Neurocomputing 443 (2021) 12–25

Contents lists available at ScienceDirect

Neurocomputing

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

Brief papers

A binary harmony search algorithm as channel selection method for motor imagery-based BCI

Bin Shi [a,b,1], Quan Wang [c,1], Shuai Yin [a,b], Zan Yue [a,b], Yaping Huai [d,*], Jing Wang [a,b,*]

[a] Institute of Robotics and Intelligent System, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[b] Shaanxi Key Laboratory of Intelligent Robots, Xi'an 710049, China
[c] Department of Occupational Therapy, Shanghai Yangzhi Rehabilitation Hospital (Shanghai Sunshine Rehabilitation Center), Shanghai 201619, China
[d] Department of Rehabilitation Medicine, Shenzhen Longhua District Central Hospital, Shenzhen 518000, China

ARTICLE INFO

Article history:
Received 26 September 2020
Revised 11 January 2021
Accepted 22 February 2021
Available online 10 March 2021
Communicated by Zidong Wang

Keywords:
Brain-computer interface (BCI)
EEG channel selection
Motor imagery
BHS algorithm

ABSTRACT

Background: Channel selection is a key topic in brain-computer interface (BCI). Task-irrelevant and redundant channels used in BCI may lead to low classification accuracy, high computational complexity, and inconvenience for application. By selecting optimal channels, the performance of BCI could enhance significantly.
Method: In this paper, a new binary harmony search (BHS) is proposed to select the optimal channel sets and optimize the system accuracy. The BHS is implemented on the training data sets to select the optimal channels and the test data sets are used to evaluate the classification performance on the selected channels. The sparse representation-based classification, linear discriminant analysis, and support vector machine are performed on the common spatial pattern (CSP) features for motor imagery (MI) classification.
Results: Two public EEG datasets are employed to validate the proposed BHS method. The paired *t*-test is conducted on the test classification performance between the BHS and traditional CSP with all channels. The results reveal that the proposed BHS method significantly improved classification accuracy as compared to the conventional CSP method ($p < 0.05$).
Conclusion: This study proposed the BHS method to select the optimal channels in MI -based BCI. On the one hand, the results confirm the validity of the BHS algorithm as a channel selection method for motor imagery data. On the other hand, the BHS method with costing shorter computation time relatively yields a better average test accuracy than the steady-state genetic algorithms. The proposed method could significantly improve the practicability and convenience of the BCI system.

1. Introduction

Brain-machine interface (BCI) is a system used to measure and convert electrophysiological indicators from the brain into commands to operate external devices [1]. It helps to accomplish tasks with the intention of users, such as selecting alphabets, moving the robotic arm, operating a wheelchair, controlling the mouse, moving a paralyzed limb with a neuro-prosthesis [2–5].

Compared with other existing neuroimaging modalities [6], electroencephalogram (EEG) has higher time resolution, affordability, availability, and portability, so it is more commonly used and has been studied in depth. In the study of EEG-based BCIs, much work has focused on motor imagery based BCI. By imagining the movement of different parts of the body, the subject can voluntarily regulate the mu or beta rhythm of the sensorimotor cortex [7].

One of the main challenges of BCI is that it is an individual-dependent system. In other words, even if the same paradigm and environment are applied, each person's brain different areas will be activated. The way to solve this challenge is to use as many electrodes as possible. Even so, using many electrodes brings about other problems, such as low classification accuracy, high computational complexity, and much setup time in some applications [8,9]. Consequently, channel selection is usually vital for EEG signal classification.

Channel selection belongs to a broader field of feature selection. The most commonly used EEG channel selection is filtering and wrapping methods. Filter technique only evaluates the relevance

* Corresponding authors at: Institute of Robotics and Intelligent System, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (Jing Wang).
E-mail addresses: huaiyaping@163.com (Y. Huai), wangpele@gmail.com (J. Wang).
[1] These authors contributed equally to this paper.

https://doi.org/10.1016/j.neucom.2021.02.051
0925-2312/

inside the channel set and is independent of the classifier. Nevertheless, filtering methods have two potential disadvantages [10]. The first disadvantage is that features are not relevant when considered individually. Features may become relevant when combined with other features. The second disadvantage is the possibility of selecting individual related features that may cause redundancy [11]. The Wrapper-based approach usually selects the channel coupled to a specific classifier. The applied classifier plays a vital role in the performance of the Wrapper-based approach [12]. Compared with filter-based methods, wrapper-based methods are relatively low computation cost and classification-dependent.

Channel selection is mathematically a problem of combinatorial optimization and analytical intractable [13]. Jong has shown that as a stochastic method, the genetic algorithm performs better than many deterministic optimization methods in high-dimensional space [14]. Meta-heuristic algorithms could be utilized to surmount the limitations of filter-based methods by evaluating the sub-set of variables based on their optimization function. The main advantage of using *meta*-heuristics algorithms to solve optimization problems can search in a set of possible solutions and find the optimal solution by iteratively trying to improve the feature subset according to the given quality metric [13]. In recent years, *meta*-heuristic algorithms have been widely used in EEG channel selection. Various genetic algorithms have effectively solved the problem of channel selection in the application of the BCI system [15–18]. He et al. proposed a Rayleigh coefficient (RC) maximization-based genetic algorithm (GA) for channel selection in the motor-imagery BCI system, which effectively improved the classification accuracy and computational load [17]. Kee et al. combined three multi-objective genetic algorithms (GAs) with wapper-based methods to improve channel reduction rate and classification accuracy [18].

Harmony Search (HS) is a new *meta*-heuristic algorithm firstly developed by Geem et al. [19] in 2001. HS algorithm and its variants have brought about widespread attention and been successfully applied to various practical structural optimization problems, such as overtopping breakwater [20], wireless sensor networks [21], aircraft stiffened panel [22]. In addition, most of the studies have reported that HS and its variants are superior to GA in terms of the optimal solution, convergence speed, and computation time[23,24]. Most previous studies on harmony search have focused on solving optimization problems in discrete or continuous spaces, and only a few studies have studied binary problems [24–27]. The BHS can show better performance than GA in some applications [23,24]. Ren et al. proposed a novel feature selection approach based on binary harmony search (BHS) which could avoid local convergence and identify multiple solutions because of the stochastic nature of BHS. Kim et al. employed the binary harmony search (BHS) algorithm to tackle the max-cut problem, which obtained better results compared with the generational GA and steady-state GA [28].

GA had widely applied to the EEG channel selection. However, there are relatively few studies devoted to applying the BHS to address channel selection. The present study proposes a binary harmony search algorithm with the wrapper method to select the optimal channels. The BHS is implemented on the training data sets to select the optimal channels and the test data sets are used to test the classification performance on the selected channel. The sparse representation-based classification (SRC), linear discriminant analysis (LDA), and support vector machine (SVM) are performed on the CSP features for MI classification. Two datasets are utilized to evaluate the performance of channel selection of the BHS. The classification performance of the BHS has also compared with the steady-state genetic algorithms (SSGA) method [18].

Therefore, this paper is organized into five sections with this section as the introduction. The proposed method and applied datasets are presented in section Method. The results are described in detail in section Results where the results of channel selection, classification accuracy, convergence, and computation time comparison are shown. In section Discussion, the parameter analysis of BHS and comparison between BHS and existed channel selection methods are discussed. The conclusions are provided in section 5.

2. Method

2.1. Description of the data

Data 1 (BCI Competition IV datasets 1): EEG data were measured from 7 healthy subjects through 59 electrodes. Subject "a" and "f" executed left-hand motor imagery tasks and foot while the rest subjects conducted right-hand and left-hand MI tasks. During each run, the visual cues are presented as left, right, or down arrows. The cue continues to display 4 s, during which the subjects are instructed to execute the corresponding motor imagery tasks. These periods are shown in the center of the screen interleaved with 2 s of a blank screen and 2 s of a fixed cross. The calibration data with a sampling rate of 100 Hz in this dataset consists of two runs, each of which contains 100 single trials. In this study, the first run of the calibration data was used as training data and test data and its numbers for all subjects are 80 and 20, respectively. The experimental paradigm of each trial is presented in Fig. 1. More detailed information about this database can be found at the following website: http: //www. bbci.de/competition/IV/.

Data 2 (BCI Competition III datasets IVa): This dataset is recorded from 5 healthy subjects ("aa", "al", "av", "aw", "ay") at 118 EEG channels. The sampling rate is 100 Hz. Each subject performs MI tasks from three classes (left-hand, right-hand, right-foot) and the cues are presented for a period of 3.5 s. This dataset only provides the EEG data of right-hand and right-foot. These visual cues were presented intermittently with random lengths ranging from 1.75 s to 2.25 s, during which the subject could relax. For all subjects, the data with a sampling rate of 100 Hz contained 140 trials, in which the first 110 trials were used for training and the rest of 30 trials were utilized for testing in this study. The experimental paradigm of each trial is presented in Fig. 2. More detailed information about this database can be found at the following website: http: //www.bbci.de/ competition/iii/.

2.2. Data preprocessing

The EEG data after the visual cue was used (2.5–6 s for data 1 and 0.5–3.5 s for data 2) [8]. Common average reference (CAR) and band-pass temporal filtering are conducted. CAR is to reduce artifacts and noise [29] and its form as follows:

$$x_i = x_i - \frac{1}{M}\sum_{n=1}^{M} x_n \tag{1}$$

where the ith channel of the EEG signal, M is the number of channels.

Min-max normalization [30] is performed independently for each channel for each trial. Min-max normalization is defined as follows:

$$x_{i,j} = \frac{x_{i,j} - min_{x_i}}{max_{x_i} - min_{x_i}} \tag{2}$$

where the j-th value of i-th channel of the EEG signal $x_{i,j}$. Every value of the EEG signal is transformed into between 0 and 1 to reduce the effect of amplitude difference between channels. Then, the EEG data

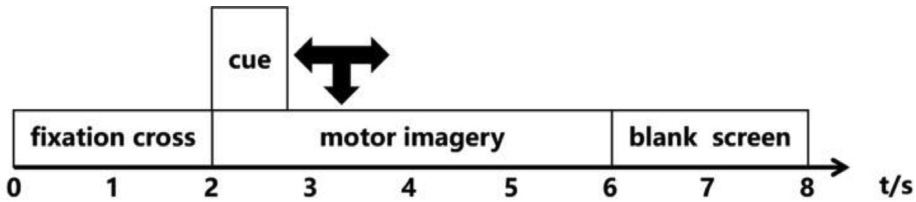


Fig. 1. The timing scheme of each trial of data 1.

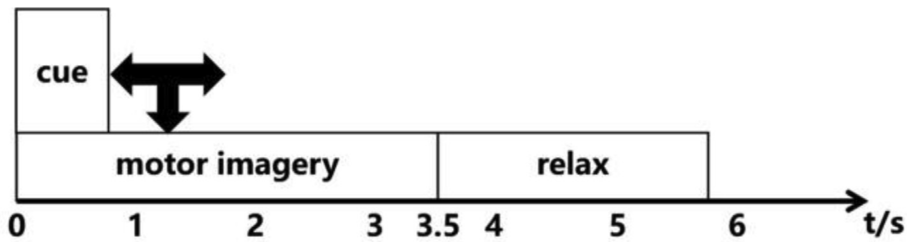


Fig. 2. The timing scheme of each trial of data 2.

were band-pass filtered employing a fifth-order Butterworth band-pass filter from 8 Hz to 40 Hz [31].

2.3. Common spatial pattern

CSP has been widely utilized in motor imagery-based BCI systems [32]. Its principle is to find the optimal projection direction through spatial projection so that the variance of one class of signal is the largest in a certain direction, and the variance of the other class of signal is the smallest, achieving that the difference between the two classes is the largest. CSP is realized by the simultaneous diagonalization of two classes of signal covariance matrices. The CSP algorithm is described as follows:

The preprocessed EEG data is represented as a matrix $X \in R^{N\times T}$. X represents the EEG signal of a trial and its dimensions $N \times T$, where N and T are the number of channels and of sampling points for per channel, respectively. $X_d,\ d \in \{1,\ 2\}$ represents the signal of class d. The normalized covariance matrix of the data is $C_d = (X_d {X_d}^T)/tr(X_d {X_d}^T)$. The mean covariance of each class is $\bar{C}_d$, Then the sum of the mean covariance matrices of the two classes can be expressed as $C_c = \bar{C}_1 + \bar{C}_2$. $C_c$can be decomposed as $C_c = U_c \lambda_c U_c^T$ where λ_c is the diagonal matrix of eigenvalues. Matrix $C_c = U_c \lambda_c U_c^T$ of eigenvectors U_c is transformed by P, where$P$ is the whitening matrix: $P = \sqrt{\lambda_c^{-1}} U_c^T$. Then, the $\bar{C}_1$ and $\bar{C}_2$ can be transformed into the following expressions:

$$R_1 = P\bar{C}_1 P^T \quad \text{and} \quad R_2 = P\bar{C}_2 P^T \tag{3}$$

R_1 and R_2 share the same eigenvectors, and the sum of the corresponding eigenvalues is 1. If $R_1 = B\lambda_1 B^T$, then

$$R_2 = B\lambda_2 B^T \ \text{ and } \ \lambda_1 + \lambda_2 = I \tag{4}$$

This means that in the direction with the largest eigenvalue of R_1, the eigenvalue of R_2 is the smallest, and vice versa and the difference between the two classes of signals is the largest. Projection matrix W can be obtained from $W = B^T P$, W is an $N{\times}N$ matrix whose rows are called spatial filters and columns are called spatial patterns. Thus, the EEG data of a single trial X is projected by W to obtain the new signal $Z = WX$.

After CSP processing, the new signal generated by filtering the first few rows of W (corresponding to the maximum eigenvalues of R_1) has the maximum variance when it belongs to class 1 and the minimum variance when it belongs to class 2. The new signal generated by filtering the last rows of W (corresponding to the smallest eigenvalues of R_1) is the opposite. The variance of the new signal generated by W to the optimal spatial filter (i.e., the first and last m rows of W) is logarithmically and normalized, and used as the feature. In this study, the value of m is 1. The features f_p, could be expressed as follows:

$$f_p = log\left(\frac{Var(Z_p)}{\sum_{i=1}^{2m} Var(Z_i)}\right) \tag{5}$$

Features are extracted from two classes of preprocessed EEG samples with CSP and these features and the corresponding labels can be used to train the classifier.

2.4. Classification

2.4.1. Sparse representation classification (SRC)

Shin et.al proposed the sparse representation-based classification (SRC) framework for the EEG MI-based BCI system [33]. SRC has been widely used to solve the problem of EEG signal classification. An important step for SRC is the design for an appropriate dictionary matrix A. The dictionary matrix A is formed by $A = [A_1, A_2]$, where A_1 and A_2 are two component dictionary matrices corresponding to two classes of features matrices. The training feature is represented by each column vector $a \in R^{2m\times 1}$, where a is equal to f_p. Let y denote a testing vector with the same dimension as a. Fig. 3 shows the linear sparse representation model for SRC. y can be sparsely represented as a linear combination of some a as follows: $y = \sum_{d=1,2} k_{d,1} a_{d,1} + k_{d,2} a_{d,2} + \cdots + k_{d,K_d} a_{d,K_d}$,where $k_{d,l}, l = 1,\ 2\ , ...,\ K_d$ are scalar coefficients and K_d is the number of column vectors of dictionary matric. The SRC method included two steps. The first step is to sparsely represent y using A by L1 norm minimization: $\underset{k}{min} \|k\|_1$ subject to $y = AK$, Where $A \in R^{2m\times(K_1+K_2)}$ is the dictionary and K is a scalar coefficient vector. In this study, orthogonal matching pursuit (OMP) can be used to solve this optimization problem [34–36]. The second step is to classify y via the minimum residual method. First, we define two temporary vectors $\mathbf{Q}_1 = \left[k_1, k_2, \ldots, k_{K_1}, 0, 0, \ldots, 0_{K_2}\right]^T$ and

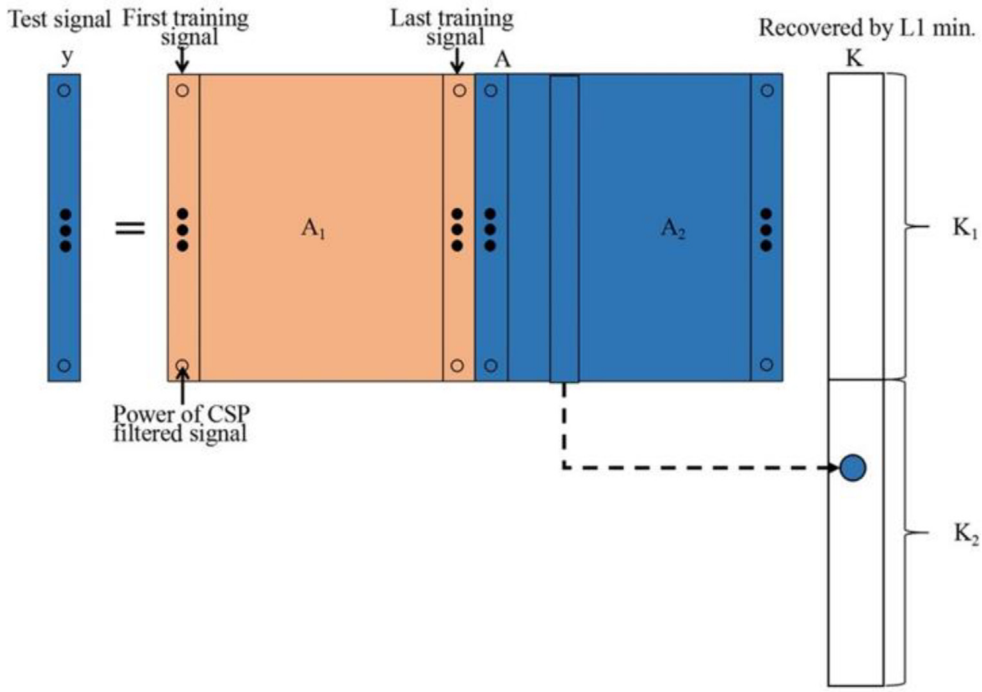


Fig. 3. Linear sparse representation model for SRC.

$\mathbf{Q}_2 = [0, 0, \ldots, 0_{K_1}, k_1, k_2, \ldots, k_{K_2}]^T$. Then, the residual norms for the two classes is obtained as follows:

$$r_1(y) = \|y - A\mathbf{Q}_1\|_2 \tag{6}$$

$$r_2(y) = \|y - A\mathbf{Q}_2\|_2 \tag{7}$$

Finally, the class of y is identified by comparing the values of $r_1(y)$ and $r_2(y)$ as

$$\text{class } (y) = \min_d \; r_d(y) \tag{8}$$

2.4.2. Support vector machine (SVM)

As a very well-known classification method, SVM was widely used in the field of MI-based BCI system [37,38]. If a feature sample data $A \in R^d$ from two classes could be expressed as $WA + b = 0$. $W \in R^d$ is the weight vector and b is the scalar. The SVM is to find the decision hyperplane by solving the following optimization problem [39]:

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2}\|w\|^2 + C\textstyle\sum_n \xi_n \\ \text{subject to} \quad & t_n(w^T\Phi(y_n) + b) \geq 1 - \xi_n \\ & \xi_n \geq 0, \quad n = 1, \ldots, N \end{aligned} \tag{9}$$

where y_n is the training feature vector, $t_n \in \{1, 2\}$ is the class information and n indicates the training trial number. ξ and C are a slack variable and a regularization parameter. In the SVM optimization problem, mapping function $\Phi(\cdot)$ can be used to map an inseparable feature vector onto a higher dimensional space using a kernel function $K(x, y)$. In BCI research, the Radial Basis Function (RBF) kernel is widely used [37]. Therefore, in this study, we consider an RBF kernel based SVM for comparison of the classification performance with the SRC method. For SVM algorithms, we use the MATLAB Toolbox (LIBSVM) [40].

2.4.3. Linear discriminant analysis (LDA)

LDA, as a linear classification method, is widely applied to the BCI field [37]. The LDA approach reported by Fisher aimed to find the optimal direction (w_L) to project sample data and maximize Fisher's ratio [41]:

$$J(w_L) = \frac{{w_L}^T S_B w_L}{{w_L}^T S_W w_L} \tag{10}$$

where S_B and S_W are denoted the between-class scatter matrix and within-class scatter matrix, respectively, which are expressed as follows:

$$S_B = (m_2 - m_1)(m_2 - m_1)^T \text{ and } S_W = \sum_i (x - m_i)(x - m_i)^T \tag{11}$$

where x is the input feature vector and m_i is the group mean of the feature vectors in class i.

2.5. Binary-coding harmony search algorithm

Harmony search (HS) [24] is a new *meta*-heuristic optimization method that simulates the music improvisation process, in which musicians improvise the pitch of instruments in order to find the perfect harmonic state. In this study, the binary coding method replaced the float encoding method in HS since candidate values of each variable only selected 0 and 1. The optimization problem can be described as follows:

$$\text{Minimizing } f(\mathbf{x}) \text{Subject to} \quad x_i \in \{0, 1\} \quad x_i \in \mathbf{x}, \; i = 1, 2, \cdots, n \tag{12}$$

In (13), $f(\mathbf{x})$ is the objective function; $\mathbf{x}$ is the set of each decision variable x_i; n is the number of decision variables.

The detail implementation of BHS is expressed as follows:

(1) Initializing the algorithm parameters. The parameters include the harmony memory size (HMS), the harmony memory considering rate (HMCR).

(2) Initializing the harmony memory. The HM denoted as $\mathrm{HM} = [x^1\ x^2\ \cdots x^{HMS}]^T$ is randomly initialized within the feasible solution space.

$$\mathrm{HM} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_n^1 \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \cdots & \vdots \\ x_1^{\mathrm{HMS}} & x_2^{\mathrm{HMS}} & \cdots & x_n^{\mathrm{HMS}} \end{bmatrix} \begin{bmatrix} f(\mathrm{x}^1) \\ f(\mathrm{x}^2) \\ \vdots \\ f(\mathrm{x}^{\mathrm{HMS}}) \end{bmatrix} \tag{13}$$

(3) Improvising new harmony. A new harmony denoted as $x' = (x_1', x_2', \cdots, x_{n-1}', x_n')$ is improvised by using the harmony memory consideration rule and randomization, which are determined by the pre-defined HMCR. In this study, the pitch adjustment operator is abolished which is the same as the improvisation process [24]. The detailed process is expressed as follows:

$$x_i' = \begin{cases} x_i' \in \{x_i^1, x_i^2, \cdots, x_i^{\mathrm{HMS}}\}; & if \;\; U(0,1) \le HMCR \\ x_i' \in \{0,1\}; & otherwise \end{cases} \tag{14}$$

where x_i' is the i th element of the new harmony candidate.

(4) Updating the harmony memory. The worst one in the HM is replaced by the new generated x_i' when it outperformed to the worst one.

(5) Judging the termination criterion. The iterative search is terminated and the optimal solution is output when the stopping criterion is satisfied. Otherwise, (3) and (4) are repeated. Fig. 4 shows the BHS algorithm flowchart.

2.6. Channel selection using binary harmony search algorithm

2.6.1. Design of harmony memory (HM)

In this study, the binary harmony memory algorithm is binary coded, in which the length of every harmony vector is equal to the number of channels available in each data set. Thus, the length of each harmony vector is 59 and 118 for data 1 and data 2, respectively. Each decision variable x_i represents a specific channel. If the element holds a value of 1, the corresponding channel is only selected. The initial harmony memory consists of HMS harmony vectors. Fig. 5 shows the binary code in HM and illustrates the channel selection.

2.6.2. Fitness function

The BHS operates on an HM and selects channels with optimizing a fitness function. The objective of channel selection is that the classification accuracy of the system should be uncompromised or

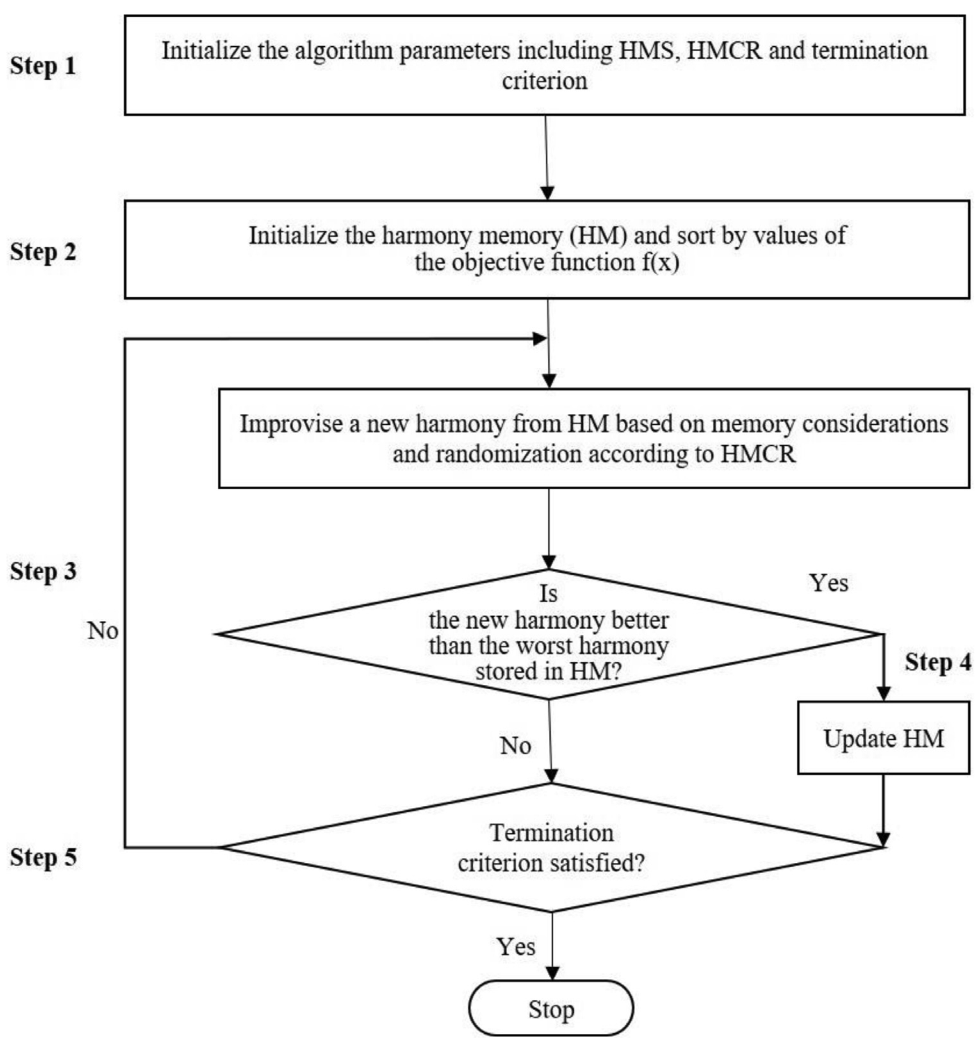


Fig. 4. Flowchart of the BHS algorithm.

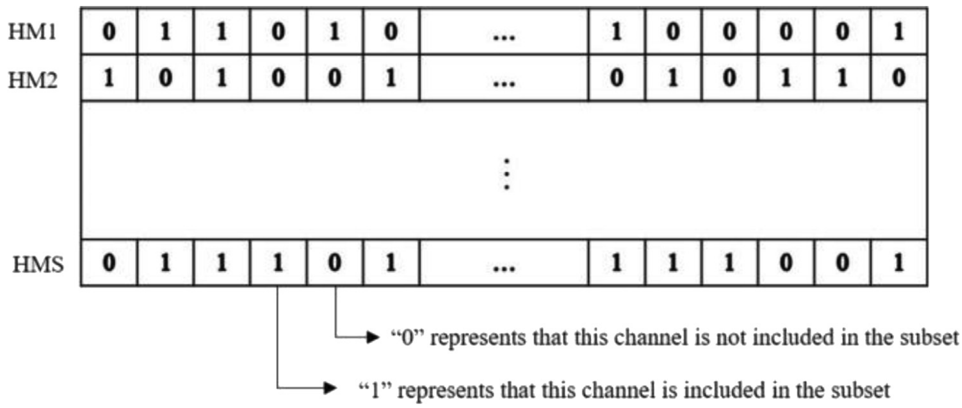


Fig. 5. Binary coding in harmony memory and the illustration of channel selection.

improved as much as possible based on reducing channels. The fitness function of binary quantum-behaved particle swarm optimization is defined as the weighted sum of classification error rate and relative number of channels [42]. The expression is as follows:

$$f(\boldsymbol{x}) = w_1 f_1(\boldsymbol{x}) + w_2 f_2(\boldsymbol{x}) \tag{15}$$

where $f(\boldsymbol{x})$ is the fitness function. $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ are classification error rate and the relative number of channels. The weights w_i are normalized, i.e.$\sum_{i=1}^{2} w_i = 1$. In our study,

$f_1(\boldsymbol{x})$ is the error rate of 10-fold cross-validation of training data and $f_2(\boldsymbol{x})$ is derived by dividing the number of channels chosen in the harmony vector (x) by the total number of channels in data set, i.e. the length of the binary string, n. Since the numerical value of $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ ranges from 0 to 1, so does that of $f(\boldsymbol{x})$.

Fig. 6 presents the structure of the channel selection using the BHS method. The proposed BHS method can be described in detail as follows:

Step (1) The original EEG data are divided into training data and test data according to the principle of 8–2. The training data is mainly used for the selection of the optimal channel set, while the test data is used for the performance of the selected optimal channel set. After that, the pre-processing of training data and test data mainly includes CAR, Min-max normalization, and frequency band filtering (8–40 Hz).

Step (2) Set iteration number, t = 1. The parameters include HMS, HMCR. HMS solutions are randomly generated to form the first HM_1. In our case, each solution embodies the selected channels. In addition, the stopping criterion is the maximum number of iterations.

Step (3) The fitness of solutions in HM_1 is evaluated. As each solution represents the selected channels, the EEG training data from selected channels will be extracted from the recording pre-processed EEG signal. Subsequently, feature extraction is performed by CSP and the 10-fold cross-validation accuracy of training data is calculated by SRC, LDA, and SVM. Finally, the fitness value of each harmony is calculated by the Eq. (15).

Step (4) A new harmony is improvised by the pre-defined HMCR. Moreover, the fitness value of newborn harmony is obtained by using the same method as in **step** (3).

Step (5) The worst one in the HM is replaced by the new generated x_i' when it outperformed the worst one.

Step (6) If the stopping criterion is not satisfied, the algorithm returns to Step (4) with t = t + 1. The algorithm will terminate if the stopping criterion is fulfilled. The optimal channels, the CSP filters, and classifiers are obtained.

Step (7) The test data are preprocessed by the selected optimal channel. Afterward, the test features are extracted by CSP filters. Finally, the test accuracy was calculated by the corresponding classifier obtained in Step (6).

3. Experimental results

In this section, we conduct a series of experiments to demonstrate the effectiveness of the proposed method. All experimental simulations are implemented by using MATLAB R2019b on a Windows personal computer with Core i5-9500H 3.00 GHz CPU and RAM 8.00 GB. The parameter values of the BHS and SSGA are listed in Table 1. In this study, we performed steady-state genetic algorithm (SSGA) for channel selection to compare the performance of the binary harmony search algorithm (BHS) in a BCI system. We used a population of P = 50 random chromosomes and set the maximum number of iterations to 500 (NI = 500). The SSGA employs the roulette wheel which is a selection mechanism widely used for GA. A single point crossover is applied for two-parent chromosomes and the crossover probability is 0.9 (P_c = 0.9). The bitwise mutation denoted as P_m is used and P_m is 0.05. In addition, we set HMS to 10, following from literature [28] and the HMCR is 0.95. The specific parameter analysis of BHS will be further discussed below.

From Eq. (15), it is shown that w1 and w2 were used to regulate the error rate of training data and the relative number of channels. Thus, w1 and w2 are very important for the fitness function. The left column in Fig. 7 presents the effect of w2 on the number of selected channels and system accuracy for data 1. System accuracy is the mean tenfold cross-validation accuracy of the training data 1 for subjects a-g. The value at the top of each bar in the subgraph is the average number of channels selected of subjects a-g for each w2. For BHS methods of different classifiers (SRC, LDA, SVM), the accuracy drastically decreases with the increase in w2 beyond 0.2. Therefore, the w2 was set to 0.2 in this study and w1 was set to 0.8. The right column in Fig. 7 presents the effect of w2 on the number of selected channels and system accuracy for data 2. Similar to data 1, the value at the top of each bar and right vertical axis are the average number of channels selected of subjects aa-ay for each w2 and mean tenfold cross-validation accuracy of the training data 2. With the increase in w2 beyond 0.4, the accuracy drastically decreases. Therefore, the value of w2 was set to 0.4 and the w1 was set to 0.6. In order to make a fair comparison, the optimal values of w1 and w2 obtained by BHS algorithm are also applicable to SSGA.

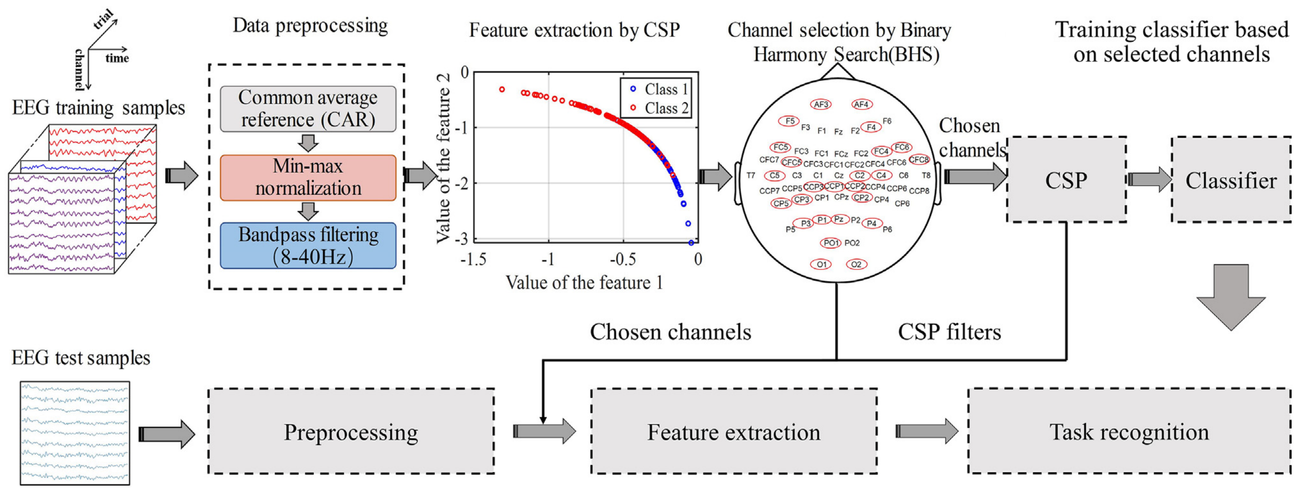


Fig. 6. The structure of the proposed method in this study.

Table 1
Parameters of the BHS and SSGA.

Algorithm	Parameter setting
BHS	HMS = 10, HMCR = 0.95, NI = 500
SSGA	P = 50, P_c = 0.9, P_m = 0.05, NI = 500

3.1. Channel selection

Fig. 8 shows the distributions of the selected channels by BHS based on different classifiers that are marked. For seven subjects of datasets 1, the deeper the yellow, the channel is selected by many subjects; conversely, the deeper the blue, the channel is not selected by the subjects. As shown in the first row of Fig. 8, the distribution selected channels located in the sensorimotor area which consistent with motor imagery of left hand, right hand, and foot. Moreover, for subjects aa to ay, the distribution of selected channels located in the left and central motor cerebral cortex area corresponding to motor imagery of the right hand and foot. It can be seen from Fig. 8 that the channel distribution selected by the BHS algorithm of different classifiers is different, but the overall distribution trend conforms to the neurophysiological principle.

3.2. Test results comparison

Table 2 summarizes the performance of all subjects in the test data. The traditional CSP method with all channels is compared with the SSGA and BHS based on different classifiers (SRC, LDA, and SVM) on two datasets. All of the BHS and SSGA algorithms relatively increased the average accuracy compared with the traditional CSP method. Moreover, the proposed BHS algorithm with fewer channels further yielded higher average accuracies rate than those of the SSGA and the traditional CSP methods. The average accuracy rate improvements achieved by BHS based on SRC were 13.9% and 1% in comparison with the CSP and SSGA based on SRC method. For LDA classifier, the average accuracy improvements obtained by BHS were 3.4% and 0.8% in comparison with the CSP and SSGA. Furthermore, the average accuracy rate improvements yielded by BHS based on SVM were 7.3% and 1.4% in comparison with the CSP and SSGA. The BHS based on three classification method obtained significant improvement in classification accuracy compared to CSP ($p < 0.05$). In summary, although the classification accuracy of BHS based on the three classifiers is different, the proposed BHS method achieved higher overall test classification accuracy over the traditional CSP methods for MI classification.

3.3. Convergence

Fig. 9 presents the convergence curve of BHS and SSGA based on different classifiers (SRC, LDA, SVM) in 500 iterations on the subject g and ay. We can observe that although the initial fitness value of SSGA is less than BHS, BHS produces a sharp decrease in fitness value than SSGA at the early iteration of the search. Moreover, it can be seen from the two convergence graphs that the convergence value of BHS has better fitness value than the SSGA. As shown in Fig. 9, SRC-based BHS and SSGA algorithms can obtain smaller fitness values at the end of iterations. In general, compared with SSGA algorithm, channel selection method based on BHS algorithm can obtain better fitness values on different classifiers.

3.4. Feature extraction

To assess the effectiveness of the proposed BHS algorithm, we explore the impact of the selected channels by BHS and SSGA algorithms on test features. Fig. 10 presents a two-dimensional feature distribution map for each class obtained by using BHS and SSGA based on SRC in subjects g and ay. As shown in Fig. 10, the two-dimensional feature distribution of each class in subject g (the number of samples = 20) and subject ay (the number of sample = 30). All test samples are scattered and fitted by Gaussion distribution. The blue asterisks and red circles represented the features samples of the two classes, respectively. It can be seen from the results in that, compared with the features obtained by the traditional CSP method without channel selection, the test features yielded by the proposed BHS algorithm in this paper can more beneficial for performing classification tasks. Furthermore, we note that the features separability of BHS outperformed the SSGA algorithm, which explains the better classification performance of BHS-SRC in Table 2.

3.5. Computational time comparison

The computational time of BHS and SSGA based on three classifiers (SRC, LDA, and SVM) methods on two datasets is shown in Fig. 11. The running time mainly includes preprocessing and

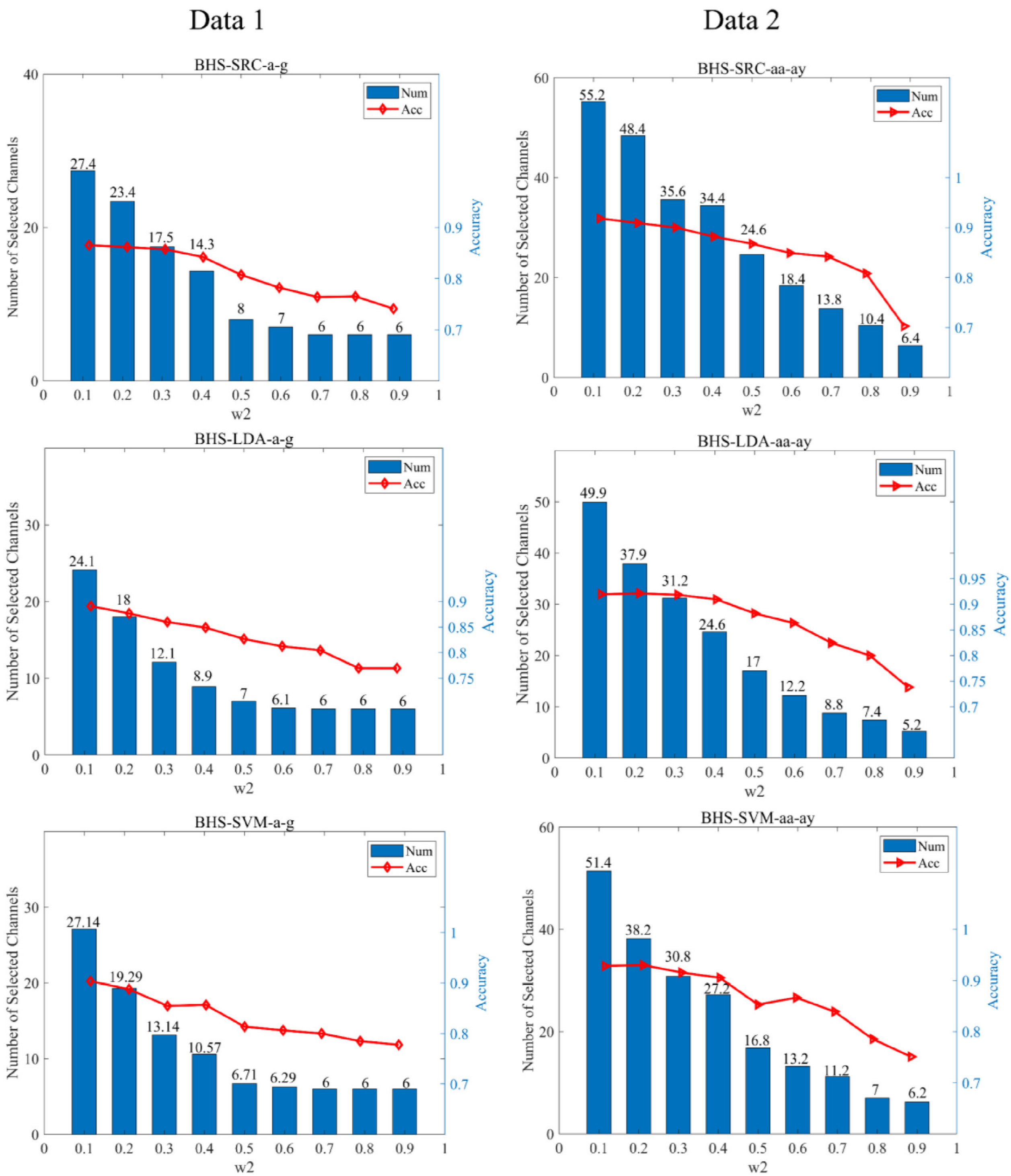


Fig. 7. The effect of w2 on the number of selected channels and system accuracy for data 1 and data 2.

channel selection without test running time. The results show that the computational time of the SSGA based on different classifiers is larger than BHS algorithm. This further demonstrates that the proposed BHS algorithm is easier to implement with low complexity. Moreover, due to the long running time of SVM algorithm, the BHS and SSGA spend more time to select the optimal channels on the two data sets. Finally, it is shown that the running time of each method is related to the sample size. Because dataset 2 has a large sample size, it takes the longest time.

4. Discussion

4.1. Parameter analysis of BHS

In the proposed BHS method, the HMCR and HMS of BHS played an important role in the optimal channel selection. Specifically, the larger the value of HMCR, the larger the convergence rate of BHS. In addition, a small HMCR value will increase the diversity of the algorithm. Thus, the value of HMCR is of great importance for EEG channels selection based on the BHS algorithm. We investi-

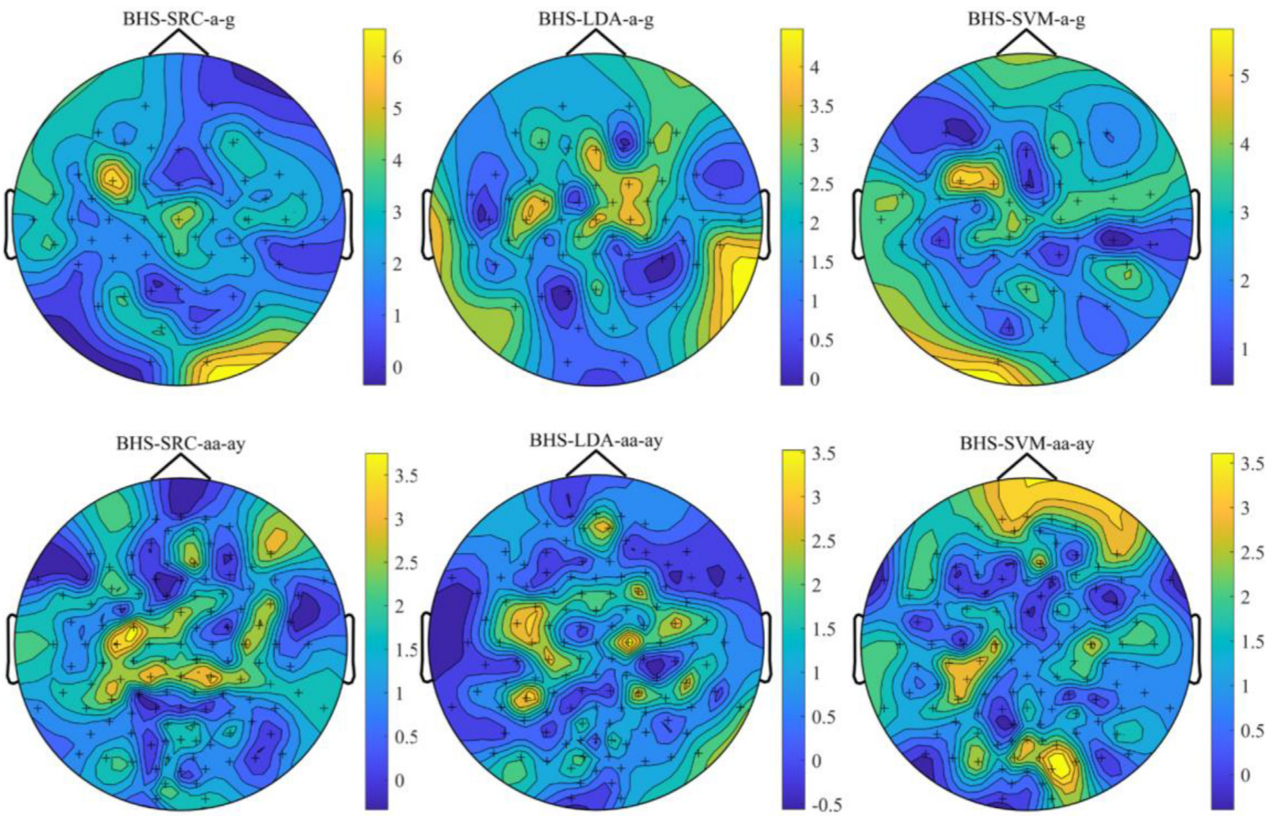


Fig. 8. The distributions of the selected channels by BHS based on different classifiers for two datasets. The deeper the blue, the fewer channels are selected, and the deeper the yellow, the more choices are made.

Table 2
Comparison of accuracy and selected channel number of different methods on all subjects.

Subjects	All channels			BHS-SRC		BHS-LDA		BHS-SVM		SSGA-SRC		SSGA-LDA		SSGA-SVM	
	SRC	LDA	SVM												
	Acc	Acc	Acc	Acc	Num	Acc	Num	Acc	Num	Acc	Num	Acc	Num	Acc	Num
a	50.0	67.5	50.0	80.0	17	67.5	15	62.5	21	82.5	20	70.0	16	60.0	19
b	62.5	70.0	57.5	70.0	20	70.0	13	70.0	19	65.0	24	70.0	17	72.5	25
c	50.0	45.0	57.5	60.0	28	57.5	19	55.0	28	60.0	23	62.5	24	67.5	21
d	90.0	90.0	90.0	90.0	19	90.0	24	85.0	24	92.5	24	87.5	25	70.0	25
e	77.5	90.0	77.5	90.0	25	85.0	20	80.0	21	95.0	24	82.5	18	75.0	21
f	50.0	55.0	52.5	72.5	24	57.5	15	65.0	22	65.0	21	62.5	14	62.5	22
g	67.5	75.0	50.0	85.0	18	87.5	14	75.0	29	85.0	26	75.0	17	72.5	20
aa	60.0	65.0	53.3	70.0	33	70.0	16	71.7	31	71.7	44	68.3	26	66.7	44
al	80.0	81.7	81.7	96.7	22	83.3	24	81.7	21	93.3	25	85.0	35	80.0	33
av	58.3	60.0	55.0	70.0	34	65.0	25	68.3	32	65.0	33	60.0	35	71.7	38
aw	73.3	80.0	80.0	85.0	41	80.0	38	73.3	43	85.0	37	81.7	38	65.0	34
ay	76.7	81.7	71.7	93.3	32	88.3	29	76.7	32	90.0	40	86.7	28	83.3	37
Mean	**66.3**	**71.7**	**64.7**	**80.2**	**26.1**	**75.1**	**21**	**72.0**	**26.9**	**79.2**	**28.4**	**74.3**	**24.4**	**70.6**	**28.3**
Std	13.3	13.9	14.4	11.5	7.6	11.9	7.3	8.5	6.9	12.9	7.9	10.1	8.3	6.8	8.5
p-value	–	–	–	0.006		0.022		0.013		0.012		0.084		0.073	

p value denotes the paired *t* test between test classification results of traditional CSP method with all channels and other method. Acc (%): test classification accuracy. Num : the number of selected channels.

gated the effect of the value of HMCR on the fitness value in the BHS based on SRC classifier method and HMCR tuned from 0.7 to 0.99. w1 and w2 are 0.8 and 0.2 for dataset 1. For dataset 2, w1 and w2 are 0.6 and 0.4. Moreover, the HMS is 10. Each result in the figure is an average of 10 independent runs. As can be seen from Fig. 12, For dataset 1 and 2, the average fitness value was the smallest when the HMCR was 0.95.

BHS generates new solution based on the HM. Thus, the HMS influences the performance of the BHS algorithm. To evaluate its influence on the fitness value, a set of the HMS values, i.e. 1,10,20,30,40, and 50, were studied. Meanwhile, HMCR = 0.95 was adopted. Fig. 12 shows the experimental results of the 10 independent runs and each value corresponding to HMS is an average fitness value. According to the results in Fig. 12, it is obvious that better fitness values are different for subjects of data 1 with the increasing of the HMS. In addition, the minimum fitness values are consistent for subjects of data 2. However, we can find that the average fitness values of all subjects are the smallest when HMS is 10. Therefore, HMS = 10 is the best value in this work.

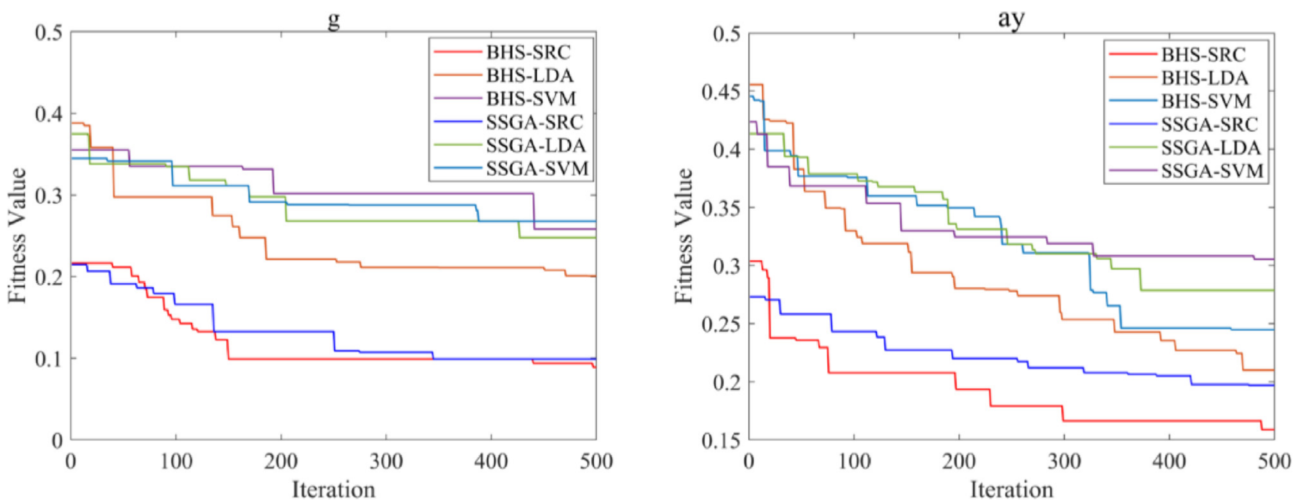


Fig. 9. Convergence graph of the BHS and SSGA on subject g and ay.

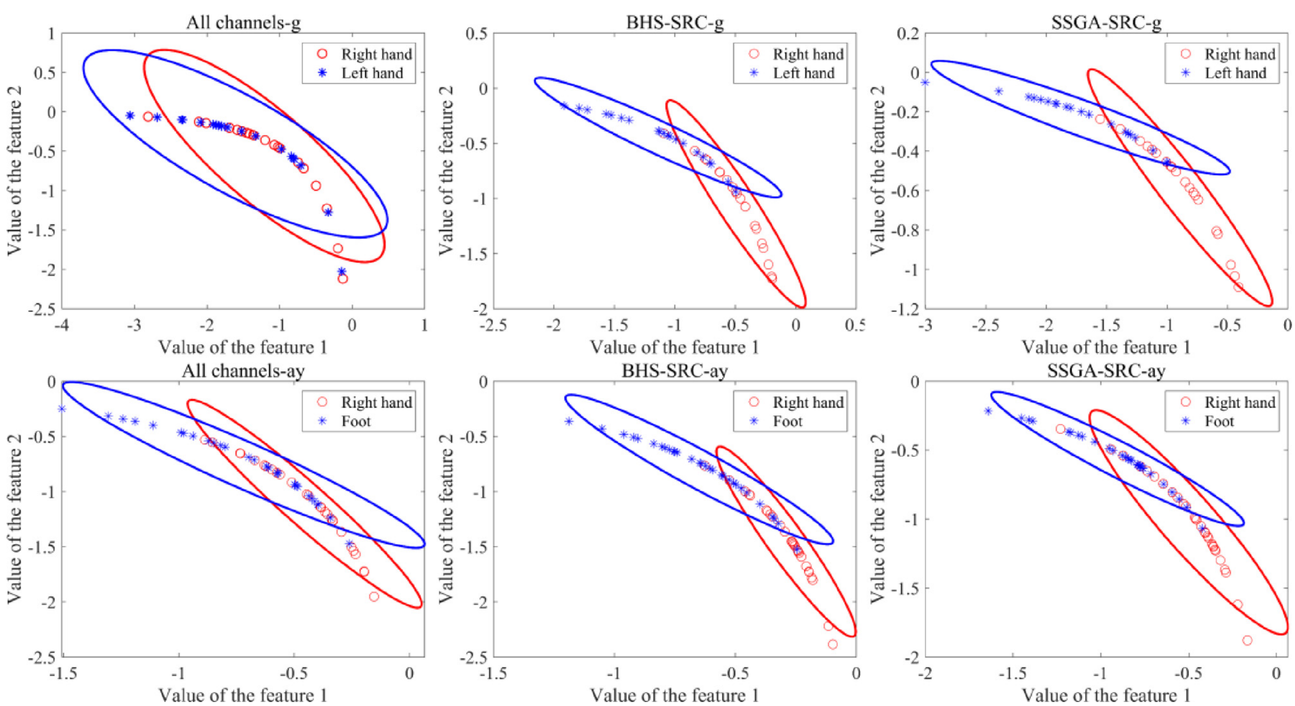


Fig. 10. A two-dimensional feature distribution map for each class obtain by using BHS and SSGA based on SRC. All test samples are scatted and fitted by Gaussian distribution.

4.2. Comparison between BHS and SSGA

In this section, we mainly compare and analyze the test performance of BHS and SSGA on all subjects. From the results of Table 2, compared with the SSGA method, the BHS algorithm based on SRC, LDA and SVM improved the average test accuracy by 1%, 0.8%, and 1.4%, respectively. To analyze the statistical significance of the classification differences, we perform the paired *t*-test for the test accuracy of BHS and SSGA. The obtained p-value of classification accuracy are 0.17 (SRC), 0.27 (LDA), and 0.25 (SVM), which indicates that the differences of test accuracy between BHS and SSGA are not statistically significant. However, the BHS algorithm achieved fewer channels than SSGA on three classification method. Fig. 9 presents convergence of the BHS and SSGA on subject g and ay. It can be seen that the BHS based on SRC, LDA, and SVM obtain the better fitness value at the end of iterations. Moreover, Fig. 11 shows the computation time comparisons between BHS and SSGA. it is important to note the BHS takes about two times less average computational time than SSGA on three classifiers. The reason why BHS is faster than SSGA in channel selection may be that BHS generates a new harmonic vector using harmony memory consideration rules. However, SSGA requires selection, crossover, and mutation operations when creating offspring. Therefore, BHS is easier to implement than SSGA.

4.3. Comparison of BHS with other channel selection methods

For BCI competition III data set IVa, the comparison of results is provided in Table 3. It can be concluded from the table that the BHS algorithm based on SRC is higher than SSGA [31] in

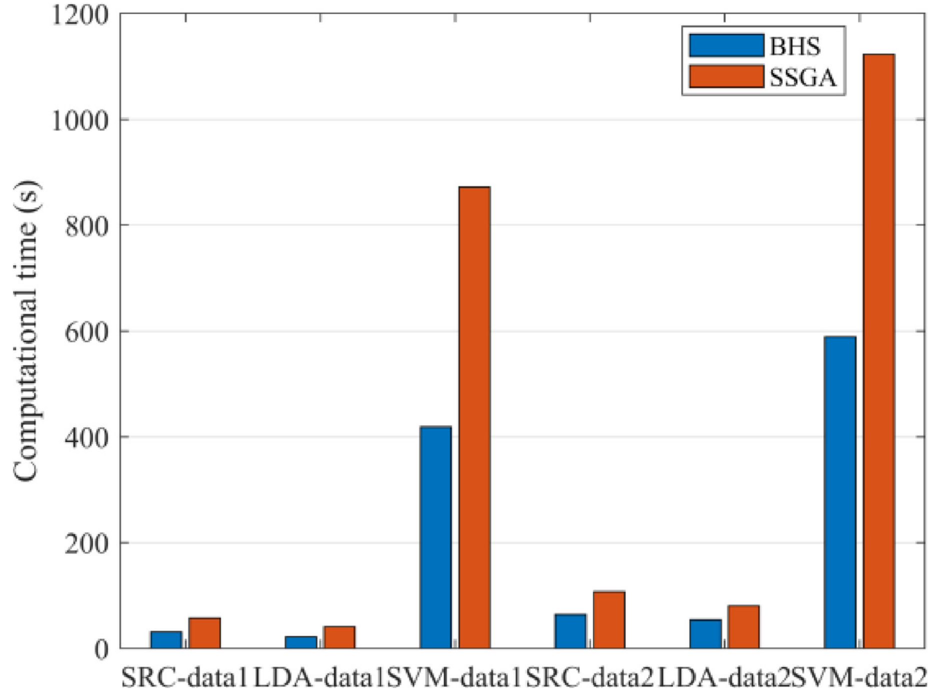


Fig. 11. Computation time comparisons between BHS and SSGA.

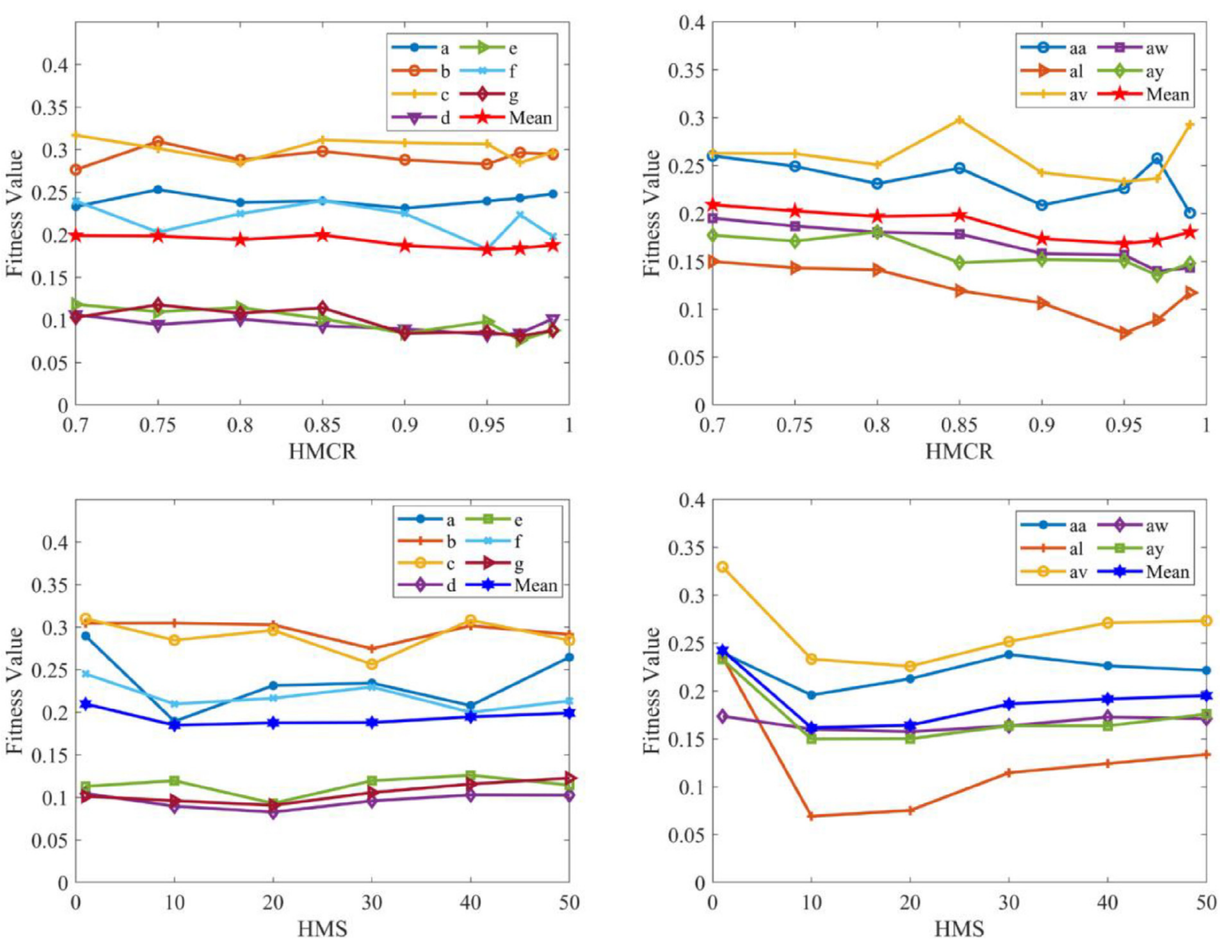


Fig. 12. The effect of HMCR and HMS on fitness value of the BHS method for all subjects.

Table 3
Comparison of BHS with existing channel selection methods.

Methods	Test_Acc (%)		Improvement_Acc (%)	Num	Computational Time (s)
	All channels	Selected channels			
SSGA [18]	77.99	82.39	4.4	21.5	–
SCSP1 [8]	73.56	82.28	8.72	22.6	–
SCSP2 [8]	73.56	79.28	5.72	7.6	–
Improved SFFS [12]	79.5	83.3	3.8	30.8	>2000
BHS-SRC	69.7	83	13.3	32.4	64.2
BHS-LDA	73.7	77.3	3.6	26.4	54.4
BHS-SVM	68.3	74.3	6	31.8	589

SCSP denotes sparse common spatial pattern, SFFS denotes sequential floating forward selection.

terms of improvement accuracy and test accuracy of selected channels, but the number of channels selected by SSGA is smaller than that of BHS based on SRC, possibly due to different objective functions. In addition, we note from the table that different classification methods produce different classification results and the number of selected channels for the BHS algorithm. Moreover, compared with the improved SFFS algorithm, the SRC-based BHS algorithm achieves similar test accuracy for selected channels and the number of channels. However, the SRC-based BHS algorithm achieves better improved accuracy. At the same time, the BHS based on three classification methods spends much less time on channel selection than Improved SFFS algorithm [25]. Furthermore, although the SRC-based BHS algorithm achieved the better test classification accuracy of the selected channels than SCSP2 [10], the SCSP2 offers a better higher channel reduction rate.

4.4. Limitations

The experimental results showed that the BHS reduced the number of channels and improved the classification accuracy compared with traditional CSP method. However, the proposed BHS method still has several limitations.

First, the BHS algorithm was adopted to tackle the channel selection for MI-based BCI in which the pitch adjustment operator (PAR) was not considered. The PAR played an important role in the HS algorithm and the PAR determines whether the new candidate needs to be refined. Therefore, future studies would be needed to improve the BHS algorithm with added to PAR. Second, the proposed BHS algorithm in this study only verifies the binary classification performance of MI-BCI systems, and the further research of the proposed BHS algorithm for channel selection to multiclass paradigms can be performed by using the one-versus-one or one-versus- rest approach. Finally, the experimental data used in this study are public data sets. Consequently, it is necessary to apply this BHS method to an online experiment of channel selection to investigate performance.

5. Conclusion

In this paper, a binary harmony search algorithm (BHS) is proposed to solve channel selection problems of MI-based BCI efficiently and effectively. Experimental studies on all subjects indicated that the proposed BHS method significantly improved classification accuracy as compared to the conventional CSP method. Moreover, the results show that the BHS method relatively yields the better average test accuracy than the SSGA method and the BHS has a shorter computation time. In summary, the proposed BHS method can significantly select optimal channels, which can improve the practicability and convenience of the BCI application. In future studies, the proposed BHS method should be assessed using more MI-EEG data and could be used for other BCI systems.

Declaration of Competing Interest

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

Acknowledgments

This work is sponsored by the Key Research and Development Plan of Shaanxi Program under Grant 2019SF-109.

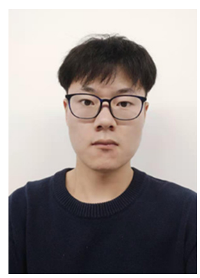
References

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain-computer interfaces for communication and control, Clin. Neurophysiol. 113 (2002) 767–791.
[2] M.A. Cervera, S.R. Soekadar, J. Ushiba, J.D.R. Millán, M. Liu, N. Birbaumer, G. Garipelli, Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis, Ann. Clin. Transl. Neurol. 5 (5) (2018) 651–663.
[3] I. Lazarou, S. Nikolopoulos, P.C. Petrantonakis, I. Kompatsiaris, M. Tsolaki, EEG-based brain-computer interfaces for communication and rehabilitation of people with motor impairment: a novel approach of the 21st century, Front. Hum. Neurosci. 12 (2018) 18.
[4] S.K. Mudgal, S.K. Sharma, J. Chaturvedi, A. Sharma, Brain computer interface advancement in neurosciences: Applications and issues, Interdiscip. Neurosurg. 20 (2020) 8.
[5] M. Rashid, N. Sulaiman, A. Majeed, R.M. Musa, A.F. Ab Nasir, B.S. Bari, S. Khatun, Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review, Front. Neurorobotics 14 (2020) 35.
[6] L.F. Nicolas-Alonso, J. Gomez-Gil, Brain computer interfaces, a review, sensors 12 (2012) 1211–1279.
[7] N. Brodu, F. Lotte, A. Lecuyer, Exploring two novel features for EEG-based brain-computer interfaces: multifractal cumulants and predictive complexity, Neurocomputing 79 (2012) 87–94.
[8] M. Arvaneh, C.T. Guan, K.K. Ang, C. Quek, Optimizing the channel selection and classification accuracy in EEG-based bci, IEEE Trans. Biomed. Eng. 58 (2011) 1865–1873.
[9] T. Alotaiby, F.E.A. El-Samie, S.A. Alshebeili, I. Ahmad, A review of channel selection algorithms for EEG signal processing, EURASIP J. Adv. Signal Process. 2015 (1) (2015), https://doi.org/10.1186/s13634-015-0251-9.
[10] Z. Shiqing, Z. Zhijin, Feature selection filtering methods for emotion recognition in Chinese speech signal, 2008 9th International Conference on Signal Processing (ICSP 2008), (2008) 1699-1702.
[11] B. Nakisa, M.N. Rastgoo, D. Tjondronegoro, V. Chandran, Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors, Expert Syst. Appl. 93 (2018) 143–155.
[12] Z.Y. Qiu, J. Jin, H.K. Lam, Y. Zhang, X.Y. Wang, A. Cichocki, Improved SFFS method for channel selection in motor imagery based BCI, Neurocomputing 207 (2016) 519–527.
[13] J.H. Yang, H. Singh, E.L. Hines, F. Schlaghecken, D.D. Iliescu, M.S. Leeson, N.G. Stocks, Channel selection and classification of electroencephalogram signals: an artificial neural network and genetic algorithm-based approach, Artif. Intell. Med. 55 (2012) 117–126.
[14] K. De Jong, Genetic algorithms: a 30 year perspective, Perspectives on adaptation in natural and artificial systems 11 (2005).
[15] K. Tavakolian, A.M. Nasrabadi, S. Rezaei, Selecting better EEG channels for classification of mental tasks, 2004 IEEE International Symposium on Circuits and Systems, (2004)537-537.

[16] J. Meng, G. Liu, G. Huang, X. Zhu, Automated selecting subset of channels based on CSP in motor imagery brain-computer interface system, in: 2009 IEEE International conference on robotics and biomimetics (ROBIO), 2009, pp. 2290–2294.
[17] L. He, Y.P. Hu, Y.Q. Li, D.L. Li, Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG, Neurocomputing 121 (2013) 423–433.
[18] C.Y. Kee, S.C. Ponnambalam, C.K. Loo, Multi-objective genetic algorithm as channel selection method for P300 and motor imagery data set, Neurocomputing 161 (2015) 120–131.
[19] Z.W. Geem, J. Kim, G. Loganathan, Music-inspired optimization algorithm harmony search, Simulation 76 (2001) 60–68.
[20] V.E. Kralli, N. Theodossiou, T. Karambas, Optimal design of overtopping breakwater for energy conversion (obrec) systems using the harmony search algorithm, Front. Energy Res. 7 (2019) 11.
[21] P.K. Poonguzhali, N.P. Ananthamoorthy, Design of mutated harmony search algorithm for data dissemination in wireless sensor network, Wirel. Pers. Commun. 111 (2020) 729–751.
[22] B. Keshtegar, P. Hao, Y.T. Wang, Q.Q. Hu, An adaptive response surface method and Gaussian global-best harmony search algorithm for optimization of aircraft stiffened panels, Appl. Soft. Comput. 66 (2018) 196–207.
[23] L. Wang, H. Ni, R. Yang, V. Pappu, M.B. Fenn, P.M. Pardalos, Feature selection based on meta-heuristics for biomedicine, Optimization Methods & Software 29 (2014) 703–719.
[24] Z.W. Geem, Harmony search in water pump switching problem, Advances In Natural Computation (2005) 751–760.
[25] Z.W. Geem, J.C. Williams, Ecological optimization using harmony search, Recent Advances On Applied Mathematics: Proceedings Of The American Conference On, Applied Mathematics (2008) 148–152.
[26] L. Wang, Y.F. Mao, Q. Niu, M.R. Fei, A multi-objective binary harmony search algorithm, Advances In Swarm Intelligence (2011) 74–81.
[27] D.X. Zou, L.Q. Gao, S. Li, J.H. Wu, Solving 0–1 knapsack problem by a novel global harmony search algorithm, Appl. Soft. Comput. 11 (2011) 1556–1564.
[28] Y.H. Kim, Y. Yoon, Z.W. Geem, A comparison study of harmony search and genetic algorithm for the max-cut problem, Swarm Evol. Comput. 44 (2019) 130–135.
[29] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.R. Muller, Optimizing spatial filters for robust EEG single-trial analysis, IEEE Signal Process. Mag. 25 (2008) 41–56.
[30] V. Gajera, Shubham, R. Gupta, P.K. Jana, An effective multi-objective task scheduling algorithm using min-max normalization in cloud computing, 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT). IEEE, 2016: 812-816.
[31] M.M. Miao, A.M. Wang, F.X. Liu, A spatial-frequency-temporal optimized feature sparse representation-based classification method for motor imagery EEG pattern recognition, Med. Biol. Eng. Comput. 55 (2017) 1589–1603.
[32] A.M. Alvarez-Meza, L.F. Velasquez-Martinez, G. Castellanos-Dominguez, Time-series discrimination using feature relevance analysis in motor imagery classification, Neurocomputing 151 (2015) 122–129.
[33] Y. Shin, S. Lee, J. Lee, H.-N. Lee, Sparse representation-based classification scheme for motor imagery-based brain-computer interface systems, J. Neural Eng. 9 (5) (2012) 056002, https://doi.org/10.1088/1741-2560/9/5/056002.
[34] D.L. Donoho, Compressed sensing, IEEE Transactions on Information Theory 52 (2006) 1289–1306.
[35] E.J. Candes, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. 59 (2006) 1207–1223.
[36] R. Baraniuk, Compressive sensing [lecture notes], IEEE Signal Process. Mag. 24 (4) (2007) 118–121.
[37] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update, J. Neural Eng. 15 (3) (2018) 031005, https://doi.org/10.1088/1741-2552/aab2f2.
[38] A. Schlögl, F. Lee, H. Bischof, G. Pfurtscheller, Characterization of four-class motor imagery EEG data for the BCI-competition 2005, J. Neural Eng. 2 (4) (2005) L14–L22.
[39] T. Sergios, C. Dionisis, K. Konstantinos, P. Aggelos, Introduction to pattern recognition: a matlab approach, (2010).
[40] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (3) (2011) 1–27.
[41] C.M. Bishop, Neural networks for pattern recognition, Agricultural Engineering International the Cigr Journal of Scientific Research & Development Manuscript Pm 12 (1995) 1235–1242.
[42] L. Zhang, Q.G. Wei, Channel selection in motor imaginary-based brain-computer interfaces: a particle swarm optimization algorithm, Journal of integrative neuroscience 2 (2019) 141–152.

Bin Shi received the M.S. in machinery manufacturing and its automation from Lanzhou University of Technology, Lanzhou, China, in 2017. He is currently pursuing his Ph.D. degree in instrumental science and technology from Xi'an Jiaotong university, Shannxi, China. His research interests includes brain-computer interface, signal processing, machine learning and pattern recognition.

Quan Wang received a master's degree in Occupational Therapy from Hong Kong Polytechnic University. He is currently serving as a section chief in the Occupational Therapy Department of the Rehabilitation Therapy Division in Shanghai Yangzhi Rehabilitation Hospital. He has more than ten years of teaching and clinical experience. His research interest includes Stroke & Neurological rehabilitation, Musculoskeletal Rehabilitation, Occupational Injury Rehabilitation. He is currently a committee member of the Chinese Association of Rehabilitation Medicine's branch of the Rehabilitation of Hand Function Committee, a youth committee member of the Chinese Association of Rehabilitation Medicine's branch of the Occupational Therapy Committee.



Shuai Yin received the B.S degree in Machinery Manufacturing and Automation from North West Agriculture and Forestry University, Shaanxi, China in 2018. He is currently pursuing his Ph.D. degree in Instrument Science and Technology in Xi'an Jiaotong University, Shaanxi, China since 2018. His research interests include lower limb rehabilitation robot, collaborative control, brain-computer interface and signal processing.

Zan Yue received the B.S degree from Wuhan University of Technology in 2015. He is currently pursuing his Ph.D. degree in Instrument Science and Technology in Xi'an Jiaotong University, Shaanxi, China. His current research interests include brain-computer interface, human-machine interface, training strategies for neurorehabilitation and virtual reality with force feedback.

Yaping Huai received the Ph.D. in Neurology from Hebei Medical University, Hebei, China in 2013. She is currently serving as Director of Department of Rehabilitation Medicine, Shenzhen Longhua District Central Hospital. In 2017, she has been the Vice-chairman, Professional Committee of Hand Function Rehabilitation, Chinese Association of Rehabilitation Medicine. Her research interests include Neurological Rehabilitation and Geriatric Rehabilitation.

Jing Wang: received the the Ph.D. in Instrument Science and Technology from Xi'an Jiaotong University, Shaanxi, China in 2008. He is an associate professor and doctoral supervisor at the School of Mechanical Engineering of Xi'an Jiaotong University, Shaanxi, China. He conducted post-doctoral research from Georgetown University in the United States from 2009 to 2011. From 2014 to 2016, he conducted post-doctoral research from the Hong Kong Polytechnic University and the Chinese University of Hong Kong, Hong Kong, China. His research interests include brain-computer interface technology, stroke rehabilitation robots and signal processing.