariance features in each time window.

The rest of this paper is organized as follows: the second section describes related theories and the system framework of our proposed method. The third section introduces the dataset and experimental setup used in this paper. The fourth section introduces the results of our experiments and comparative studies. The fifth section discusses the proposed method and looks forward to the future research direction. Finally, the sixth section puts forward the conclusion.

II. Methods

A. Filter Bank Common Spatial Patterns

The FBCSP algorithm is a machine learning method for EEG data processing [21]. The key to the traditional FBCSP algorithm is frequency filtering and spatial filtering [19]. The frequency filtering step filters the EEG into multiple frequency bands. The CSP algorithm is then used for spatial filtering to extract features.

Within CSP, the EEG is transformed linearly by using (1)

$$S_{i,j} = W_i^T Z_{i,j} \tag{1}$$

where $Z_{i,j} \in R^{c\times\kappa}$ is the EEG of the *j*-th single-trial obtained by the *i*-th band-pass filter; $S_{i,j} \in R^{c\times\kappa}$ is the obtained from $Z_{i,j}$ after spatial filtering, and W_i is the CSP projection matrix. ^{T} is transpose operator; *c* is the number of channels and κ is the number of the EEG samples for each channel.

The CSP algorithm calculates the transformation matrix by solving the eigenvalue decomposition problem

$$D_{i,1}W_i = (D_{i,1} + D_{i,2})\, W_i\Lambda_i \tag{2}$$

where $D_{i,1}$ and $D_{i,2}$ represent the estimation of the *i*-th band-pass filtered EEG covariance matrices for different motor imagery categories; Λ_i represents the diagonal matrix containing the eigenvalues of the covariance matrix $D_{i,1}$.

The spatial filtering signal $S_{i,j}$ in (1) is replaced by W_i in eq (2), so that the variance difference of the different categories of the band-pass filtered EEG signals is the largest [21]. (3) gives the *m*-pair CSP characteristics of the EEG signals obtained from the *i*-th band-pass filter

$$f_{i,j} = \log\left[\frac{diag\left(\bar{W}_i^T Z_{i,j} Z_{i,j}^T \bar{W}_i\right)}{tr\left[\bar{W}_i^T Z_{i,j} Z_{i,j}^T \bar{W}\right]}\right] \tag{3}$$

where $f_{i,j} \in R^{1\times 2m}$ and $\bar{W}_i$ represents the first *m* columns and the last *m* columns of the transformation matrix W_b; $diag(\cdot)$ represents the diagonal elements of the square matrix, and $tr(\cdot)$ represents the sum of the diagonal elements of the square matrix.

Then (4) is used to construct the FBCSP eigenvector of the *i*-th trial

$$f_i = [f_{1,i}, f_{2,i}, \ldots, f_{k,i}] \tag{4}$$

where $f_i \in R^{1\times(k\times 2m)}$ and *k* is the number of the frequency bands.

Some methods to extend the CSP algorithm to multi-category problems have been proposed [24]. A successful technique is to perform two types of CSP for all possible combinations of different categories. In the case of λ categories, this results in at least *T* spatial filters

$$T = \frac{\lambda\,(\lambda-1)}{2}, (\lambda \geq 2, \lambda \in N^*) \tag{5}$$

where N^* is a positive integer. In the case of λ categories, we take $m = 2$, that is, from each band-pass filtered EEG signal,

Algorithm 1: Iterative gradient descent algorithm for finding the Riemannian mean.

Input: N covariance matrices $\{D_i\}_{i=1}^{n} \in R^{N\times N}$, iteration threshold $\varepsilon > 0$.
Output: The Riemannian mean of n covariance matrices.
Begin
Initialize a zero matrix $\bar{D}_0 = 0 \in R^{N\times N}$;
Identity a matrix $\bar{D}_0 = I \in R^{N\times N}$;
while $\delta_R(\bar{D}, \bar{D}_0) = \|\log m(D_1^{-1}D_2)\|_F > \varepsilon$
The matrix $\bar{D}$ is assigned to $\bar{D}_0$, that is $\bar{D}_0 = \bar{D}$;
Calculate the mean $\tilde{D} = \frac{[\sum_{i=1}^{n} Log_{\bar{D}_0}(D_i)]}{n}$ of n covariance matrices projected into the tangent space;
Obtain a new mean $\bar{D}$ by projecting $\tilde{D}$ on the tangent space back to sub-manifold via (10).
End

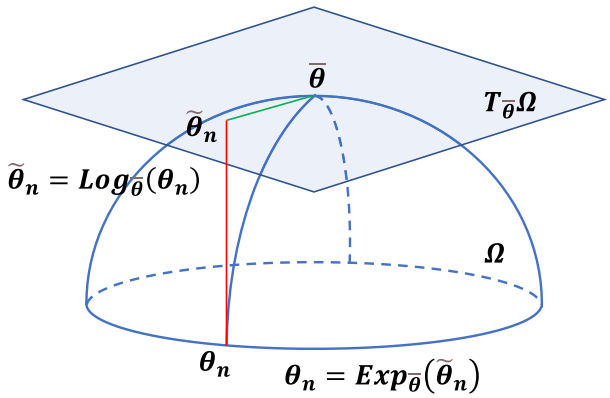


Fig. 1. Manifold of reference points and corresponding local tangent space.

we get a pair of CSP features through spatial filtering, producing a total of $2T$ features. Therefore, for a participant's EEG data, the total number of features we can get is

$$N_{feature}(\lambda) = N_T \times N_B \times 2T \tag{6}$$

where N_T and N_B represent the number of time windows [25] and frequency bands.

B. Riemannian Geometry

In MI-BCIs, Riemannian methods are typically applied to the covariance matrix of the EEG [23], [26]. These covariance matrices are real symmetric positive definite, which can form a differentiable Riemannian manifold with dimension $R(R+1)/2$, and allow mapping between the manifold and its corresponding local tangent space [27], where R is the row number of the covariance matrix.

We can select a fixed reference point $\bar{\theta} \in \Omega$ and the corresponding tangent space is denoted as $T_{\bar{\theta}}\Omega$, as shown in Fig. 1. According to the principle of Riemannian geometry, we use logarithmic mapping to project the vector from the Sub-manifold Ω to its tangent space [23] and use exponential mapping [28] to project the points on the tangent space back to the Sub-manifold

$$\tilde{\theta}_n = Log_{\bar{\theta}}(\theta_n) = \bar{\theta}^{1/2} \log m\left(\bar{\theta}^{-1/2}\theta_n\bar{\theta}^{-1/2}\right)\bar{\theta}^{1/2} \tag{7}$$

$$\theta_n = Exp_{\bar{\theta}}\left(\tilde{\theta}_n\right) = \bar{\theta}^{1/2} \exp m\left(\bar{\theta}^{-1/2}\tilde{\theta}_n\bar{\theta}^{-1/2}\right)\bar{\theta}^{1/2} \tag{8}$$

where $\log m(\cdot)$ and $\exp m(\cdot)$ are the logarithmic function and the exponential function of the matrix [29].

Generally, the covariance matrices D_1 and D_2 constructed by the CSP spatial filters are real matrices [30], so the Euclidean distance between two different covariance matrices can be expressed as

$$\delta_E(D_1, D_2) = \|D_1 - D_2\|_F \tag{9}$$

where $\delta_E(\cdot)$ is the distance between the matrices and $\|\cdot\|_F$ is the Frobenius norm [30]. The Riemannian distance of two covariance matrices can be expressed a

$$\delta_R(D_1, D_2) = \left\|\log m\left(D_1^{-1}D_2\right)\right\|_F \tag{10}$$

The difference between the two distances is that the Euclidean distance represents the shortest distance along a straight-line path, while the Riemannian distance represents the shortest path along with a geodesic (curve) search [31], [32].

However, in the case of multiple categories, the number of covariance matrices may be more than two, so we need to consider the measurement between multiple covariance matrices from the perspective of the mean. Suppose that $\{D_i\}_{i=1}^{n}$ is a set of n covariance matrices, then in Euclidean space, its arithmetic mean is

$$\mu(D_1, D_2, \ldots, D_n) = \frac{1}{n}\sum_{i=1}^{n} D_i \tag{11}$$

In the Riemannian manifold, the Riemannian mean value [33] of n covariance matrices can be defined as a matrix minimizing the sum of squares of Riemannian distances

$$\bar{D} = \vartheta(D_1, D_2, \ldots, D_n) = \arg\min_{D} \sum_{i=1}^{n} \delta_R^2(D, D_i) \tag{12}$$

It should be noted that there is no closed-form solution for this mean. However, the iterative gradient descent algorithm [34] can be used to solve this problem, as shown in **algorithm 1.**

C. Feature Extraction Based on Filter Bank and Riemannian Tangent Space

In our work, we propose the FBRTS method to extract covariance features. The framework of this method is shown in Fig. 2**.** First, we divide the EEG epoch recorded during the motor imagery period from the experimental paradigm in order to construct multiple time windows. We then preprocess the original EEG signals in each time window. Specifically, we decompose the EEG into multiple frequency bands, and then use the Riemannian method to extract the covariance features specific to each of the different frequency bands; finally, all the features are vectorized and fed into a classifier in the feature fusion stage.

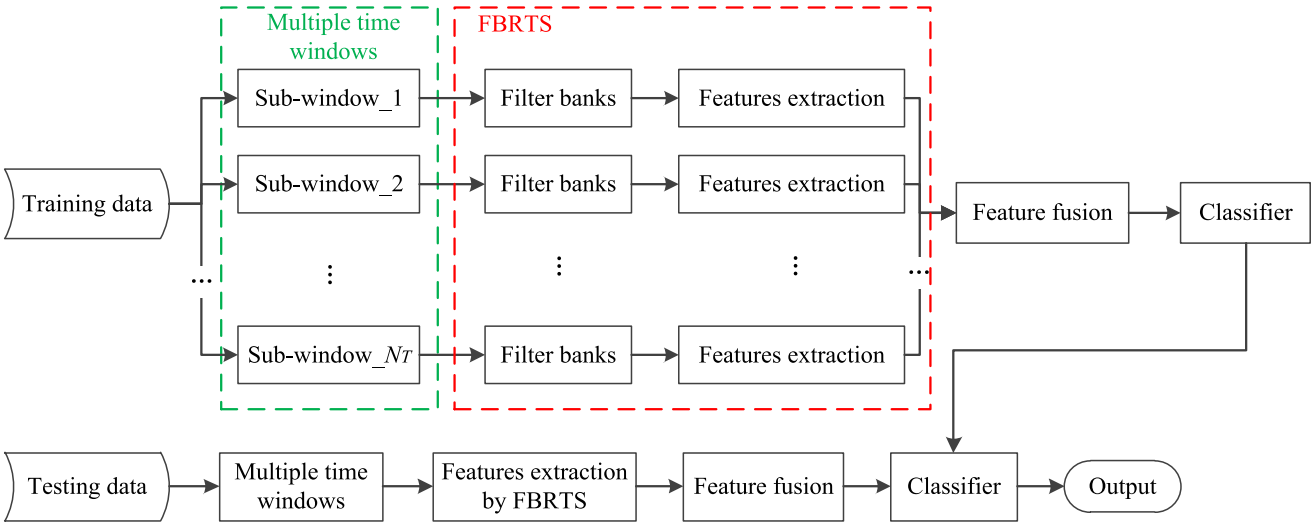


Fig. 2. The system block diagram of our proposed method.

In order to apply Riemannian features in MI-BCIs, we first need to use band-pass filters to get the filtered EEG signals from every single trial. We then estimate its covariance matrix via

$$D_\nu = \frac{1}{N_S} Z_{i,j}^T Z_{i,j}, i, j, \nu = 1, 2, \ldots, n \tag{13}$$

where N_s is the number of sampling points. The process is described further in part A of Section II. Considering the performance of the classifier, we vectorize the covariance matrix as features

$$\begin{aligned} \vec{D}_\nu &= vect\left(D_\nu\right) \\ &= \left[D_{\nu(1,1)}; \sqrt{2} D_{\nu(1,2)}; \cdots D_{\nu(N_C, N_C)}\right] \in R^{(N_C+1)N_C/2} \end{aligned} \tag{14}$$

$$\left\|D_\nu\right\|_F = \left\|vect\left(D_\nu\right)\right\|_2 \tag{15}$$

where N_c is the number of channels. Because the covariance matrix is symmetric, its non-diagonal elements need to be scaled to maintain the standard [35], as shown in eq (14).

Then, we use the Riemannian mean to calculate the Riemannian distance between multiple covariance matrices. The Riemannian feature extraction process is shown in **algorithm 2.**

As mentioned in part A of section II, under the joint action of multiple time windows and filter banks, for a single participant's EEG data, the total number of features we can get is

$$N_{feature}\left(R\right) = N_T \times N_B \times N_C\left(N_C + 1\right)/2 \tag{16}$$

D. Support Vector Machine

The Support vector machine (SVM) is a classical classification method [36], which is widely used in MI-BCIs. The principle of the SVM is to find an optimal hyperplane, which can maximize the distance between training data points between two categories. In addition, the SVM has a strong generalization ability for new unknown data objects and has a flexible decision boundary. Therefore, we use the SVM as the classifier in this paper.

Algorithm 2: The Riemannian feature extraction method.

Input: The filtered EEG signal
$Z_{i,j} \in R^{N_C \times N_S},\ i, j, \nu = 12, \ldots, n.$
Output: $N_C(N_C+1)/2$ Riemannian features.
Begin
Calculate the covariance matrix D_v of the EEG $Z_{i,j}$ via eq (13);
Use the iterative gradient descent algorithm to calculate the Riemannian mean between n covariance matrices;
Calculate $N_C(N_C+1)/2$ Riemannian features via eq (14) and eq (15);
End

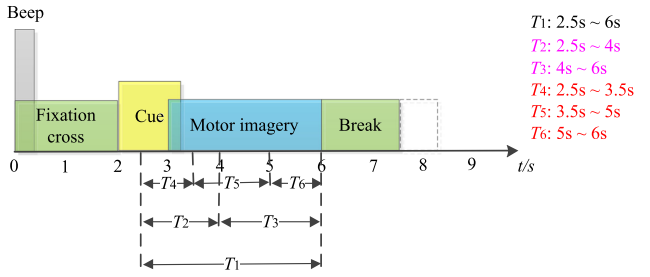
In addition, through the experimental analysis, we select the linear kernel as the kernel function of the SVM model.

The SVM identifies the optimal separating hyperplane defined by $f\ (x) = \boldsymbol{\omega}^T\ x + b$, where b is the deviation and $\boldsymbol{\omega}$ is the weight vector. The hyperplane of the SVM maximizes the class separation and minimizes the classification error. The edge distance is the sum of the distances calculated to the nearest positive sample and the nearest negative sample. In other words, the hyperplane maximizes the edge distance. Therefore, the SVM model can be expressed in the following form

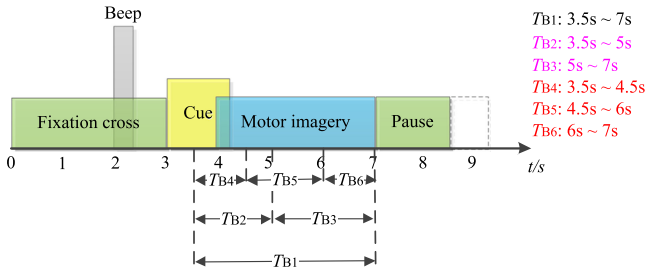
$$\min_{\omega, b, \xi_i} \underbrace{\frac{1}{2}\|\boldsymbol{\omega}\|^2}_{Regularized} + C \cdot \underbrace{\sum_{i=1}^{n} \xi_i}_{Loss}$$

$$s.t. y_i\left(\boldsymbol{\omega}^T \boldsymbol{x}_i + b\right) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \ldots, n \tag{17}$$

where ξ is the relaxation variable, and the eigenvector has a corresponding relaxation variable, which indicates the degree that the eigenvector does not meet the constraint; C is the penalty parameter of the error term ($C > 0$). Support vector machines try to strike a balance between minimizing the regularization term and minimizing the classification error.



(a) BCI competition IV dataset 2a



(b) BCI competition IV dataset 2b

Fig. 3. The time flow and time windows division of each trial.

III. Experimental Study

A. Data Set Description

BCI competition IV dataset 2a: This dataset [37] contains EEG data recorded from 9 healthy participants that performed four motor imagery tasks: left-hand, right-hand, feet, and tongue imaginary movement. Each participant completed two sessions, one for training and the other for evaluation [38]. A total of 25 measurement channels were used to record the dataset, including 22 EEG channels and 3 monopolar electrooculograms (EOG) channels.

BCI competition IV dataset 2b: This dataset [38] included EEG data of 9 healthy participants. All participants were required to complete two different types of motor imagery tasks (left hand and right hand). Each participant has five sessions. The first two sessions are completed without feedback, and the other three sessions are completed with feedback. In addition, among the five sessions, the first three are used for training and the last two are used for testing. The dataset uses 3 electrodes to collect EEG data.

For our evaluation, we only used 22 EEG channels in dataset 2a and 3 EEG channels in dataset 2b. These were recorded with a sampling frequency of 250Hz and were band-pass filtered between 0.5 to 100Hz. The time flow of each trial is shown in Fig. 3. In dataset 2a, at the beginning of each trial, a fixation cross was displayed on a black screen and a short warning beep was issued. Then, a cue in the form of an arrow pointing left, right, down, or up appeared on the screen (corresponding to the four tasks: left-hand, right-hand, feet, and tongue motor imagery). The participants performed the corresponding motor imagery tasks in response to the given cue until the second fixation cross disappeared from the screen. In dataset 2b, at the beginning of each trial, a fixed cross will appear on the screen. Then, a cue in the form of an arrow pointing left or right appeared on the screen (corresponding to the two tasks: left-hand and right-hand motor imagery), and then the participants need to perform corresponding MI tasks according to the arrow direction.

TABLE I
Division of Filter Banks and Bandwidth Selection

Frequency range (Hz)	Bandwidth (Hz)				
	2 (n=1)	4 (n=2)	8 (n=3)	16 (n=4)	32 (n=5)
4~40	4~6	4~8	4~12	4~20	4~36
	6~8	6~10	6~14	6~22	6~38
	8~10	8~12	8~16	8~24	8~40
	…	…	…	…	/
	38~40	36~40	22~40	24~40	/

Note: Bandwidth $= 2^n$ Hz; step size $= 2$ Hz.

B. Experiment Setting

In this study, we extracted the data from an epoch lasting from 0.5s before each motor imagery cue until the end of the motor imagery period (2.5s $\sim$ 6s). We then divided this epoch of between 2.5s $\sim$ 6s into six different time windows, as shown in Fig. 3. To select the number of filter banks, we consider the mixed-use of different bandwidths and select $2^n (n = 5)$ Hz as the bandwidth. To search for the optimal frequency window, we use a step size of 2 Hz. We use a 5th order Butterworth filter to construct the filter bank in the range of 4$\sim$40 Hz, as shown in Table I.

The time window should first include the execution time of the motor imagery tasks, and the length of the time window should not be too short. Too short time windows will increase the number of time windows. According to eq (16), the number of time windows will increase the dimension of the feature, resulting in a significant increase in the calculation time. Therefore, we select the period from 0.5s before the start of the motor imagery tasks to the end of the motor imagery tasks as the time window division range (2.5s $\sim$ 6s), that is, T_1, and the length of other time windows shall not be less than 1s. In addition, according to the short before long and bisection symmetry rules, T_2 and T_3 choose 2s and 2.5s respectively, and $T_4 \sim T_6$ choose 1s, 1.5s, and 1s respectively. Similarly, six time windows ($T_{\mathrm{B1}} \sim T_{\mathrm{B6}}$) corresponding to dataset 2b can be obtained.

Since the BCI competition dataset 2a has four categories of motor imagery tasks, we consider using the One-Versus-Rest (OVR) method [39] to solve the multi-category classification problem. The idea of the OVR method is to transform multiple categories into two categories. In the training phase, for the sample data of λ categories, it is necessary to train λ SVM binary classifiers. The sample data belonging to the i-th SVM sub classifier is marked as the positive category, and the other sample data not belonging to the i-th category is marked as the negative category. The values of each discriminant function are calculated for the test data in the test phase. If only one classifier outputs a positive value, then the result can be directly judged as the corresponding classifier number. Otherwise, the category

TABLE II
CLASSIFICATION ACCURACY (MEAN AND STANDARD DEVIATION IN PERCENTAGE) ACHIEVED WITH THE FBRTS METHOD AND OTHER MI CLASSIFICATION METHODS ON DATASET 2A

Participant	Method					
	FBCSP	HSS-ELM	WaSF ConvNet	TSLDA	CSP-LCD	FBRTS
A01	76.9	82.1	73.4	79.2	72.4	**86.1**
A02	57.2	47.2	50.2	47.5	58.3	**65.2**
A03	84.9	80.2	86.3	85.6	82.1	**90.0**
A04	51.3	**66.2**	55.1	59.9	52.3	63.8
A05	52.9	42.1	68.7	49.4	62.5	**75.6**
A06	46.5	44.3	49.3	54.7	**61.4**	52.4
A07	86.3	85.6	73.5	86.3	82.9	**91.1**
A08	84.1	76.4	86.2	83.4	88.5	**89.0**
A09	82.2	78.3	77.8	85.2	83.4	**86.5**
mean±std	69.1±15.7	66.9±16.6	68.9±13.5	70.1±15.9	71.5±12.5	**77.7±13.6**
p-value	0.005	0.013	0.001	0.032	0.037	/
kappa	0.59	0.62	0.63	0.57	0.61	0.71

corresponding to the maximum value of the discriminant function is selected as the category of the test data. In addition, it should be noted that all experiments used a tenfold cross-fold validation scheme.

IV. EXPERIMENT RESULTS

A. Comparison With Other State-of-the-Art Methods

In order to evaluate the performance of the FBRTS method, we use BCI competition datasets 2a and 2b to compare the classification accuracy of the FBRTS method with other state-of-the-art methods.

1) *HSS-ELM:* A hierarchical semi-supervised extreme learning machine (HSS-ELM) method, which uses the deep structure of a hierarchical ELM to learn features automatically [40].
2) *WaSF ConvNet:* Wavelet spatial filter convolution network (WaSF ConvNet). This method uses the space-time-frequency joint features of EEG data based on time-frequency and spatial filters [41].
3) *TSLDA:* Tangent space linear discriminant analysis (TSLDA). In this method, the covariance matrix is mapped to tangent space and vectorized as a Euclidean object. [22].
4) *CSP-LCD:* A method combining CSP and local characteristic-scale decomposition (CSP-LCD). The method builds a functional brain network from EEG signals and fuses the features extracted by CSP and LCD algorithms in the frequency and spatial domains [42].

Table II shows the accuracies achieved by our FBRTS method and each of the other methods we compare it to in dataset 2a. Compared with the traditional FBCSP method, our FBRTS method has significantly improved accuracy with all participants ($p<0.05$, paired t-test), and the average classification accuracy is 8.6% higher than the FBCSP method. In addition, compared with other state-of-the-art methods, our FBRTS method performs the best in most participants. Our FBRTS method achieved better classification accuracy than the WaSF ConvNet method for all participants ($p < 0.05$, paired t-test). However, for participants A04 and A06, our FBRTS method still needs to be improved. Our results show that our FBRTS method can make good use of the characteristics of the Riemannian method to improve the performance of the algorithm. Furthermore, the addition of multiple time windows and filter banks not only solves the problem of frequency band dependence but also effectively expands the number of features so that the data can be classified more accurately. Besides, we also calculated the kappa value of each method in all participants. Kappa value [43] is a consistency test index, which can measure the effect of classification and judge whether the predicted results of the model are consistent with the actual classification results. The results show that the kappa value of the FBRTS method is 0.71, which is better than other methods, which shows that the FBRTS method can indeed obtain a better classification effect. For dataset 2b, the FBRTS method also achieved significant results in classification accuracy (86.9% $\pm$ 6.03) and kappa value (0.7) compared with other state-of-the-art methods ($p < 0.05$, paired t-test), as shown in Table III. These experimental results show that the FBRTS method is not only effective for four classification problems in the MI-BCI system but also shows good performance for two classification problems.

The training time is very important when evaluating our proposed FBRTS method. The training time of each method affects its performance to a certain extent. Therefore, we calculated the training time of each method on datasets 2a and 2b. As shown in **Table** IV, the training time of the FBRTS method on datasets 2a and 2b is 18.67s and 14.46s respectively. Although the FBRTS method requires less training time than the latter three methods (WaSF ConvNet, TSLDA, and CSP-LCD), compared with the first two methods (FBCSP and HSS-ELM), the training time of the FBRTS method is longer. It may be caused by the high-dimensional features extracted by FBRTS, which needs further experimental analysis. The above experimental results show that the training time of the FBRTS method is at the medium level among similar methods, and needs further improvement.

In recent years, deep learning technology has achieved good results in the field of MI-BCI, especially in multi-classification problems. Zhang *et al.* [44] proposed a hybrid depth network framework (CNN-LSTM) to improve the classification accuracy of four types of MI-EEG signals. CNN-LSTM obtained 83% classification accuracy and 0.8 kappa value on dataset 2a. Chen *et al.* [45] proposed a deep spatiotemporal feature learning

TABLE III
CLASSIFICATION ACCURACY (MEAN AND STANDARD DEVIATION IN PERCENTAGE) ACHIEVED WITH THE FBRTS METHOD AND OTHER MI CLASSIFICATION METHODS ON DATASET 2B

Participant	Method					
	FBCSP	HSS-ELM	WaSF ConvNet	TSLDA	CSP-LCD	FBRTS
B01	68.6	64.4	73.8	76.3	75.1	**82.4**
B02	61.7	64.3	64.2	68.9	64.2	**75.2**
B03	81.2	82.7	85.7	86.4	**88.7**	86.9
B04	90.4	86.7	**96.2**	94.2	95.4	95.2
B05	82.3	85.2	85.2	88.1	86.2	**89.7**
B06	63.7	64.5	68.5	72.3	74.2	**80.2**
B07	84.2	77.3	88.3	89.2	90.2	**90.5**
B08	87.2	83.5	90.1	**92.8**	88.2	91.2
B09	74.5	80.6	81.5	87.3	89.1	**91.1**
mean±std	77.1±9.8	76.6±8.97	81.5±9.97	83.9±8.61	83.5±9.45	**86.9±6.03**
p-value	0.001	0.001	0.009	0.022	0.034	/
kappa	0.60	0.59	0.63	0.65	0.66	0.70

TABLE IV
TRAINING TIME (SECOND) CONSUMED USING THE FBRTS METHOD AND OTHER MI CLASSIFICATION METHODS ON DATASET 2A AND 2B

Method	FBCSP	HSS-ELM	WaSF ConvNet	TSLDA	CSP-LCD	FBRTS
Dataset 2a	13.91	16.83	36.43	25.27	26.72	18.67
Dataset 2b	9.21	11.53	24.48	19.65	19.47	14.46

method based on MI-BCI (FBSF-TSCNN), which achieved 72% classification accuracy and 0.627 kappa value on dataset 2a. Compared with these deep learning methods, our proposed FBRTS method has certain competitive advantages in classification accuracy (77.7%) and kappa value (0.7), but it may not achieve the best effect. However, in terms of training time, the training time required by CNN-LSTM and FBSF-TSCNN on dataset 2a is 28s and 24.37s respectively, which is more time consumed than FBRTS method. These results show that FBRTS method still has a good competitive advantage compared with deep learning method.

B. Optimal Numbers of Time Windows and Filter Banks

We now consider the influence of the number of filter banks and time windows on the performance of the FBRTS method. We also attempt to identify the most suitable number of filter banks and time windows for the FBRTS method. It should be noted that the number of filter banks will change as the bandwidth changes. Therefore, it is meaningless to simply compare the impact of the number of filter banks on performance under different bandwidths. Instead, we consider the way of accumulating filter banks for analysis. Specifically, when we analyze the performance using filter banks with a bandwidth of 2 Hz, we do not replace the filter banks but add a filter bank with the bandwidth of 4 Hz onto the basis of the original filter bank, and then carry out the experiment. We iterate this process until all filter banks with different bandwidths are added to the experiment. In this way, with the progress of each group of experiments, the number of filter banks always keeps increasing. The results are shown in Fig. 4.

Fig. 4 illustrates that, as the number of filter banks in different time windows increases, the classification accuracy of the FBRTS method also increases. Therefore, we measure Pearson correlation coefficients between the number of filter banks and the classification accuracy. The results are listed in Table V. There is a significant positive correlation between the number of filter banks and the classification accuracy ($p < 0.01$) under each of the different time windows.

TABLE V
RESULT OF PEARSON CORRELATION ANALYSIS BETWEEN THE NUMBER OF FILTER BANKS AND CLASSIFICATION ACCURACY IN DIFFERENT TIME-WINDOWS

Time-windows	Correlation	p-value
1	0.999	**
2	0.996	**
3	0.974	**
4	0.993	**
5	0.997	**
6	0.977	**

Note: ** denotes a significance of $p < 0.01$.

This shows that increasing the number of filter banks has a positive impact on the FBRTS method. However, with the increase in the number of time windows, the accuracy of the FBRTS method under different filter banks first increases and then decreases. One possible reason is that the increase in the number of time windows introduces redundancy in the feature set, which makes the performance of the system decline. However, with a continuous increase in the number of time windows, the performance of the system improves as the number of features increases and achieves optimal performance in three time windows (T_1, T_2, T_3).

C. Comparison With the FBCSP Method

1) The number of filter banks and time windows: In order to further analyze the FBRTS method, we observe the performance

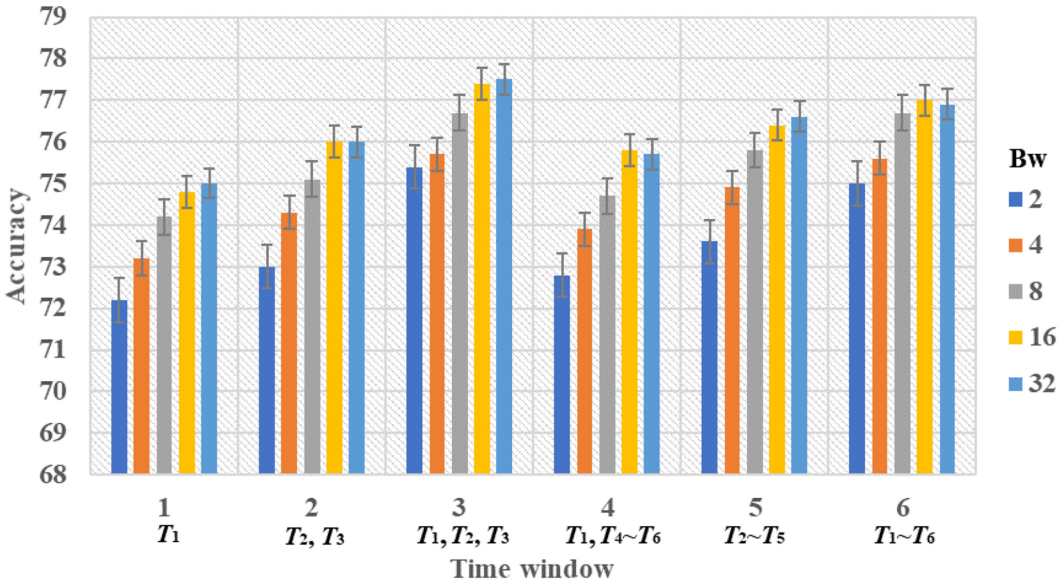


Fig. 4. The influence of different time-windows and filter banks on classification accuracy (Bw denotes Bandwidth).

of the FBRTS and the FBCSP methods under different numbers of filter banks and time windows. These results are shown in Fig. 5. Please note that the filter bank in Fig. 5 also adopts the mode of accumulation filter banks, as described in part B of section IV.

From the comparison of the two methods, it can be seen that the accuracy of the FBRTS is better than that of the traditional FBCSP method. This effect is particularly strong for participant A05. For this participant, the FBRTS method achieves an accuracy of more than 60% under a range of different numbers of filter banks and time windows, but the performance of the traditional FBCSP is mostly about 50%, or even lower.

Overall, the FBRTS method produces better performance than the FBCSP method. However, this improvement is not as obvious for other participants as for participant A05. The above results also confirm the performance seen for participants A05 and A07 in Table II. The accuracy achieved for participant A05 is 52.9% and 75.6% respectively under the FBRTS and FBCSP methods. The accuracy achieved for participant A07 is 86.6% and 91.1% respectively under the two methods. Taken together, this indicates that the FBRTS method can improve both the best and the worst performing participants when compared to the FBCSP method. Furthermore, it can be seen that the FBRTS method can achieve or even surpass the best performance of the FBCSP method by using fewer filter banks and time windows (as shown in Fig. 5). This further demonstrates the superiority of the FBRTS method.

2) Features distribution: We use the visualization method to display the features extracted by the two methods, as shown in Fig. 6. It should be noted that due to the use of more filter banks and time windows, the dimensionality of the features extracted by the two methods is relatively high. Therefore, we use the t-distributed stochastic neighbor embedding (t-SNE) method [46] to visualize the features extracted by two different methods in two-dimensional space.

The t-SNE method integrates dimensionality reduction and visualization. It is based on the SNE visualization method. The t-SNE method solves the problems of crowded sample distributions and unobvious boundaries in SNE visualization. Consequently, it is a better method of dimension reduction and visualization. In Fig. 6, for participant A05, the features of "feet" and "tongue" obtained by the FBCSP method are highly fused, which makes it difficult for the classifier to distinguish them in the process of classification, resulting in low accuracy. The FBRTS method can better separate the features into four categories. Therefore, it can obtain better performance compared with the FBCSP method. Similarly, the FBCSP features during the "left-hand" and "feet" imagery tasks performed by participant A07 are also partially fused, while the four features identified by the FBRTS method are more distinctive. Therefore, the classification performance achieved with the FBRTS method is better than the FBCSP method, but there is still interference from outliers.

V. Discussion

Identifying an optimal feature extraction method is a key problem in the research of MI-BCIs [18]. How to extract the most effective features for discrimination of multi-category motor imagery tasks is also a highly relevant and challenging research problem. The CSP algorithm is one of the most widely used feature extraction methods in MI-BCIs. It shows excellent performance for the problem of binary classification, but it still has significant limitations when applied to the field of multi-category classification. The traditional CSP method is also limited by variability in the operational frequency band of the EEG and by noise interference [18], [19], which leads to poor generalization performance of CSP.

Both SBCSP [20] and FBCSP [21] have been proposed as solutions to improve the CSP method. However, they are both

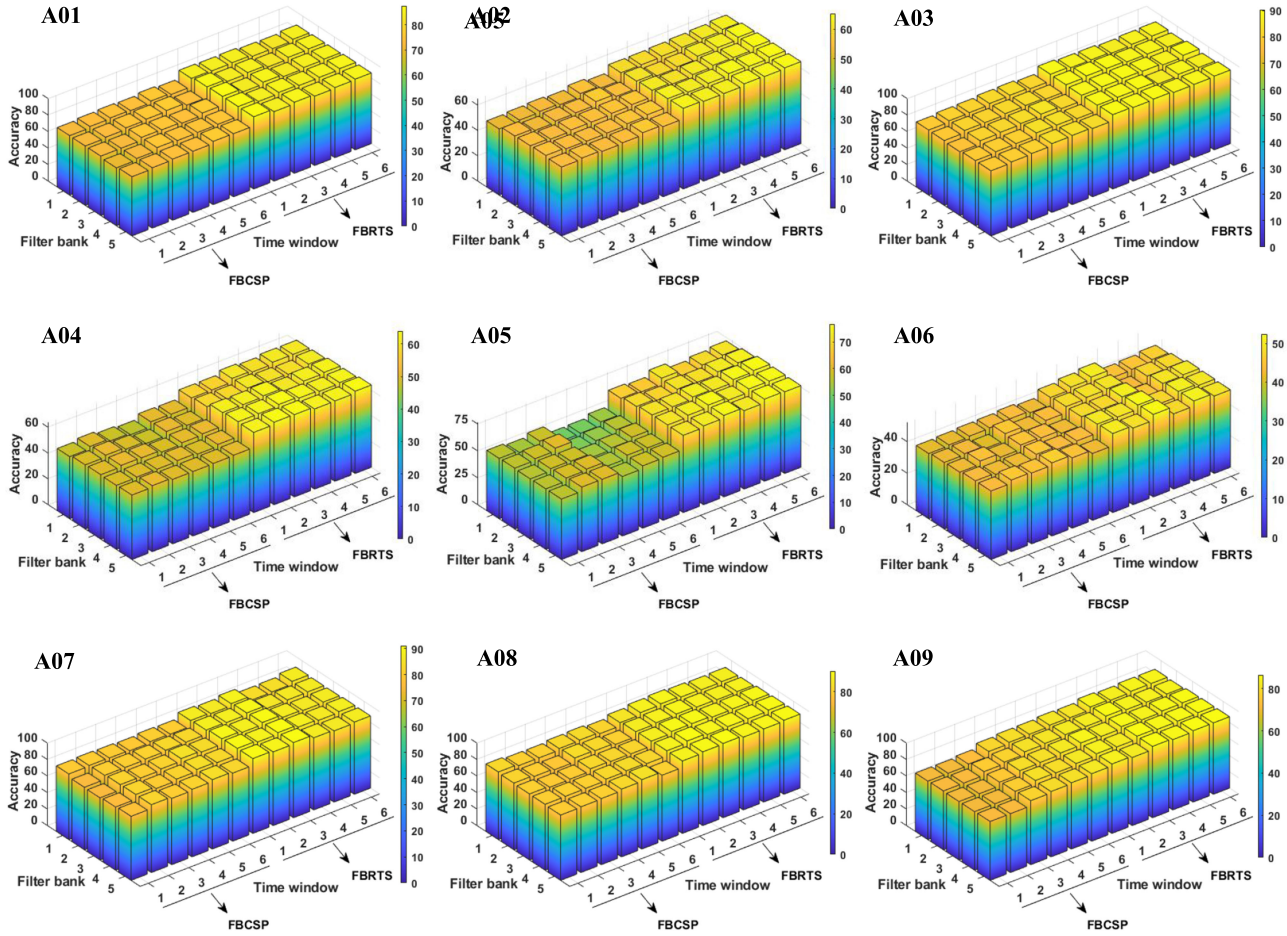


Fig. 5. The performance comparison for all participants using the FBRTS and the FBCSP methods under different numbers of filter banks and time-windows (Filter bank: *n* represents 2^n, n = 1, 2, …, 5).

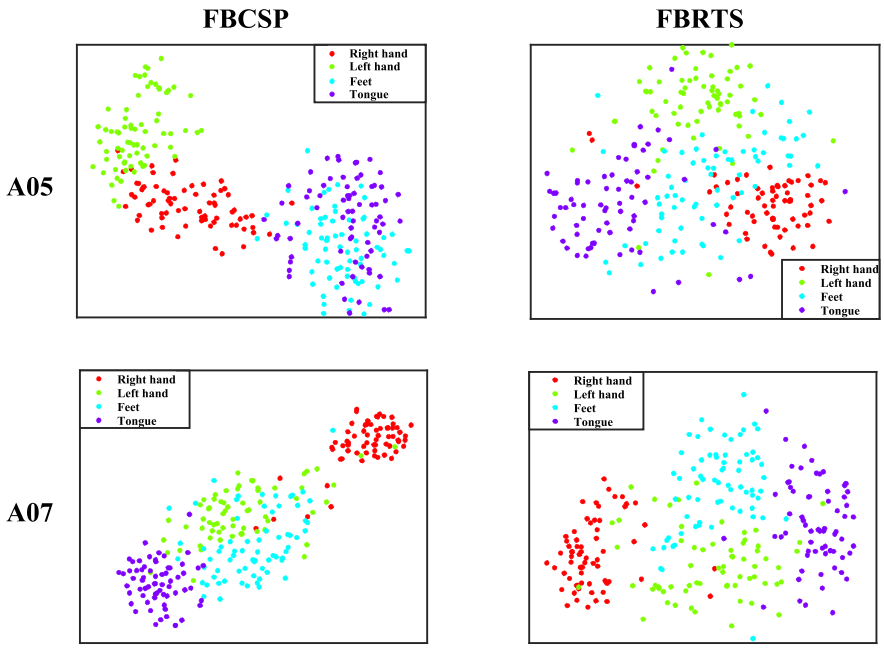


Fig. 6. Comparison of features distributions for participants A05 and A07 under the two methods.

limited by variability in the operational frequency band of the EEG. They are also susceptible to the same problem that the covariance matrix of the traditional CSP method is susceptible to, namely noise interference. Thus, some Riemannian methods have been proposed to extract more robust features and have achieved improved results. In this work, we propose to use the FBRTS method to solve the problem of variability in the operational frequency band and noise interference in multi-category classification problems. The FBRTS method combines filter banks with a Riemannian method in multiple time windows to attempt to extract optimal features. After combining with a linear SVM classifier, the method shows good results.

The FBRTS method uses the Riemannian method to construct robust features and expands the dimensionality of multiple category features through using multiple time windows and filter banks. In addition, we use the classic OVR strategy. This strategy shows significant advantages compared with other methods, as shown in Table II. We find that the number of filter banks has a positive impact on the performance of the FBRTS method, and the increase in the number of time windows may introduce redundancy characteristics, which will reduce the performance of the FBRTS method. In Fig. 4, we illustrate how the optimal number of filter banks and time windows can be found.

In addition, we compare the FBRTS method with the traditional FBCSP method. The results show that the FBRTS method achieves or even exceeds the best performance of the FBCSP method with only a few filter banks and time windows. On the other hand, we use the t-SNE method to visualize the features obtained by the two methods in two-dimensional space. The results show that the FBRTS method can achieve better performance than the FBCSP method. This further demonstrates the effectiveness of the FBRTS method.

Although the FBRTS method has achieved better performance in all our metrics, there are still some areas in which it can be further improved. First, the FBRTS method still exhibits poor performance with a small number of participants in our test dataset, such as participants A04 and A06. In addition, due to the introduction of multiple filter banks and time windows, the dimensionality of the features extracted by the FBRTS method is relatively high. In this study, we do not use any additional feature selection methods to further select sub-sets of features. Consequently, when the dimensionality of the selected features is high the classification problem is considerably more challenging. To attempt to resolve this, we choose relevant parameters for the classifier (such as regularization parameters and penalty terms) according to the empirical method. However, this does not necessarily result in optimal parameters for SVM classifiers, as used with the FBRTS method. Therefore, the dimensionality of the selected features has an impact on the performance of the system. In the future, we will consider introducing a further feature selection method to select a subset of features extracted by the FBRTS method in order to further reduce the dimensionality of features and reduce the burden on the classifiers. On the other hand, we will consider introducing a parameter optimization method to automatically select the optimal classifier parameters to further improve the performance of the FBRTS method in the MI-BCI system.

The proposed FBRTS method is an exploration in the field of machine learning and has made certain advantages compared with other state-of-the-art methods in the same field. However, in recent years, deep learning technology and transfer learning [47] methods have made great development in the field of brain-computer interface. Especially in the MI-BCI system of multi-classification problems, deep learning technology, and transfer learning method have been used by more scholars and researchers, and achieved remarkable results. Therefore, the FBRTS method integrating deep learning technology or transfer learning method will be a new direction worthy of thinking and researching in the future.

VI. Conclusion

Feature extraction is one of the key steps to improve the performance of MI-BCIs. However, traditional feature extraction methods cannot completely overcome the problems of variability in the operational frequency band of the EEG and noise interference. Methods such as CSP and FBCSP suffer degraded system performance due to these problems.

In this study, we propose a novel FBRTS method, which combines multiple filter banks and the Riemannian tangent space. First, we band-pass filter the original EEG into several frequency bands. In each frequency band, we use a spatial filter to get the covariance matrix of the original signal. This allows us to solve the problem of variability in the operational frequency band of the EEG. Then, we use the Riemannian method to map the covariance matrix to the Riemannian tangent space. In the Riemannian tangent space, the covariance matrix is vectorized and treated as a Euclidean object. Finally, we use the OVR classification strategy to generate the result through the linear kernel SVM classifier.

Experimental results on the BCI IV dataset 2a and dataset 2b demonstrate that the average classification accuracy of our FBRTS method is 77.7% and 86.9% in datasets 2a and 2b respectively. Furthermore, the FBRTS method can obtain more distinctive features than the FBCSP method in two-dimensional embedding space. However, the FBRTS method needs to be further improved in terms of feature selection and parameter optimization in future work. In conclusion, the FBRTS method can effectively aid the research and development of MI-BCI systems.

References

[1] H. Si-Mohammed *et al.*, "Towards BCI-based interfaces for augmented reality: Feasibility, design and evaluation," *IEEE Trans. Vis. Comput. Graph.,* vol. 26, no. 3, pp. 1608–1621, Mar. 2020.

[2] J. Jin, Z. Chen, R. Xu, Y. Miao, X. Wang, and T.-P. Jung, "Developing a novel tactile P300 brain-computer interface with a cheeks-stim paradigm," *IEEE Trans. Biomed. Eng.,* vol. 67, no. 9, pp. 2585–2593, Sep. 2020.

[3] D. Zhang, K. Chen, D. Jian, and L. Yao, "Motor imagery classification via temporal attention cues of graph embedded EEG signals," *IEEE J. Biomed. Health Informat.,* vol. 24, no. 9, pp. 2570–2579, Sep. 2020.

[4] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-Norm and dempster-shafer theory," *IEEE Trans. Neural Netw. Learn. Syst.,* vol. 32, no. 11, pp. 4814–4825, Nov. 2021.

[5] S. M. M. Martens and J. M. Leiva, "A generative model approach for decoding in the visual event-related potential-based brain-computer interface speller," *J. Neural Eng.,* vol. 7, no. 2, Apr. 2010, Art. no. 026003.

[6] D. Saravanakumar and M. R. Reddy, "A high performance hybrid SSVEP based BCI speller system," *Adv. Eng. Informat.,* vol. 42, Oct. 2019, Art. no. 100994.

[7] J. Jin *et al.*, "Bispectrum-based channel selection for motor imagery based brain-computer interfacing," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 28, no. 10, pp. 2153–2163, Oct. 2020.

[8] A. Schloegl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005," *J. Neural Eng.,* vol. 2, no. 4, pp. L14–L22, Dec. 2005, Art. no. Pmid 16317224.

[9] Y. Rong, X. Wu, and Y. Zhang, "Classification of motor imagery electroencephalography signals using continuous small convolutional neural network," *Int. J. Imag. Syst. Technol.,* vol. 30, no. 3, pp. 653–659, Sep. 2020.

[10] F. Cincotti *et al.*, "The use of EEG modifications due to motor imagery for brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 11, no. 2, pp. 131–133, Jun. 2003.

[11] X. Wu, B. Zhou, Z. Lv, and C. Zhang, "To explore the potentials of independent component analysis in brain-computer interface of motor imagery," *IEEE J. Biomed. Health Informat.,* vol. 24, no. 3, pp. 775–787, Mar. 2020.

[12] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Netw.,* vol. 118, pp. 262–270, Oct. 2019.

[13] J. Meng, X. Sheng, D. Zhang, and X. Zhu, "Improved semisupervised adaptation for a small training dataset in the brain-computer interface," *IEEE J. Biomed. Health Informat.,* vol. 18, no. 4, pp. 1461–1472, Jul. 2014.

[14] N. J. Hill *et al.*, "Classifying event-related desynchronization in EEG, ECoG and MEG signals," in *Proc. Joint Pattern Recognit. Symp.*, 2006, pp. 404–413.

[15] C. Li, T. Jia, Q. Xu, L. Ji, and Y. Pan, "Brain-computer interface channel-selection strategy based on analysis of event-related desynchronization topography in stroke patients," *J. Healthcare Eng.,* vol. 2019, Aug. 2019, Art. no. 3817124.

[16] J. Jin *et al.*, "The study of generic model set for reducing calibration time in P300-based brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 28, no. 1, pp. 3–12, Jan. 2020.

[17] Y. Jiao *et al.*, "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health Informat.,* vol. 23, no. 2, pp. 631–641, Mar. 2019.

[18] U. Talukdar, S. M. Hazarika, and J. Q. Gan, "Adaptive feature extraction in EEG-based motor imagery BCI: Tracking mental fatigue," *J. Neural Eng.,* vol. 17, no. 1, Feb. 2020, Art. no. 016020.

[19] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2008, pp. 2390–2397.

[20] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (SBCSP) for brain-computer interface," in *Proc. 3rd Int. IEEE/EMBS Conf. Neural Eng.*, 2007, pp. 204–207.

[21] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) algorithm using online adaptive and semi-supervised learning," in *Proc. Int. Joint Conf. Neural Netw.*, 2011, pp. 392–396.

[22] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain-computer interface classification by Riemannian geometry," *IEEE Trans. Biomed. Eng.,* vol. 59, no. 4, pp. 920–928, Apr. 2012.

[23] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neurocomputing,* vol. 112, pp. 172–178, Jul. 2013.

[24] Z. Y. Chin, K. K. Ang, C. Wang, C. Auan, and H. Zhang, "Multi-class filter bank common spatial pattern for four-class motor imagery BCI," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2009, pp. 571–574.

[25] J. Feng *et al.*, "Towards correlation-based time window selection method for motor imagery BCIs," *Neural Netw.,* vol. 102, pp. 87–95, Jun. 2018.

[26] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in brain-computer interfaces: A review," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 25, no. 10, pp. 1753–1762, Oct. 2017.

[27] W. Förstner and B. Moonen, "A metric for covariance matrices," in *Geodesy-the Challenge of the 3rd Millennium*, Berlin, Germany: Springer, 2003, pp. 299–309.

[28] M. Berger, *A Panoramic View of Riemannian Geometry*. Berlin, Germany: Springer, 2012.

[29] D. Wu, B. J. Lance, V. J. Lawhern, S. Gordon, T.-P. Jung, and C.-T. Lin, "EEG-based user reaction time estimation using Riemannian geometry features," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 25, no. 11, pp. 2157–2168, Nov. 2017.

[30] M. Hersche *et al.*, "Fast and accurate multiclass inference for MI-BCIs using large multiscale temporal and spectral features," in *Proc. 26th Eur. Signal Process. Conf.*, 2018, pp. 1690–1694.

[31] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM J. Matrix Anal. Appl.,* vol. 26, no. 3, pp. 735–747, 2005.

[32] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Geometric means in a novel vector space structure on symmetric positive-definite matrices," *SIAM J. Matrix Anal. Appl.,* vol. 29, no. 1, pp. 328–347, 2007.

[33] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int. J. Comput. Vis.,* vol. 66, no. 1, pp. 41–66, Jan. 2006.

[34] P. T. Fletcher and S. Joshi, "Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors," in *Computer Vision Mathematical Methods Methods Biomedical Image Analysis*, M. Sonka, I. A. Kakadiaris, and J. Kybic, Eds., Berlin, Germany: Springer, 2004, pp. 87–98.

[35] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 30, no. 10, pp. 1713–1727, Oct. 2008.

[36] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.,* vol. 10, no. 5, pp. 988–999, Sep. 1999.

[37] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008–Graz data set A," *Inst. Knowl. Discov. Lab. Brain-Comput. Interfaces Graz Univ. Technol.,* vol. 16, pp. 1–6, 2008.

[38] M. Tangermann *et al.*, "Review of the BCI competition IV," *Front. Neurosci.,* vol. 6, 2012, Art. no. 55.

[39] G. Dornhege, B. Blankertz, G. Curi, and K. R. Muller, "Increase information transfer rates in BCI by CSP extension to multi-class," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 733–740.

[40] Q. She, B. Hu, Z. Luo, N. Thinh, and Y. Zhang, "A hierarchical semi-supervised extreme learning machine method for EEG recognition," *Med. Biol. Eng. Comput.,* vol. 57, no. 1, pp. 147–157, Jan. 2019.

[41] D. Zhao, F. Tang, B. Si, and X. Feng, "Learning joint space-time-frequency features for EEG decoding on small labeled data," *Neural Netw.,* vol. 114, pp. 67–77, Jun. 2019.

[42] Q. Ai *et al.*, "Feature extraction of four-class motor imagery EEG signals based on functional brain network," *J. Neural Eng.,* vol. 16, no. 2, Apr. 2019, Art. no. 026032.

[43] J. Carletta, "Assessing agreement on classification tasks: The Kappa statistic," *Comput. Linguistics,* vol. 22, no. 2, pp. 249–254, Jun. 1996.

[44] R. Zhang, Q. Zong, L. Dou, and X. Zhao, "A novel hybrid deep learning scheme for four-class motor imagery classification," *J. Neural Eng.,* vol. 16, no. 6, Dec. 2019, Art. no. 066004.

[45] J. Chen, Z. Yu, Z. Gu, and Y. Li, "Deep temporal-spatial feature learning for motor imagery-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 28, no. 11, pp. 2356–2366, Nov. 2020.

[46] L. van der Maaten, "Accelerating t-SNE using tree-based algorithms," *J. Mach. Learn. Res.,* vol. 15, pp. 3221–3245, Oct. 2014.

[47] F. Fahimi, Z. Zhang, W. B. Goh, T.-S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *J. Neural Eng.,* vol. 16, no. 2, Apr. 2019, Art. no. 026007.