

# Application of artificial bee colony algorithm in feature optimization for motor imagery EEG classification

Minmin Miao<sup>1</sup> · Aimin Wang<sup>1</sup> · Feixiang Liu<sup>1</sup>

Received: 18 July 2016 / Accepted: 20 March 2017 / Published online: 31 March 2017  
© The Natural Computing Applications Forum 2017

**Abstract** Motor imagery (MI) tasks evoke event-related desynchronization (ERD) and synchronization (ERS); the ERD-/ERS-related features appearing at specific channels are frequency and time localized. Therefore, optimal channels, frequency band and time interval are of great significance for MI electroencephalography feature extraction. In this paper, channel selection method based on linear discriminant criteria is used to automatically select the channels with high discriminative powers. In addition, the concept of artificial bee colony algorithm is first introduced to find the global optimal combination of frequency band and time interval simultaneously without prior knowledge for common spatial pattern features extraction and classification. Experimental results demonstrate that this scheme can adapt to user-specific patterns and find the relatively optimal channels, frequency band and time interval for feature extraction. The classification results on the BCI Competition III Dataset IVa and BCI Competition IV Dataset IIa clearly present the effectiveness of the proposed method outperforming most of the other competing methods in the literature.

**Keywords** Brain–computer interface (BCI) · Common spatial pattern (CSP) · Artificial bee colony (ABC) · Motor imagery (MI) · Electroencephalogram (EEG)

---

✉ Aimin Wang  
wangam@seu.edu.cn

Minmin Miao  
15006187659@sina.cn

Feixiang Liu  
220152688@seu.edu.cn

<sup>1</sup> School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China

## 1 Introduction

Brain–computer interfaces (BCIs) use electroencephalography (EEG) [1] or other electrophysiological measures of brain activity like magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI) and electrooculography (ECoG) to establish a nonmuscular channel for sending messages and commands to external devices. Motor imagery (MI)-based BCI uses the information correlated with amplitude modulations of sensory motor rhythms (SMR) to reflect the motor intention of a subject [2]. The modulation of SMR generates contra lateral preponderant event-related desynchronization (ERD) and synchronization (ERS) which are short lasting amplitude attenuation and enhancements in the rhythmic components of EEG [3]. Currently, ERD-/ERS-based BCIs have received a lot of attentions due to their potential application in motor rehabilitation and assisting for the motor function impaired patients [4, 5]. However, there still remain some challenges and barriers to use this technology easily and effectively for the intended beneficiaries. One of them is that ERD/ERS analysis for different subjects has proven to be complex, as ERD/ERS occurs in different parts of the cortex, at different frequencies, and during different time intervals due to individual distinctions in physiology, anatomy and brain state, which leads to difficulty when extracting features for classification [6].

Feature extraction plays an important role for the performance of ERD-/ERS-based BCI, and common spatial pattern (CSP) is one of the most popular methods that have proven to be very useful in extracting the spatial information of EEG activity [7]. CSP aims to find the optimal spatial filters onto which the difference of projected scatters between two EEGs is maximized. However, some preprocessing steps like the band-pass filtering and time

interval selection have great effect on final CSP feature extraction [8]. As described above, the reactive frequency band and time interval are always user specific, and spatial adaptation of EEG-based BCI system was first investigated in [9]. It has been shown that using subject-specific parameters such as discriminable frequency bands, temporal segments of maximum separability and channels with high discriminative powers can improve the performance of a BCI system [3].

In order to accurately capture the band-power changes resulting from ERD/ERS, most of the CSP refinements are focused on the proper selection of frequency band, which is typically subject specific and can hardly be determined manually [1, 10, 11]. Besides the selection of discriminable frequency bands, automated channel selection is also essential to enhance the performance of CSP feature extraction and subsequent classification by removing task-irrelevant and redundant channels [12]. Moreover, the optimal observation period in MI classification depends on subjects and some attempts to determine the optimal temporal segment automatically have been proposed [13–15]. All these mentioned works have shown effectiveness in their fields. However, most of these methods above only extract spatial, frequency, frequency–temporal or spatial–temporal information of EEG, respectively; they do not work on all three of the spatial–frequency–temporal domains, simply expressing MI features in one (or two) domain(s) while disregarding the other(s) will result in loss of information that may contribute to more accurate MI classification.

Recently, some approaches that engage comprehensive considerations over the spatial–frequency–temporal features have been reported. Wang et al. [16, 17] proposed two approaches for motor imagery classification that utilized information from time, frequency, and spatial domains. In literature [18], the time and frequency axes were first divided into basic grids and then merged together based on task-discriminant maximization. In each adapted frequency–time grid, a respective selection was conducted to pick channels with high discriminative information. In order to extract discriminative features as much as possible, Higashi and Tanaka [19] proposed a method called common spatio-time–frequency patterns (CSTFP) for finding the time windows as well as the coefficients of the band-pass filters and spatial weights. In another technique proposed by Ince et al. [3], the most discriminative time segments were first found by implementing a merge/divide strategy along the time axis, and a frequency domain clustering was then used to select the most discriminative frequency bands. As we can see, in most of the methods above, frequency or temporal decomposition is implemented. Since the bandwidth of a frequency bin or the length of a time segment is usually

constant, although some merge strategies are used, a precise optimal frequency band or time interval is still hard to determine. Moreover, the frequency band selection is still independent of the time interval selection in [3]. Obviously, the pre-decided time segments would affect the detection of optimal frequency parameter to a large extent. As mentioned above, ERD and ERS are frequency and time localized, so the most discriminative information is always determined by the combined effects of both time interval and frequency band. The sequential procedure in above works may fail to find the best combination of frequency–time parameters for CSP feature extraction and subsequent classification, and the mere optimization of time segments without the simultaneous consideration of frequency band in the first step may lead to a local optimum.

Focused on the proper selection of optimal channels, frequency band and time interval for different people without any prior knowledge, we manage to enhance MI EEG classification accuracy with two coordinated methods: spatial optimization by channel selection, and the simultaneous searching of user-specific frequency–time parameters that locate the best discriminative signal components. The channel selection method will select the most discriminative EEG channels which are highly subject dependent, thus removing redundant channels and reducing the computational load for the CSP algorithm. In essence, the ultimate goal for searching user-specific frequency–time parameters is convergent to the optimization problem that how to find the global optimal combination of frequency band and time interval for CSP features extraction and classification. Obviously, there are infinite combinations for this issue; an exhaustive manual search is highly time-consuming and unreasonable. Artificial bee colony (ABC) algorithm is inspired by foraging behavior of bee and proposed by Karaboga and Basturk [20]. ABC algorithm has been successfully applied to various areas and proved higher efficiency than some other optimization algorithms [21]; however, it has never been used in this field. The original ABC algorithm is developed for solving multivariable function optimization problem, while in our work the selection of frequency–time parameters is precisely a multivariable optimization problem, so we carefully design a frequency–temporal parameters optimization strategy and integrate it into different phases of ABC algorithm. To our best knowledge, we are the first to examine the efficiency of applying ABC concept to frequency–temporal feature optimization for MI EEG classification.

This paper is organized as follows: In the next section, we present the data description and a detailed explanation of our method. Section 3 describes experimental results and discussion. Finally, conclusions are given in Sect. 4.

## 2 Materials and methods

### 2.1 Data description

In this study, three EEG datasets are used for evaluation; they are obtained from two publicly available sources, which are the BCI Competition III [22] and BCI Competition IV [30]. The detailed descriptions of three datasets are presented in this section.

#### 2.1.1 BCI Competition III Dataset IIIa

In this dataset, EEG signals were recorded from three subjects (“K3B,” “K6B,” “L1B”) with a 64-channel amplifier; however, EEG data of only 60 channels were provided for the competition. The EEG data were sampled at 250 Hz and filtered between 1 and 50 Hz. The subjects sat on a comfortable chair and had to perform left-/right-hand, tongue or foot movement imagination according to the cue on the screen. Since the classification method is developed for binary classes in this paper, only the samples of left hand and right hand are used. For each trial, the first 2 s were quite, at  $t = 2$  s an acoustic stimulus indicated the beginning of the trial and a cross was displayed; then from  $t = 3$  s, an arrow appeared on the fixation cross indicating the imagined movement to be executed until the cross disappeared at  $t = 7$  s. There are 180 trails in total for “K3B,” 90 trials per task. There are 120 trails in total for “K6B” and “L1B,” 60 trials per task. The numbers of trials for training and testing of each subject are as following: subjects (left hand, right hand) “K3B” (45, 45), “K6B” (30, 30) and “L1B” (30, 30) are for training and the remaining are for evaluation.

#### 2.1.2 BCI Competition III Dataset IVa

In this dataset, EEG signals were recorded from five healthy subjects (“aa”, “al”, “av”, “aw”, “ay”) through 118 electrodes. In each trial, a visual cue was shown for 3.5 s, during which three kinds of motor imageries were performed, that is, left hand, right hand and right foot. The motor imageries of right hand and right foot were provided for the competition. For this dataset, there were 100 and 1000 Hz versions. In our study, the 1000-Hz version was chosen and it was band-pass filtered between 0.05 and 200 Hz. There are 280 trials in total, 140 trials per task for each subject. The numbers of trials for training and testing vary across subjects as follows: subjects (left hand, right foot) “aa” (80, 88), “al” (112, 112), “av” (42, 42), “aw” (30, 26) and “ay” (18, 10) are for training and the remaining are for evaluation.

#### 2.1.3 BCI Competition IV Dataset IIa

In this dataset, nine subjects participated in the collection of EEG signals and EEG signals were recorded from 22 electrodes [30]. Subjects were instructed to perform four types of MI: left hand, right hand, both feet and tongue. Two sessions on different days were recorded for each subject. Only one session contained the class labels for all trials, whereas the other one session was used for evaluation. Each session was comprised of six runs and one run consisted of 48 trials (12 for each class). Therefore, one session consisted of 288 trials totally. At the beginning of each trial ( $t = 0$  s), a fixation cross appeared on the black screen. After two seconds ( $t = 2$  s), a cue in the form of an arrow pointing either to the left, right, down or up (corresponding to one of the four classes left hand, right hand, foot and tongue) appeared and stayed on the screen for 1.25 s. Subjects were instructed to perform motor imagery tasks according to the cue until the fixation cross disappeared at  $t = 6$  s. The EEG signals were band-pass filtered between 0.5 and 100 Hz and sampled at 250 Hz.

### 2.2 Methods

The proposed EEG signal analysis system includes both training and testing phases. Training is performed in several steps to obtain the necessary information for the corresponding testing, such as optimal channels, optimal frequency band, optimal time interval and CSP filters. These steps contain data preprocessing, channel selection, ABC-based frequency–temporal parameters selection and feature extraction. Then, the information obtained in the training phase is applied for testing directly. The same data preprocessing method is implemented in both training and testing phases.

#### 2.2.1 Data preprocessing

Prior to temporal filtering, the continuous experimental data are intercepted into single-trial data at first, and common average reference (CAR) [18] is adopted to re-reference them. For all three datasets, considering the usual frequency band of ERD and ERS, a fifth-order Butterworth band-pass filter is used to remove noises over 40 Hz and slow baseline signal under 5 Hz. The sampling rate of BCI Competition III Dataset IVa is 1000 Hz, in order to reduce the computational cost of subsequent processing without omitting important time series information; this dataset is down sampled to 250 Hz.

### 2.2.2 Channel selection

In order to achieve decent system performance, channel selection method is usually required to remove irrelevant and redundant channels [12]. In addition, the computational burden of subsequent time interval and frequency band selection can be reduced significantly by channel reduction. In this step, the fisher's linear discriminant criteria (FDC) value between two classes is calculated to quantify the discriminative power of the channel. Since the ERD/ERS phenomenon is time localized, temporal segmentation is done at first, it uses rectangular time windows, length of each  $t$ -segment is set to 250 and 100 data points for BCI Competition III Dataset IIIa (250 Hz  $\times$  8 s) and BCI Competition III Dataset IVa (250 Hz  $\times$  3.5 s) empirically, and the neighboring  $t$ -segments are 50% overlapping. For each channel, log-power is computed as the feature of each segment, as  $p_{ch,t}$  in (1), where  $x_{ch,t}$  is signal segment of  $t$ -segment  $t$  of channel  $ch$  and  $\text{var}()$  denotes variance of the variable.

$$p_{ch,t} = \log(\text{var}(x_{ch,t})) \quad (1)$$

Then,  $p_{ch,t}$  of all trials are grouped according to MI classes, two sets of  $p_{ch,t}$  of all trials in two classes are defined as  $p_{ch,t}^1$  and  $p_{ch,t}^2$ , respectively. FDC value between two classes is calculated in (2), where  $m_1$  and  $m_2$  are means of  $p_{ch,t}$  of two MI classes, and  $\varphi_{ch,t}$  is the FDC value of  $t$ -segment  $t$  of channel  $ch$ .

$$\varphi_{ch,t} = \frac{(m_1 - m_2)^2}{\text{var}(p_{ch,t}^1) + \text{var}(p_{ch,t}^2)} \quad (2)$$

After that, for each  $ch$ , the maximum of FDC values of all  $t$ -segments is served as the FDC value of this channel. Finally, channels with high FDC values are picked out. In this study, we sort the FDC values in descending order and the first  $K$  corresponding channels are selected. Note that EEG signals are recorded from only 22 electrodes in BCI Competition IV Dataset IIa, and too few channels may not include enough information for high performance; hence, channel selection is not implemented for this dataset.

### 2.2.3 Feature extraction and classification

In the procedure of ABC-based frequency–temporal parameters selection, average classification accuracy (CA) of cross-validation (CV) will be served as the objective measure of performance; therefore, we describe the details of feature extraction and classification before automatic frequency band and time interval selection.

CSP algorithm is selected for feature extraction due to its good performance in discriminating two classes of EEG signals. We assume that we have two classes of MI EEG signals, each class includes some trials in which each single trial is represented as a  $C \times Q$  matrix  $\mathbf{X}$ , where  $C$  is

the number of EEG channels and  $Q$  is the number of sampled time points per channel. For each class, the average covariance matrix can be obtained as follows:

$$\Psi_i = \frac{1}{N_i} \sum_{j \in N_i} \frac{\mathbf{X}_j \mathbf{X}_j^T}{\text{trace}(\mathbf{X}_j \mathbf{X}_j^T)}, \quad i \in \{1, 2\} \quad (3)$$

where  $N_i$  is the number of trials corresponding to  $i$ th class,  $T$  denotes the transpose operator, and  $\text{trace}(\mathbf{X}\mathbf{X}^T)$  gives the sum of the diagonal elements of  $\mathbf{X}\mathbf{X}^T$ . The CSP algorithm projects  $\mathbf{X}$  to spatially filtered  $\mathbf{Z}$  as

$$\mathbf{Z} = \mathbf{W}\mathbf{X} \quad (4)$$

where the rows of the projection matrix  $\mathbf{W}$  are the spatial filters.  $\mathbf{W}$  is generally computed by solving the following eigenvalue decomposition problem.

$$\Psi_1 \mathbf{W} = \Psi_2 \mathbf{W} \mathbf{\Lambda} \quad (5)$$

where  $\mathbf{\Lambda}$  is the diagonal matrix that contains the eigenvalues of  $\Psi_2^{-1} \Psi_1$ . Since the eigenvalues in  $\mathbf{\Lambda}$  indicate the ratio of the variances under two conditions, the first and the last  $m$  rows of  $\mathbf{W}$ , corresponding to the  $m$  largest and  $m$  smallest eigenvalues, are generally extracted as the most discriminative filters. Afterward, the variances of the projected signals are calculated as the features.

It should be noted that, in the process of optimal frequency–temporal parameter searching based on ABC algorithm, the average classification accuracy of cross-validation is served as the objective measure of performance and we have to run the whole system to compute classification accuracies for many iterations. In order to reduce the computing burden,  $K$  is set to 16 empirically and we extract eigenvectors corresponding to the largest and smallest eigenvalues to form the spatial filter. Afterward, the variances of the projected signals are calculated directly to form a two-dimensional feature vector. It is widely known that the first element of the feature vector is larger than the second element for one class and smaller than the second element for the other class. In this study, a simple comparison between these two elements of the feature vector is done to discriminate the classes in the procedure of frequency–temporal parameter searching.

While in the testing phase, we extract the eigenvectors corresponding to the largest and smallest  $m$  eigenvalues to form the filter, the choice for  $m$  depends on the data. For both BCI Competition III Dataset IIIa and BCI Competition III Dataset IVa, we set  $m$  to 1 and a simple sign function classifier is utilized to discriminate the classes as:

$$\begin{cases} \text{if } \text{sgn}(\text{var}(\mathbf{Z}(1,:)) - \text{var}(\mathbf{Z}(2,:))) = 1, & \mathbf{X} \in R_1 \\ \text{if } \text{sgn}(\text{var}(\mathbf{Z}(1,:)) - \text{var}(\mathbf{Z}(2,:))) = -1, & \mathbf{X} \in R_2 \end{cases} \quad (6)$$

where  $R_1$  and  $R_2$  denote class one and class two, respectively. For BCI Competition IV Dataset IIa, in order to

compare with the method proposed in [14] under the same condition, we also set  $m$  to 2, LDA [23] classifier is used for this dataset.

#### 2.2.4 Artificial bee colony algorithm

Among evolutionary approaches, the ABC algorithm has been found to perform better or at a same level than others like genetic algorithm (GA), differential evolution (DE), particle swarm optimization (PSO) or evolution strategies (ES) on a large set of unconstrained test functions by Karaboga and Akay [24]. The ABC algorithm simulates the bee colony behavior in food search activity and contains three groups of artificial bees: employed bees, onlooker bees and scout bees. The search procedures are carried out by those bees iteratively and can be summarized as follows:

At the initial phase of the algorithm, it generates a group of food positions randomly. A food position represents a possible solution, and each position has several parameters to be optimized. The quality of a food position, i.e., the amount of nectar, is evaluated by its fitness value. Then, the algorithm calculates fitness values of all these food positions.

At the employed bees phase, each employed bee is associated with a food position at first. Then, the employed bees search new positions using their current food positions. For each employed bee, the fitness value of the new food position is calculated and compared with the fitness value of the old one, and the food position with the larger fitness value is recorded.

At the onlooker bees phase, the algorithm calculates the selection probabilities of all the food positions by their fitness values. The food position with the largest probability is selected by onlooker bees and they make modifications on this food position. Note that, for each onlooker bee, the fitness value of the new food position is also compared with the fitness value of the old one, and the better food position is recorded.

At the scout bees phase, when an employed bee cannot improve a food position within a certain number of trials, it abandons its current food position and becomes a scout bee which will find a random food position.

For all the three kinds of bees, the food position with the largest fitness value is recorded as the temporary best solution in each cycle. The best food position is iteratively updated for certain iterations. The general structure of ABC algorithm is shown in Fig. 1.

#### 2.2.5 An ABC-based frequency–temporal parameter optimization scheme

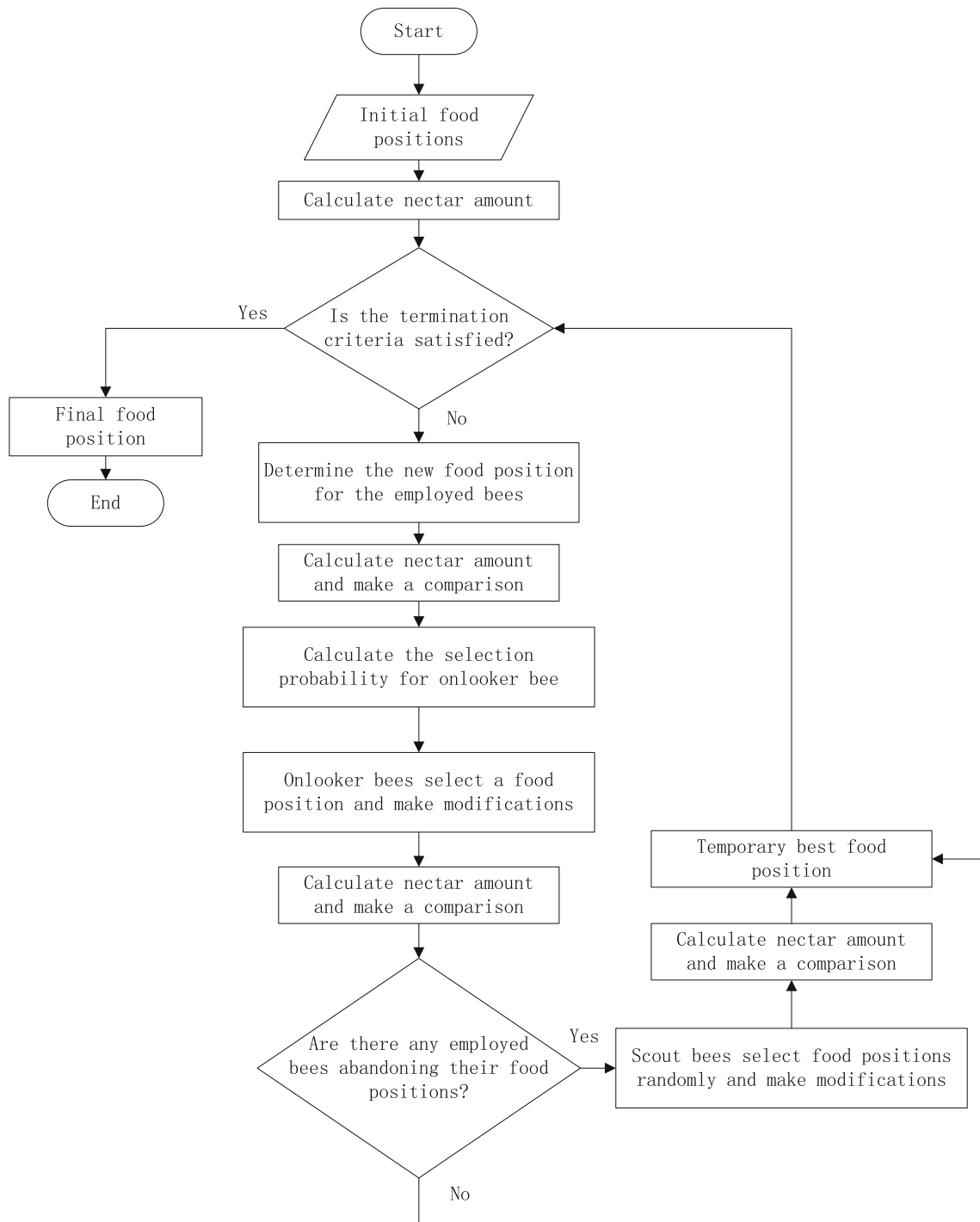
In the BCI application, the labels of the training set are known, while the labels of the testing set are unknown. In this section, based on the training set, we apply the concept

of ABC algorithm to find the optimal frequency band  $\{f_{\text{start}}, f_{\text{start}} + f_{\text{width}}\}$  and time interval  $\{t_{\text{start}}, t_{\text{start}} + t_{\text{length}}\}$  simultaneously for robust CSP feature extraction, with  $f_{\text{start}}$  denoting the low bound of band-pass filter,  $f_{\text{width}}$  denoting the frequency band width of band-pass filter,  $t_{\text{start}}$  denoting the low bound of time interval and  $t_{\text{length}}$  denoting the length of time interval. Obviously, the dimension of a solution (food position) is four, with each component representing  $f_{\text{start}}$ ,  $f_{\text{width}}$ ,  $t_{\text{start}}$  and  $t_{\text{length}}$ . For all the three datasets, the searching ranges of food position components are listed in Table 1. For all the three datasets, if  $f_{\text{start}} + f_{\text{width}} \geq 40$ , then  $f_{\text{width}} = 40 - f_{\text{start}}$ . For BCI Competition III Dataset IIIa, if  $t_{\text{start}} + t_{\text{length}} \geq 8 \times 250$ , then  $t_{\text{length}} = 8 \times 250 - t_{\text{start}}$ . For BCI Competition III Dataset IVa, if  $t_{\text{start}} + t_{\text{length}} \geq 3.5 \times 250$ , then  $t_{\text{length}} = 3.5 \times 250 - t_{\text{start}}$ . It should be noted that the time segment of 0–4 s of EEG after the onset of the visual cue was used for BCI Competition IV Dataset IIa in literature [14]; we also use this time segment for analysis. Therefore, for this dataset if  $t_{\text{start}} + t_{\text{length}} \geq 4 \times 250$ , then  $t_{\text{length}} = 4 \times 250 - t_{\text{start}}$ .

In order to evaluate the quality of a food position, the objective function should be defined to calculate the fitness value. In our study, two kinds of objective functions are used. One is the average classification accuracy of cross-validation. In this study, fivefold CV is implemented on the training set of BCI Competition III Dataset IIIa and BCI Competition IV Dataset IIa. While for BCI Competition III Dataset IVa, since some subjects have a small amount of training data, i.e., subject “ay” has only 28 training samples, 5 × 2-fold CV is used. The other objective function is  $R$ -squared ( $R^2$ ) [13] between two tasks.

According to the principle of ABC algorithm, our algorithm consists of four phases, namely the initialization phase, the employed bees phase, the onlooker bees phase and the scout bees phase. After the initialization phase, the other three phases are invoked iteratively for  $D$  cycles, and the best food position is obtained in the end. The ABC-based approach for optimal frequency band and time interval selection can be denoted as below.

Step 1. Initialization phase: In this phase, we first create a population of initial food positions according to the colony size  $S$ ; a food position  $v_i (i = 1, 2, \dots, S)$  represents a feasible solution for the optimization problem, i.e., an initial combination of frequency–time parameters. Each position  $v_i$  has four parameters  $v_{i,j}, j = 1, 2, 3, 4$  to be optimized, where the first component defines  $f_{\text{start}}$ , the second component defines  $f_{\text{width}}$ , the third component defines  $t_{\text{start}}$ , and the fourth component defines  $t_{\text{length}}$ . Each component is initialized with a random integer value within its searching range as described in Table 1. Then, the quality of each food position is evaluated; for the  $i$ th food



**Fig. 1** Flowchart of ABC algorithm

**Table 1** Searching ranges of food position components for all the three datasets

Dataset	$f_{start}$ (Hz)	$f_{width}$ (Hz)	$t_{start}$ (sample points)	$t_{length}$ (sample points)
III–IIIa	5–30	5–30	$3 \times 250$ – $5 \times 250$	$1 \times 250$ – $5 \times 250$
III–IVa	5–30	5–30	$0 \times 250$ – $3 \times 250$	$1 \times 250$ – $3 \times 250$
IV–IIa	5–30	5–30	$0 \times 250$ – $3 \times 250$	$1 \times 250$ – $3 \times 250$

position, we feed the four parameters  $[v_{i,1}, v_{i,2}, v_{i,3}, v_{i,4}]$  into the fitness function to obtain the fitness value  $\eta_i$ . After the initialization phase, totally  $S$  initial combinations of frequency–time parameters are obtained, and the best one with the largest fitness value  $\eta_{\text{best}}$  is recorded. Then, all the  $S$  initial food positions will be iteratively updated by employed bees, onlooker bees and scout bees for  $D$  cycles to find the optimal frequency band and time interval for CSP feature extraction and classification; each phase is described as follows.

**Step 2. Employed bees phase:** The number of employed bees is  $S/2$ ; in our work, they are associated with the first  $S/2$  food positions in the initialization phase. Each employed bee searches a new food position based on its current food position; for the  $i$ th employed bee, the new food position is obtained by

$$\omega_{i,j} = v_{i,j} + \phi(v_{i,j} - v_{k,j}) \tag{7}$$

$$\omega_{i,g} = v_{i,g}, \quad g \neq j \tag{8}$$

where  $k$  is a random food position index ( $k \neq i$ ) and ( $k \in [1, S/2]$ ),  $j$  is a parameter chosen randomly and  $\phi(v_{i,j} - v_{k,j})$  returns a random integer value within  $[-|v_{i,j} - v_{k,j}|, |v_{i,j} - v_{k,j}|]$ . Then, the fitness value of  $\omega_i$  is calculated, and greedy selection is applied: If  $\omega_i$  has better fitness value than  $v_i$ , the food position  $v_i$  is replaced by  $\omega_i$  and its trial counter  $\tau$  is reset to zero, otherwise  $v_i$  is retained and  $\tau$  increases by one.

**Step 3. Onlooker bees phase:** An onlooker bee selects a food position depending of a probability  $p_i$  as:

$$p_i = \frac{\eta_i}{\sum_{i=1}^S \eta_i} \tag{9}$$

where  $\eta_i$  is the fitness value of the  $i$ th food position. Once that the onlooker bees have selected the food position with largest value of  $p_i$ , they generate new food positions using (7). However, for onlooker bees,  $k \neq i$  and  $k \in [S/2, S]$ . The new food positions are evaluated, and the same greedy selections are applied.

**Step 4. Scout bees phase:** The purpose of this phase is to avoid being trapped into local optimum that usually occurs after running many iterations. When a food position has not improved within a number of trials as:

$$\tau > \textit{Limit} \tag{10}$$

where *Limit* is the maximum limit search number, this food position is replaced by a random food position.

**Step 5.** After each cycle, the food position with the best fitness value so far is recorded as the temporary best food position. If the number of iterations reaches  $D$ , the algorithm stops and the temporary best position at this moment becomes the final optimal solution of the system.

Three control parameters are included in the proposed algorithm above. The colony size  $S$ , the predefined

maximum limit search number *Limit* and the maximum iteration number  $D$  are the parameters. It should be noted that  $S$  is closely related to  $D$ . If  $S$  is big, it is more likely that the best solution can be found in each iteration and the ABC algorithm can converge to global optimal solution within a small number of iterations, so  $D$  can be relatively small, while the computation time of each iteration will be relatively long. The purpose of setting *Limit* is to avoid being trapped into local optimum. If *Limit* is too small, then the searching ability of the employed bee will be weakened, while a big setting of *Limit* means that the employed bee may be trapped into a certain local optimum for too many iterations. In our practice, when  $S$  is set to 50, the number of iteration epochs for convergence is less than 15 for each subject in all three datasets, so we set  $D$  to 15, *Limit* is set to 5 according to literature [25].

### 3 Results and discussion

In this section, we present the feasibility and effectiveness of our proposed method by comparing the classification performance with other competing methods in the literature and analyzing the experimental results on three public datasets.

#### 3.1 Classification performance comparisons with other methods

##### 3.1.1 Results of BCI Competition III Dataset IVa

For this dataset, we firstly used the same training set provided in BCI Competition to perform feature optimization and the performance of our method was evaluated by measuring the ratio of trials correctly classified to the total number of trials in testing set for each subject. Under this circumstance, the training sets of subject “aa”, “al”, “av”, “aw” and “ay” contain 168, 224, 84, 56 and 28 trials, respectively. The challenge of BCI Competition III dataset IVa is to validate the efficiency of algorithms to deal with various sizes of training dataset. In Table 2, we present the effectiveness of our method by comparing its classification results with those of other recently reported methods, namely the standard CSP [8], common spatial spectral pattern (CSSP) [26], filter bank CSP (FBCSP) [27], discriminative FBCSP (DFBCSP) [11], optimal spatio-spectral filter network (OSSFN) [28], Bayesian spatio-spectral filter optimization (BSSFO) [29], channel selection with GA [12], space–time–frequency adaption [3], channel configuration with spectrally weighted common spatial pattern (SWCSP) [31], spatio-spectral-cognitive fuzzy interface system (SS-CFIS) [32] and optimized sparse spatio-spectral filtering (OSSSF) [33]. For the OSSFN, two

**Table 2** Comparison of classification accuracies (%) by standard competition procedure, *SD* standard deviation. (BCI Competition III Dataset IVa)

Methods	aa	al	av	aw	ay	Mean	SD
CSP [8]	66.96	89.29	52.55	47.77	52.38	61.79	16.98
CSSP [26]	79.46	92.86	52.55	91.52	51.59	73.60	20.33
FBCSP [27]	69.64	80.36	47.96	55.36	48.41	60.35	14.21
DFBCSP [11]	69.64	82.14	54.08	50.89	48.41	61.03	14.41
OSSFNwFBCSP [28]	75.00	83.93	53.06	74.11	48.81	66.98	15.22
OSSFNwDFBCSP [28]	75.00	83.93	52.05	74.11	47.22	66.46	15.93
BSSFO [29]	79.46	94.64	57.65	91.96	53.57	75.46	19.06
Channel selection with GA [12]	74.44	98.47	70.78	90.52	83.44	83.53	11.37
s-t-f adaption [3]	88.40	91.10	82.10	81.20	88.90	86.34	4.41
SWCSP [31]	83.0	100	73.5	82.1	91.5	86.02	10.08
SS-CFIS [32]	82.14	100	63.27	83.04	60.32	77.75	16.24
OSSSF [33]	77.68	100	77.04	94.28	92.14	88.23	10.33
Our method	85.71	98.21	76.57	95.09	91.67	89.45	8.56

approaches are considered, each of which utilizes FBCSP (OSSFNwFBCSP) or DFBCSP (OSSFNwDFBCSP) in composing a filter band. Note that all the methods in Table 2 are evaluated by standard competition procedure, except for OSSSF in which the training trials are increased to 140 for subjects “aw” and “ay.” Table 2 shows that, for our proposed method, the final classification accuracies of subject “aa”, “al”, “av”, “aw” and “ay” are 85.71% ( $K = 16$ ), 98.21% ( $K = 16$ ), 76.57% ( $K = 16$ ), 95.05% ( $K = 32$ ) and 91.67% ( $K = 16$ ), respectively. The results of the first seven algorithms are presented in [29]. We also note that our algorithm provides a mean classification accuracy of 89.45%, the average accuracy improvements achieved by the proposed method are 27.66, 15.85, 29.1, 28.42, 22.47, 22.99, 13.99, 5.92, 3.11, 3.43, 11.70 and 1.22% in comparison with CSP, CSSP, FBCSP, DCSP, OSSFNwFBCSP, OSSFNwDCSP, OSSFN, channel selection with GA, space–time–frequency adaption, SWCSP, SS-CFIS and OSSSF, respectively. The result indicates that an overall consideration on space–time–frequency domains can improve the classification ability effectively, and our approach outperforms other competing methods which are under the same condition significantly. According to BCI Competition 2005 final results, only the top one winner shows better performance than our method. Note that the high performance of the winner is achieved by applying additional techniques like boosting or online classifier training to tackle the problem of the small training sets. However, these techniques are out of the scope of this paper. All these results show that the proposed algorithm can generalize well with little amount of training data.

We notice that, instead of using the exact splitting of this dataset in BCI Competition, some other researchers performed cross-validation for evaluation. To compare with these methods under the similar condition, we used

$10 \times 10$  cross-validation to evaluate the ability of our proposed method. The results of  $10 \times 10$  cross-validation are obtained by repeating the tenfold cross-validation [7] for 10 times. In this study, the optimal channels and frequency–temporal parameters are obtained from the first validation procedure and used for other validations to simplify computation. Table 3 reports the comparison results of the classification accuracy rates for our proposed method and the eleven existing algorithms for BCI Competition III Dataset IVa. As shown in Table 3, the highest mean for all five subjects is obtained by the optimal allocation (OA)-based approach [7], which is 96.62%. Further looking at the performance comparison in Table 3, it is noted that our proposed is ranked sixth among all the twelve algorithms in terms of the average accuracy. Five existing methods perform better than our method; these methods are as follows: (1) OA-based approach proposed by Siuly and Li [7]; (2) iterative spatio-spectral patterns learning (ISSPL) proposed by Wu et al. [34]; (3) channel configuration with SWCSP proposed by Meng et al. [31]; (4) DFBCSP proposed by Thomas et al. [11]; and (5) cross-correlation and logistic regression (CC–LR) with threefold cross-validation proposed by Siuly and Wen [35]. Note that, for OA-based approach and CC–LR algorithm, CSP technique is not employed for feature extraction. Hence, there still remain three CSP styled approaches outperforming our method. For ISSPL, the reason might be that the temporal filter, channels, time windows and the number of spatial filter are tuned according to the winning entry of the dataset [34]. For channel configuration with SWCSP, like the authors said, the reason why high classification accuracy is obtained might be that they find the most proper channel configuration by comparing twelve fixed channel configurations [31]. Overall, our method still performs better than most of the other CSP styled approaches.



**Table 3** Comparison of average classification accuracies (standard deviation) (%) obtained by cross-validation

Methods	aa	al	av	aw	ay	Average
OA-based approach [7]	99.17 (2.63)	98.29 (2.20)	100 (0.0)	91.13 (8.03)	94.51 (9.04)	96.62 (3.72)
ISSPL [34]	93.57 (3.28)	100 (0.0)	79.29 (7.10)	99.64 (1.13)	98.57 (3.01)	94.21 (2.9)
Algorithm of [31]	94.2 (4.5)	99.2 (1.8)	78 (8.5)	97.7 (2.9)	95.6 (3.8)	93 (4.3)
DFBCSP [11]	90.2 (5.6)	98.7 (2.9)	77.8 (9.9)	97.9 (3.7)	94.2 (5.3)	91.8 (5.5)
CC–LR (threefold) [35]	96.57 (2.65)	82.9 (5.35)	100 (0.0)	96.6 (3.21)	82.9 (0.69)	91.79 (2.38)
Our method	91.2 (4.81)	97 (3.13)	76 (8.12)	94.3 (4.74)	95.3 (4.22)	90.76 (5)
CSP [31]	91.5 (5.4)	99.2 (1.8)	70.9 (8.2)	96.9 (2.8)	94.7 (3.8)	90.6 (4.4)
CC–LR (tenfold) [36]	95.93 (11.44)	91.2 (10.5)	88.34 (19.61)	90.08 (14.27)	85.92 (8.49)	90.29 (12.9)
FBCSP [11]	93.1 (5.8)	99 (2.4)	69 (14.2)	95.1 (8.9)	93.8 (9.7)	90.0 (8.2)
SBCSP (MC) [37]	89.3 (5.6)	98.6 (1.8)	70.4 (5.3)	95.7 (4.0)	95.7 (2.8)	89.9 (3.9)
SBCSP (RBE) [37]	90.8 (4.5)	97.8 (3.4)	69 (7.3)	95.8 (3.3)	95 (3.4)	89.7 (4.4)
SWCSP [31]	86.8 (6.3)	97.5 (2.9)	76.8 (9.4)	93.9 (3.8)	92.5 (3.9)	89.5 (5.3)

Methods are listed in descending order in terms of average classification accuracy (BCI Competition III Dataset IVa)

### 3.1.2 Results of BCI Competition IV Dataset IIa

The experiment of this dataset was performed in two parts. In the first part, ABC algorithm was performed on the training data (the first session) to determine the optimal frequency–temporal parameters. Note that there are in total 288 trials in one session and 72 trials per class. Since one session contains four types of MI: (1) left hand (LH); (2) right hand (RH); (3) feet (FT); and (4) tongue (TO), there are six pairs of MI tasks for binary classification, namely LH/RH, LH/FT, LH/TO, RH/FT, RH/TO and FT/TO. For each pair of MI task, the optimal frequency–temporal parameters are selected individually using ABC-based approach, and then CSP algorithm and classifier training is implemented according to these parameters. In the second part of the experiment, the session-to-session performance of the proposed method was computed on the evaluation data (the second session). Note that the multi-class classification result was obtained based

on the one-against-one [38] approach in this study. Since the kappa coefficient [14] was used as a performance measure in the BCI Competition IV, it was used in this part of the experiment to measure the maximum kappa value evaluated on the entire single-trial EEG from the onset of the fixation cross. The kappa value was computed from the onset of the fixation cross to the end of the cue for every point in time across all the trials of the evaluation data.

The performance of our method is compared with optimal spatial–temporal patterns (OSTP) [14], OSTP using the LDA classifier (denoted FLD), OSTP using the mutual information-based sequential forward selection (MISFS) algorithm (denoted SFS), OSTP using manual selection of time segment and frequency on the training data (denoted Manual), FBCSP [27], DFBCSP [11] and results of the 2nd, 3rd and 4th placed submissions for the competition.

Table 4 shows the results of the session-to-session transfer from the training data to the evaluation data for

**Table 4** Performance comparison of the proposed method with other typical approaches (including the winners of the competition)

Methods	Subjects									Average
	1	2	3	4	5	6	7	8	9	
OSTP	0.731	0.398	0.787	0.574	0.412	0.255	0.829	0.75	0.62	0.595
FBCSP	0.676	0.417	0.745	0.481	0.398	0.273	0.773	0.755	0.606	0.569
DFBCSP	0.736	0.375	0.718	0.329	0.245	0.366	0.727	0.778	0.685	0.551
SFS	0.583	0.306	0.718	0.491	0.440	0.259	0.769	0.718	0.671	0.550
Our method	0.6481	0.3657	0.6632	0.5046	0.3241	0.2963	0.7188	0.6354	0.6458	0.5336
2nd winner	0.69	0.34	0.71	0.44	0.16	0.21	0.66	0.73	0.69	0.52
FLD	0.634	0.324	0.653	0.560	0.306	0.153	0.769	0.532	0.514	0.494
Manual	0.528	0.319	0.662	0.435	0.190	0.282	0.560	0.708	0.657	0.483
3rd winner	0.38	0.18	0.48	0.33	0.07	0.14	0.29	0.49	0.44	0.31
4th winner	0.46	0.25	0.65	0.31	0.12	0.07	0.00	0.46	0.42	0.30

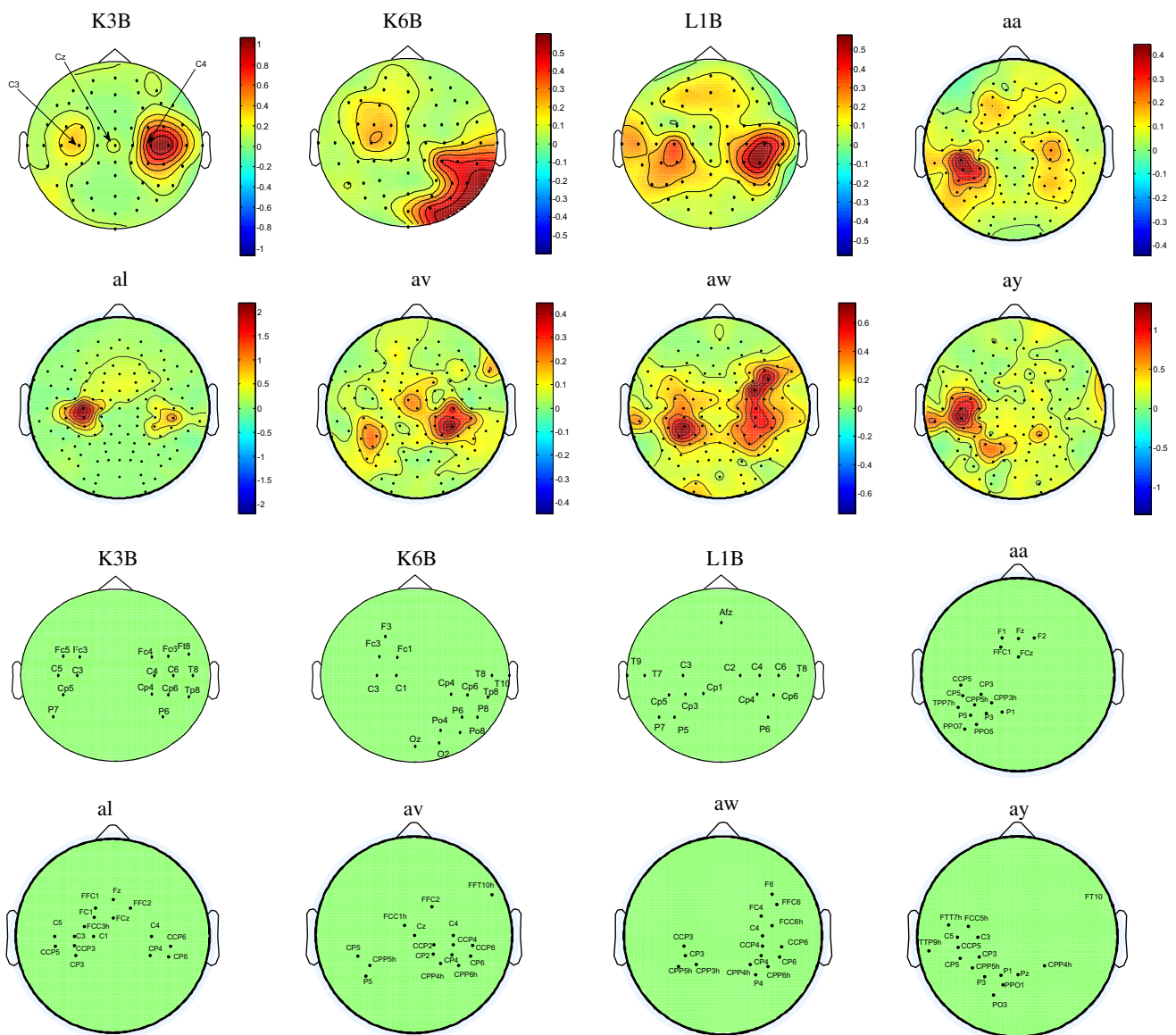
The numbers represent Cohen’s Kappa values, and methods are listed in descending order in terms of average Kappa value (BCI Competition IV Dataset IIa)

BCI Competition IV Dataset IIa. From Table 4, we observe that our method yields an average kappa value of 0.5336 and is ranked fifth among all the ten algorithms. Four existing methods perform better than our method; these methods are as follows: (1) OSTP; (2) FBCSP; (3) DFBCSP; (4) SFS. It should be noted that FBCSP algorithm is the top one winner of BCI Competition IV Dataset IIa, and OSTP, FBCSP, DFBCSP, SFS, FLD and Manual are proposed by the same research team. The results also show that our method yields better classification performance when compared to the 2nd, 3rd and 4th placed submissions for the competition. Note that our method outperforms FLD which also uses LDA classifier; hence, our results are still to be improved by designing more

sophisticated machine learning techniques for classification. However, the main aim of this paper is optimizing CSP features on spatial–frequency–temporal domains, so our forthcoming research issue would be focused on the classification step.

### 3.2 Analysis of channel selection

In order to understand the underlying reasons that spatial adaptation yields better results and to provide a physiological basis for discrimination, the channel selection method was used to calculate the discriminative power of each channel, and the topographical distributions are illustrated in Fig. 2 for each subject of both BCI Competition III Dataset IIIa and



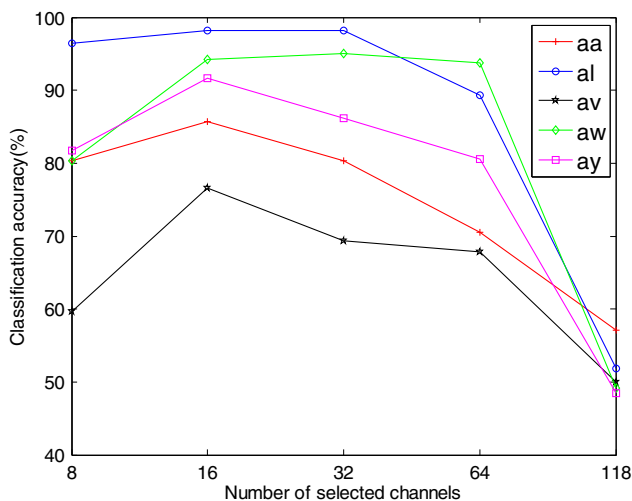
**Fig. 2** Topographical maps of channels' discriminative power distributions and selected channels of eight subjects in BCI Competition III: Dataset IIIa and Dataset IVa

IVa. As shown in Fig. 2, for most subjects the channels with high discriminative powers are located in neighboring area of C3, Cz and C4 locations, except for subject “K6B.” However, because of individual variability across different subjects, the spatial pattern of one subject is different from other subjects. According to the discriminative powers of channels, an optimal combination of channels with high discriminative powers was selected. Figure 2 also shows the topographies of selected channels of each subject, here  $K = 16$ . As serial processing is employed on signal of each channel, it is obvious that the computational burden is reduced when the number of channels decreases. Aiming to evaluate the effectiveness of channel selection, different

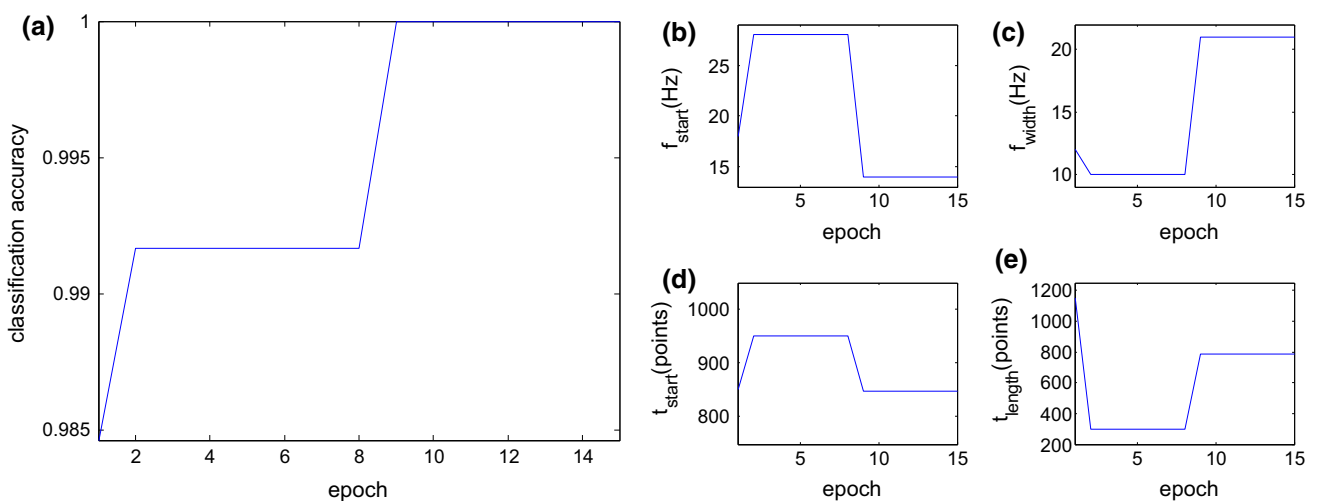
numbers of channels were selected for classification; we set  $K$  to 8, 16, 32, 64 and 118. Figure 3 presents the classification performance by standard competition procedure under different settings of  $K$  for each subject of BCI Competition III Dataset IVa. Figure 3 shows that the accuracies decrease dramatically when  $K$  is bigger than 32. The reason might be that the nonevent-related channels add ambiguous information to the dataset, thus resulting in separability attenuation of different classes. Therefore, channel selection can improve the separability effectively by eradicating a significant amount of redundant information.

### 3.3 Effects of different fitness functions

To evaluate the performance of two different fitness functions, both CA and  $R^2$  were used for BCI Competition III Dataset IIIa. Figure 4 presents the evolution procedures of classification accuracy and best food position for subject “K3B” when CA is used as the fitness function. Figure 4 (a) shows that CA value (fitness value) inclines to a constant after 9 epochs, note that the numbers of epochs for convergence are also less than 15 for other subjects in all three datasets according to the experimental results, which demonstrate that ABC has good convergence. Since CA and  $R^2$  are used for optimal frequency band and time interval selection, the resulted frequency band and time interval will be of generalization for feature extraction and classifier training. Table 5 shows the performance of two approaches on BCI Competition III Dataset IIIa using CA and  $R^2$  as the fitness functions of ABC. The obvious differences of the frequency band and time interval are given in Table 5 for all subjects when two different fitness functions are used. The results show that the mean test



**Fig. 3** Classification accuracies of subject “aa”, “al”, “av”, “aw” and “ay” by standard competition procedure when the number of selected channels ( $K$ ) is varied



**Fig. 4** ABC iteration procedure using CA as fitness function for subject “K3B”. **a** The fitness value evolution, **b** the evolution procedure of the low bound of band-pass filter  $f_{start}$ , **c** the evolution

procedure of the frequency band width of band-pass filter  $f_{width}$ , **d** the evolution procedure of the low bound of time interval  $t_{start}$ , **e** the evolution procedure of the length of time interval  $t_{length}$

**Table 5** Classification performance and optimal parameters selected by ABC-based approach using CA and  $R^2$  as the fitness functions

Subjects	Train accuracy (%)		Test accuracy (%)		Frequency band (Hz)		Time interval (sample points)		Computational time (s)	
	CA	$R^2$	CA	$R^2$	CA	$R^2$	CA	$R^2$	CA	$R^2$
K3B	100 ± 0	–	98.89	84.44	14–35	20–30	847–1633	800–1160	4700.1	1624.3
K6B	88.82 ± 7.39	–	61.67	58.33	19–34	19–28	925–1375	900–1325	3047.3	1057.7
L1B	97.42 ± 3.59	–	95	86.67	15–31	12–40	825–1325	1150–1775	3180.3	973.7
Mean	95.41 ± 3.66	–	85.187	76.48	N/A	N/A	N/A	N/A	3642.6	1218.6

Column “train accuracy” contains results of classification accuracies on the training set; in the case of CA, all accuracies are obtained by fivefold cross-validation, so results are presented in “mean ± Std” form, while in the case of  $R^2$ , train accuracy is not calculated. Column “Test accuracy” contains results of classification accuracies on the testing set. Column “Frequency band” and column “Time interval” contain results of optimal frequency band and time interval selected automatically by the proposed method. Column “Computational time” contains results of computational time for ABC iterations. (–) Not provided (BCI Competition III Dataset IIIa)

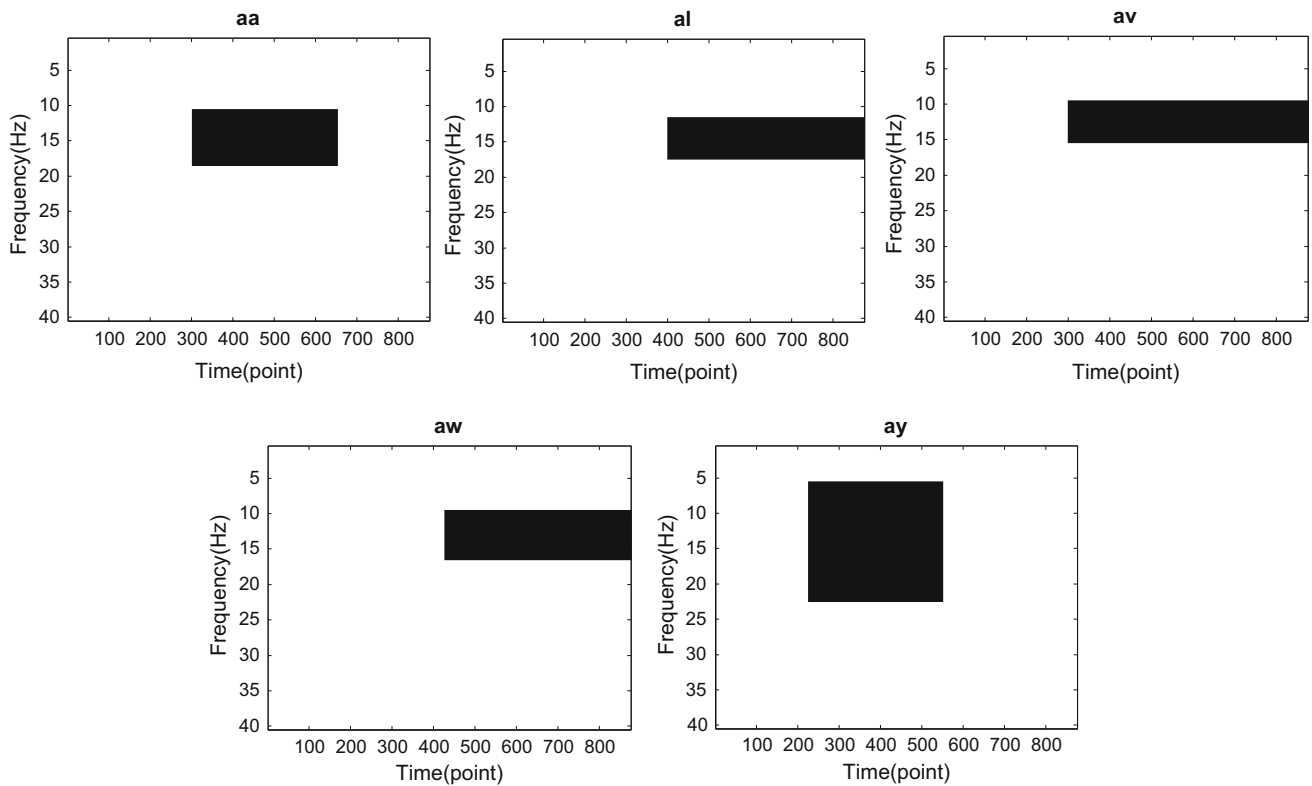
accuracy of CA approach is 85.187%, while in the case of  $R^2$ , it decreases to 76.48%. It is important to note here that the CA-based approach is bound to give better accuracy as it computes the fitness value based on CA, thus finding out the best possible frequency band and time interval for optimum classification accuracy. On the other hand, the mean computational time of CA approach for whole ABC iteration procedure is 3642.6 s, while the mean computational time of  $R^2$  approach is just 1218.6 s. The main reason is that the  $R^2$  approach computes the fitness value based on the correlation difference between different classes and hence reduces the computational burden by not running the whole system to compute classification accuracy. Additionally, a relatively long time is needed for training as we can see from the column “Computational time” of Table 5, but after training the optimal frequency band, time interval and CSP filters can be directly used in the testing phase, and therefore, the ABC-based approach is feasible for a real-time BCI application. Since higher classification accuracies are obtained when CA is used as fitness function, to better evaluate the classification performance of our approach and compare with other methods, only CA was used as fitness function for BCI Competition III Dataset IVa and BCI Competition IV Dataset IIa.

### 3.4 Analysis of frequency–temporal parameters optimization

Figure 5 shows the optimal frequency–temporal zones found by ABC-based approach using CA as the fitness value for five subjects of BCI Competition III Dataset IVa. We note that the optimal frequency bands of five subjects are all located in 8–30 Hz, which covers the mu and beta rhythms. However, the optimal frequency band of subject “ay” is broader than others. The main reason is that the small amount of training data prevents the ABC algorithm to find more precise result for subject “ay.” In addition, the optimal time segments for classification also differ among

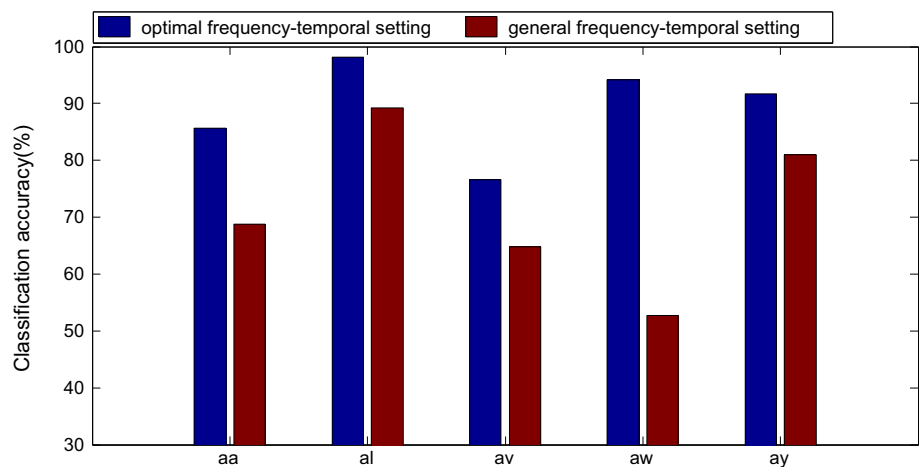
these subjects. To further assess the performance of frequency–temporal parameter optimization, we also used general frequency and time setting for comparison. The frequency band 8–30 Hz [1] [18] and time segment 0.5–2.5 s after the visual cue [11] which are generally used in MI EEG analysis were adopted. Figure 6 presents the comparison of classification accuracies obtained with optimal frequency–temporal setting and general frequency–temporal setting for BCI Competition III Dataset IVa. Figure 6 shows that, for all the subjects in BCI Competition III Dataset IVa, the results obtained with optimal frequency–time setting are much better. Although it is convenient to use general frequency or time setting for all subjects, the differences among subjects are neglected and the inclusion of extraneous signals degrades classification performance.

Figure 7 presents the optimal frequency–temporal zones of six pairs of MI tasks found by ABC-based approach for Subject 1 in BCI Competition IV Dataset IIa. Figure 7 clearly shows that though the optimal frequency bands for different pairs of MI tasks are mainly located in 8–30 Hz, they still vary a lot, e.g., the optimal frequency band for RH/FT is 10–21 Hz which is similar with the results in Fig. 5, while the optimal frequency band for LH/RH is 17–40 Hz which is similar with the results in Table 5. From our point of view, the discriminative EEG signal components locate in different frequency ranges for different pairs of MI tasks, e.g., mu rhythm (12–16 Hz) contains much discriminative information for RH/FT MI task, while beta rhythm (18–24 Hz) contains much discriminative information for LH/RH MI task. Further looking at the optimal time segments for different pairs of MI tasks in Fig. 7, we observe that though the optimal time segments for different pairs of MI tasks all start from almost 1 s (250 points) after the onset of the visual cue, the starting points still differ in a small range. Moreover, the lengths of optimal time segments are not equal for different pairs of MI tasks. We notice that nine predefined temporal band-pass filters as well as three time segments 0.5–2.5,



**Fig. 5** Optimal frequency–temporal zones (colored in *black*) found by ABC-based approach using CA as the fitness value for five subjects in BCI Competition III Dataset IVa

**Fig. 6** Comparison of classification accuracies obtained by standard competition procedure with optimal frequency–temporal setting and general frequency–temporal setting for BCI Competition III Dataset IVa

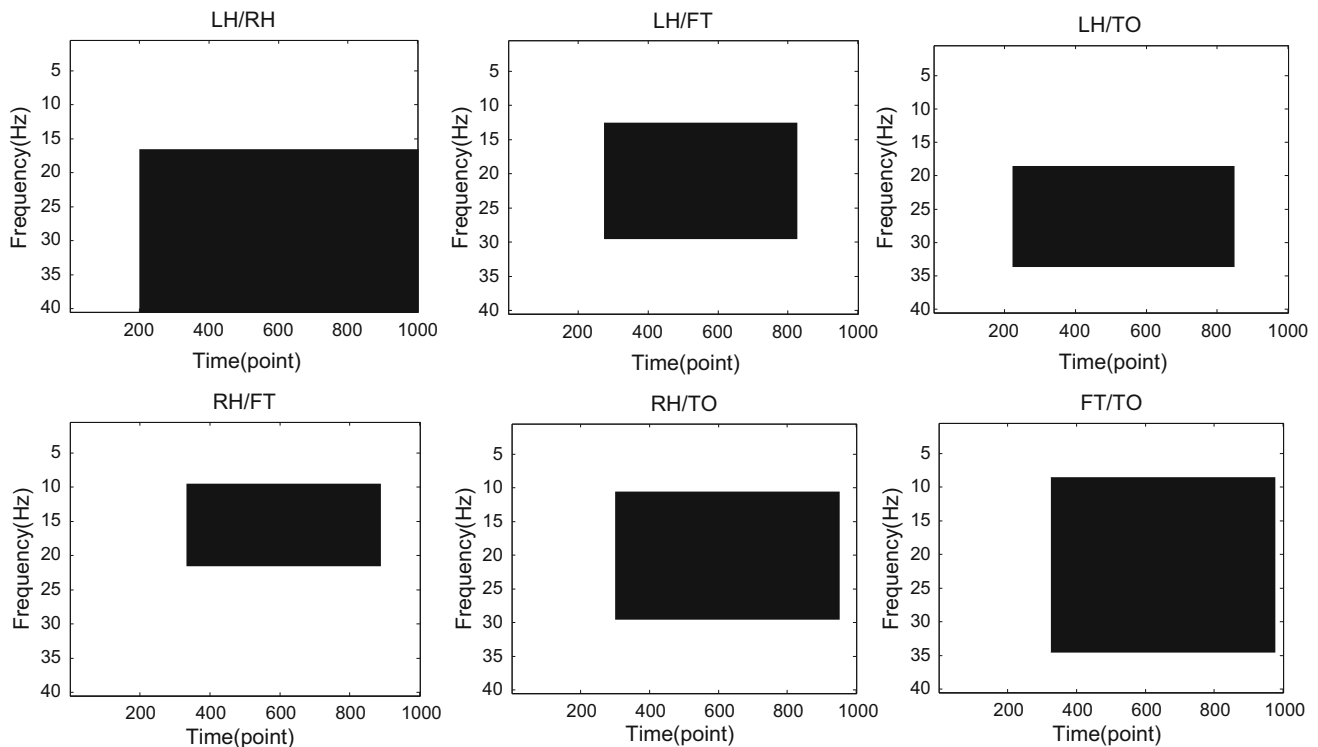


1.0–3.0 and 1.5–3.5 s are used in literature [14]; however, the optimal frequency bands and time segments in our study are determined automatically by ABC algorithm.

### 4 Conclusions

In this study, we proposed a spatial–frequency–temporal feature optimization method for motor imagery EEG pattern recognition. Channel selection method was used to

automatically select the channels with high discriminative powers. In addition, the concept of ABC algorithm was first introduced to find the global optimal combination of frequency band and time interval simultaneously without prior knowledge for CSP features extraction and classification. The classification accuracy of the algorithm, its comprehensive consideration over the spatial–temporal–frequency domains and its capability to adapt to different subjects make the proposed approach a promising candidate for future BCI systems. In current work, the control



**Fig. 7** Optimal frequency–temporal zones (colored in *black*) of six pairs of MI tasks found by ABC-based approach for subject 1 in BCI Competition IV Dataset IIa

parameters of ABC algorithm are mainly determined according to practice and the previous studies. For these parameters, the trade-off between classification accuracy and computational burden will be analyzed in the future.

**Acknowledgements** The study is supported by the Jiangsu Province Science and Technology Support Program of China (No. BE2012740). Special thanks to the reviewers for their positive and constructive comments and suggestions on our manuscript.

#### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Hu SQ, Tian QQ, Cao Y, Zhang JH, Kong WZ (2013) Motor imagery classification based on joint regression model and spectral power. *Neural Comput Appl* 23(7–8):1931–1936. doi:[10.1007/s00521-012-1244-3](https://doi.org/10.1007/s00521-012-1244-3)
- Blankertz B, Sannelli C, Haider S, Hammer EM, Kubler A, Muller KR, Curio G, Dickhaus T (2010) Neurophysiological predictor of SMR-based BCI performance. *Neuroimage* 51(4):1303–1309. doi:[10.1016/j.neuroimage.2010.03.022](https://doi.org/10.1016/j.neuroimage.2010.03.022)
- Ince NF, Goksu F, Tewfik AH, Arica S (2009) Adapting subject specific motor imagery EEG patterns in space-time-frequency for a brain computer interface. *Biomed Signal Process Control* 4(3):236–246. doi:[10.1016/j.bspc.2009.03.005](https://doi.org/10.1016/j.bspc.2009.03.005)
- Pichiorri F, Morone G, Petti M, Toppi J, Pisotta I, Molinari M, Paolucci S, Inghilleri M, Astolfi L, Cincotti F, Mattia D (2015) Brain-computer interface boosts motor imagery practice during stroke recovery. *Ann Neurol* 77(5):851–865. doi:[10.1002/ana.24390](https://doi.org/10.1002/ana.24390)
- Tavakolan M, Yong XY, Zhang X, Menon C (2016) Classification scheme for arm motor imagery. *J Med Biol Eng* 36(1):12–21. doi:[10.1007/s40846-016-0102-7](https://doi.org/10.1007/s40846-016-0102-7)
- Hamedi M, Salleh SH, Noor AM (2016) Electroencephalographic motor imagery brain connectivity analysis for BCI: a review. *Neural Comput* 28(6):999–1041. doi:[10.1162/NECO\\_a\\_00838](https://doi.org/10.1162/NECO_a_00838)
- Siuly S, Li Y (2015) Discriminating the brain activities for brain-computer interface applications through the optimal allocation-based approach. *Neural Comput Appl* 26(4):799–811. doi:[10.1007/s00521-014-1753-3](https://doi.org/10.1007/s00521-014-1753-3)
- Lotte F, Guan CT (2011) regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Trans Biomed Eng* 58(2):355–362. doi:[10.1109/tbme.2010.2082539](https://doi.org/10.1109/tbme.2010.2082539)
- Ramoser H, Muller-Gerking J, Pfurtscheller G (2000) Optimal spatial filtering of single trial EEG during imagined hand movement. *Ieee Trans Rehabil Eng* 8(4):441–446. doi:[10.1109/86.895946](https://doi.org/10.1109/86.895946)
- Dornhege G, Blankertz B, Krauledat M, Losch F, Curio G, Muller KR (2006) Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE Trans Biomed Eng* 53(11):2274–2281. doi:[10.1109/tbme.2006.883649](https://doi.org/10.1109/tbme.2006.883649)
- Thomas KP, Guan CT, Lau CT, Vinod AP, Ang KK (2009) A new discriminative common spatial pattern method for motor imagery brain-computer interfaces. *IEEE Trans Biomed Eng* 56(11):2730–2733. doi:[10.1109/tbme.2009.2026181](https://doi.org/10.1109/tbme.2009.2026181)
- Kee CY, Ponnambalam SC, Loo CK (2015) Multi-objective genetic algorithm as channel selection method for P300 and

- motor imagery data set. *Neurocomputing* 161:120–131. doi:[10.1016/j.neucom.2015.02.057](https://doi.org/10.1016/j.neucom.2015.02.057)
13. Xu P, Liu TJ, Zhang R, Zhang YS, Yao DZ (2014) Using particle swarm to select frequency band and time interval for feature extraction of EEG based BCI. *Biomed Signal Process Control* 10:289–295. doi:[10.1016/j.bspc.2013.08.012](https://doi.org/10.1016/j.bspc.2013.08.012)
  14. Ang KK, Chin ZY, Zhang HH, Guan CT (2012) Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs. *Pattern Recogn* 45(6):2137–2144. doi:[10.1016/j.patcog.2011.04.018](https://doi.org/10.1016/j.patcog.2011.04.018)
  15. Hsu WY (2011) EEG-based motor imagery classification using enhanced active segment selection and adaptive classifier. *Comput Biol Med* 41(8):633–639. doi:[10.1016/j.combiomed.2011.05.014](https://doi.org/10.1016/j.combiomed.2011.05.014)
  16. Wang T, Deng H, He B (2004) Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns. *Clin Neurophysiol* 115(12):2744–2753. doi:[10.1016/j.clinph.2004.06.022](https://doi.org/10.1016/j.clinph.2004.06.022)
  17. Wang T, He B (2004) An efficient rhythmic component expression and weighting synthesis strategy for classifying motor imagery EEG in a brain-computer interface. *J Neural Eng*. doi:[10.1088/1741-2560/1/1/001](https://doi.org/10.1088/1741-2560/1/1/001)
  18. Wu B, Yang F, Zhang JC, Wang YW, Zheng XX, Chen WD (2012) A frequency–temporal-spatial method for motor-related electroencephalography pattern recognition by comprehensive feature optimization. *Comput Biol Med* 42(4):353–363. doi:[10.1016/j.combiomed.2011.11.014](https://doi.org/10.1016/j.combiomed.2011.11.014)
  19. Higashi H, Tanaka T (2013) Common spatio-time-frequency patterns for motor imagery-based brain machine interfaces. *Comput Intell Neurosci*. doi:[10.1155/2013/537218](https://doi.org/10.1155/2013/537218)
  20. Karaboga D, Basturk B (2008) On the performance of artificial bee colony (ABC) algorithm. *Appl Soft Comput* 8(1):687–697. doi:[10.1016/j.asoc.2007.05.007](https://doi.org/10.1016/j.asoc.2007.05.007)
  21. Gao WF, Liu SY (2012) A modified artificial bee colony algorithm. *Comput Op Res* 39(3):687–697. doi:[10.1016/j.cor.2011.06.007](https://doi.org/10.1016/j.cor.2011.06.007)
  22. Blankertz B, Muller KR, Krusienski DJ, Schalk G, Wolpaw JR, Schlogl A, Pfurtscheller G, Millan JDR, Schroder M, Birbaumer N (2006) The BCI competition III: validating alternative approaches to actual BCI problems. *IEEE Trans Neural Syst Rehabil Eng* 14(2):153–159. doi:[10.1109/tnsre.2006.875642](https://doi.org/10.1109/tnsre.2006.875642)
  23. Shin Y, Lee S, Lee J, Lee HN (2012) Sparse representation-based classification scheme for motor imagery-based brain-computer interface systems. *J Neural Eng*. doi:[10.1088/1741-2560/9/5/056002](https://doi.org/10.1088/1741-2560/9/5/056002)
  24. Karaboga D, Akay B (2009) A comparative study of artificial bee colony algorithm. *Appl Math Comput* 214(1):108–132. doi:[10.1016/j.amc.2009.03.090](https://doi.org/10.1016/j.amc.2009.03.090)
  25. Saravanan M, Arulkumar PV (2015) An artificial bee colony algorithm for design and optimize the fixed area layout problems. *Int J Adv Manuf Technol* 78(9–12):2079–2095. doi:[10.1007/s00170-014-6774-7](https://doi.org/10.1007/s00170-014-6774-7)
  26. Lemm S, Blankertz B, Curio G, Muller KR (2005) Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Trans Biomed Eng* 52(9):1541–1548. doi:[10.1109/tbme.2005.851521](https://doi.org/10.1109/tbme.2005.851521)
  27. Ang KK, Chin ZY, Zhang HH, Guan CT, Ieee (2008) Filter bank common spatial pattern (FBCSP) in brain-computer interface. In: 2008 IEEE international joint conference on neural networks (IJCNN), vols 1–8, pp 2390–2397
  28. Zhang HH, Chin ZY, Ang KK, Guan CT, Wang CC (2011) Optimum spatio-spectral filtering network for brain-computer interface. *IEEE Trans Neural Netw* 22(1):52–63. doi:[10.1109/tnn.2010.2084099](https://doi.org/10.1109/tnn.2010.2084099)
  29. Suk HI, Lee SW (2013) A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces. *IEEE Trans Pattern Anal Mach Intell* 35(2):286–299. doi:[10.1109/tpami.2012.69](https://doi.org/10.1109/tpami.2012.69)
  30. Tangermann M, Muller KR, Aertsen A, Birbaumer N, Braun C, Brunner C, Leeb R, Mehring C, Miller KJ, Muller-Putz GR, Nolte G, Pfurtscheller G, Preissl H, Schalk G, Schogl A, Vidaurre C, Waldert S, Blankertz B (2012) Review of the BCI competition IV. *Front Neurosci*. doi:[10.3389/fnins.2012.00055](https://doi.org/10.3389/fnins.2012.00055)
  31. Meng JJ, Huang G, Zhang DG, Zhu XY (2013) Optimizing spatial spectral patterns jointly with channel configuration for brain-computer interface. *Neurocomputing* 104:115–126. doi:[10.1016/j.neucom.2012.11.004](https://doi.org/10.1016/j.neucom.2012.11.004)
  32. Das AK, Suresh S, Sundararajan N (2016) A discriminative subject-specific spatio-spectral filter selection approach for EEG based motor-imagery task classification. *Expert Syst Appl* 64:375–384. doi:[10.1016/j.eswa.2016.08.007](https://doi.org/10.1016/j.eswa.2016.08.007)
  33. Arvaneh M, Guan CT, Ang KK, Quek C (2014) Mutual information-based optimization of sparse spatio-spectral filters in brain-computer interface. *Neural Comput Appl* 25(3–4):625–634. doi:[10.1007/s00521-013-1523-7](https://doi.org/10.1007/s00521-013-1523-7)
  34. Wu W, Gao XR, Hong B, Gao SK (2008) Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL). *IEEE Trans Biomed Eng* 55(6):1733–1743. doi:[10.1109/tbme.2008.919125](https://doi.org/10.1109/tbme.2008.919125)
  35. Siuly Li Y, Wen P (2013) Identification of motor imagery tasks through CC–LR algorithm in brain computer interface. *Int J Bioinform Res Appl* 9(2):156–172
  36. Siuly, Li Y, Wu J, Yang J (2011) Developing a logistic regression model with cross-correlation for motor imagery signal recognition. In: 2011 IEEE/ICME international conference on complex medical engineering (CME), pp 502–507
  37. Novi Q, Guan C, Dat TH, Xue P, Ieee (2007) Sub-band common spatial pattern (SBCSP) for brain-computer interface. In: 2007 3rd international IEEE/EMBS conference on neural engineering, vols 1 and 2. doi:[10.1109/cne.2007.369647](https://doi.org/10.1109/cne.2007.369647)
  38. Kang S, Cho S, Kang P (2015) Constructing a multi-class classifier using one-against-one approach with different binary classifiers. *Neurocomputing* 149:677–682. doi:[10.1016/j.neucom.2014.08.006](https://doi.org/10.1016/j.neucom.2014.08.006)