# Bispectrum-Based Channel Selection for Motor Imagery Based Brain-Computer Interfacing

Jing Jin[ID], *Senior Member, IEEE*, Chang Liu, Ian Daly, Yangyang Miao[ID], Shurui Li[ID], Xingyu Wang, and Andrzej Cichocki[ID], *Fellow, IEEE*

*Abstract*— **The performance of motor imagery (MI) based Brain-computer interfacing (BCI) is easily affected by noise and redundant information that exists in the multi-channel electroencephalogram (EEG). To solve this problem, many temporal and spatial feature based channel selection methods have been proposed. However, temporal and spatial features do not accurately reflect changes in the power of the oscillatory EEG. Thus, spectral features of MI-related EEG signals may be useful for channel selection. Bispectrum analysis is a technique developed for extracting non-linear and non-Gaussian information from non-linear and non-Gaussian signals. The features extracted from bispectrum analysis can provide frequency domain information about the EEG. Therefore, in this study, we propose a bispectrum-based channel selection (BCS) method for MI-based BCI. The proposed method uses the sum of logarithmic amplitudes (SLA) and the first order spectral moment (FOSM) features extracted from bispectrum analysis to select EEG channels without redundant information. Three public BCI competition datasets (BCI competition IV dataset 1, BCI competition III dataset IVa and BCI competition III dataset IIIa) were used to validate the effectiveness of our proposed method. The results indicate that our BCS method outperforms use of all channels (83.8% vs 69.4%, 86.3% vs 82.9% and 77.8% vs 68.2%, respectively). Furthermore, compared to the other state-of-the-art methods, our BCS method also can achieve significantly better classification accuracies for MI-based BCI (Wilcoxon signed test, p <0.05).**

*Index Terms*— **Brain-computer interface, motor imagery, electroencephalogram (EEG), bispectrum analysis, channel selection.**

## I. INTRODUCTION

**B**RAIN-COMPUTER Interfacing (BCI) provides a new external pathway which can use brain signals to control devices (e.g., wheelchairs, robot arms, drones etc.) without involvement of the muscles [1], [2]. Electroencephalogram (EEG)-based BCI is one of the most commonly studied BCI systems. It has the advantages of being non-invasive, low cost, and highly portable [3]. Currently, P300 [4]–[7], steady-state visual evoked potential (SSVEP) [8], [9], and motor imagery (MI) [10]–[13] based BCIs are the three main research directions in the field of EEG-based BCI. Although P300 and SSVEP based BCI systems have better classification accuracies and information transfer rates (ITRs), these two kinds BCI are limited because of the need for an external stimulus [14], [15]. Compared with stimulation-based BCIs, MI-based BCIs are often more intuitive for their users. When individuals imagine movement of their body there are associated changes in the power of the oscillatory EEG. These changes are known as event-related desynchronization (ERD) and event-related synchronization (ERS) [16], and their accurate detection is the basic principle of MI based BCI.

In order to solve the problem of performance degradation due to the redundant information and noise across multiple EEG channels [17], many researchers have proposed channel selection methods for MI-based BCI. Considering the neurophysiologic knowledge, the simplest channel selection method is to extract data from EEG channels Cz, C3, and C4, which are located over the motor cortex.

Channel selection methods can be categorized as either filter [18] or wrapper [17] based methods. Filter methods usually select channels based on F score, mutual information or another criterion. In contrast, wrapper methods make use of a classifier to validate the effects of different sets of selected channels. However, those methods are most frequently used to select channels based on their temporal [18], [19] or spatial features [17].

Bispectrum analysis has also been used to extract features for MI-based EEG classification [20]–[22].

Jing Jin, Chang Liu, Yangyang Miao, Shurui Li, and Xingyu Wang are with the Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China (e-mail: jinjingat@gmail.com; liuchang@mail.ecust.edu.cn; miaoyy1991@163.com; shuruilil008@163.com; xywang@ecust.edu.cn).

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

Andrzej Cichocki is with the Skolkovo Institute of Science and Technology (Skoltech), 121205 Moscow, Russia, and also with the Department of Applied Computer Science, Nicolaus Copernicus University (UMK), 87-100 Torun, Poland (e-mail: a.cichocki@skoltech.ru).

Digital Object Identifier 10.1109/TNSRE.2020.3020975

Bispectrum analysis is a technique which can extract useful non-linear and non-Gaussian information from non-linear and non-Gaussian signals such as EEG [23]. But to the best of our knowledge, bispectrum-based features have not been used to aid EEG channel selection. Therefore, this study proposes a bispectrum-based channel selection method. We suggest that the spectral information existing in the bispectrum of MI-related EEG signals may aid identification of which channels contain redundant information. In this study, the sum of logarithmic amplitudes (SLA) and the first order spectral moment (FOSM) features extracted from bispectrum analysis are used to select channels for MI-based BCI.

The rest part of paper is organized as follows. Section II introduces our proposed method and related research. Section III describes the data used in this study and the experimental setup. Section IV and V shows the result and discussion, respectively. Section VI is the conclusion of our study.

## II. METHODS

### A. Bispectrum Analysis

Bispectrum analysis is a statistical analysis method used to study the interactions between non-linear signals [20]. It is the expectation of three frequencies which includes two direct frequency components and a complex conjugate frequency of the sum of the two direct frequencies. Moreover, for a deterministic and zero mean signal $x(n)$, its bispectrum $B(f_1, f_2)$ can be estimated by:

$$B(f_1, f_2) = E < X(f_1)X(f_2)X^*(f_1, f_2) > \quad (1)$$

where $X(f)$ is the Fourier Transform (FT) of the signal $x(n)$, $X^*(f)$ is the complex conjugate of $X(f)$ and $E < \bullet >$ is the statistical expectation operator.

In practice, both nonparametric and parametric model based techniques can be used to estimate the bispectrum [22]. Moreover, nonparametric techniques can be employed with indirect and direct methods [24]. The direct method has the advantages of being easy to implement and requiring less computational resources [25]. Therefore, we use the direct nonparametric technique to estimate the bispectrum in this study.

As the third-order spectrum, bispectrum analysis is able to increase the signal-to-noise ratio (SNR) of signals contaminated with Gaussian noise. It also can detect and quantify the quadratic phase coupling (QPC) between signals [22]. That is to say, compared to power spectrum analysis, bispectrum analysis may be more suitable for processing MI-related EEG signals, which have non-linear and non-Gaussian characteristics [26].

### B. Bispectrum Based Feature Extraction

The bispectrum of a set of signals has symmetrical properties because of the conjugate symmetry of the Fourier transform (FT). Hence, we only calculate the bispectral values on the principal domain (non-redundant area) in this study [27]. The principal domain $\Omega$ is shown as a gray area in Fig. 1.
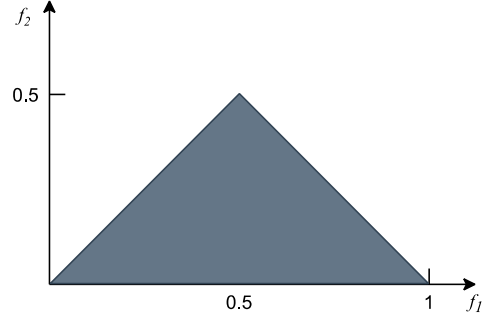


Fig. 1. Principal domain (gray area) of bispectrum estimation. The $f_1$ and $f_2$ represent two normalized frequency components used to estimated bispectrum.

It is given by the triangular region $f_1 \geq f_2 \geq 0$ and $f_1 + f_2 \leq 1$ [28]. The $f_1$ and $f_2$ represent two frequency components of EEG and the frequencies are shown normalized in Fig. 1.

We do not directly use the bispectral values in this study. In order to retain the temporal and spectral information within the MI-related EEG signals, the following two features of the bispectrum are extracted:
1) The sum of the logarithmic amplitudes.
2) The first order spectral moment.

### C. Sum of Logarithmic Amplitudes (SLA)

The SLA is a widely derived bispectral feature that can be obtained by computing the sum of absolute log bispectrum over all bisfrequencies in the principal domain [22]. Mathematically this is defined as

$$B_{Sla} = \sum_{f_1, f_2 \in \Omega} \log(|B(f_1, f_2)|) \quad (2)$$

where $B(f_1, f_2)$ denotes the 2D bispectrum on the principal domain $\Omega$.

### D. First Order Spectral Moment (FOSM)

The FOSM is the first order spectral moment of the amplitudes of the diagonal elements in the bispectrum and can be calculated according to the following formula:

$$B_{Fosm} = \sum_{n=1}^{N} n \times \log(|B(n, n)|) \quad (3)$$

where $N$ denotes the number of diagonal elements of the bispectrum.

### E. Bispectrum-Based Channel Selection (BCS)

We think the channels related to MI should contain more information in the $\mu$ (8-13 Hz) and $\beta$ (13-25 Hz) frequency bands when participants are performing MI tasks [29]. Meanwhile, compared to MI related channels, other channels may contain less information in the corresponding bands.

Hence, we can select channels by extracting bispectrum features (e.g., SLA and FOSM) of each channel for different classes of EEG data. Then, the discriminative power of the
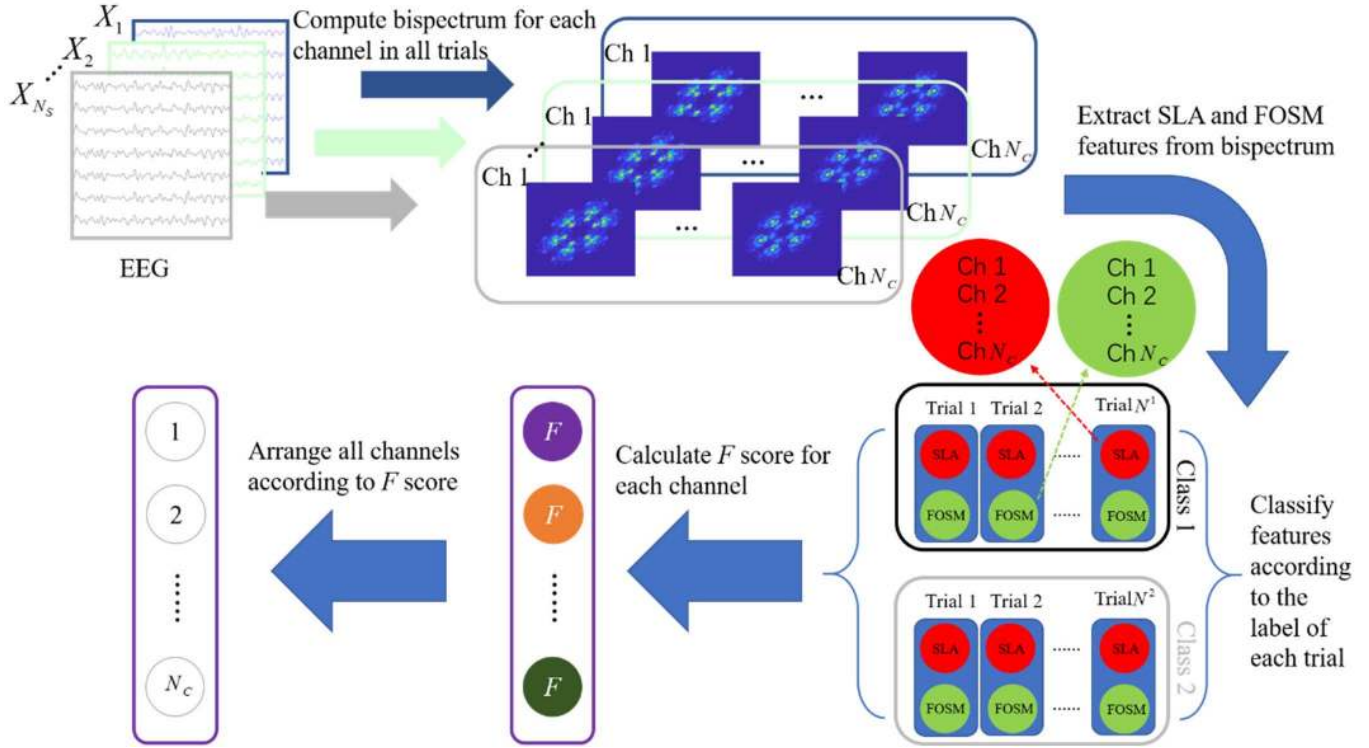
Fig. 2. Diagram of BCS method. First, each channel of EEG data estimates the bispectrum on the principal domain (Fig. 1). Then, the SLA and FOSM features are extracted from the bispectrum of each channel. Next, features are classified to two classes according to the label of data. Finally, F score can be calculated for each channel, and all channels will be arranged according to F score.

bispectrum features is estimated by the F score, which is calculated according to the following formula [18]:

$$\hat{F}_b = \frac{(\bar{B}_{Sla}^1 - \bar{B}_{Sla}^2)^2 + (\bar{B}_{Fosm}^1 - \bar{B}_{Fosm}^2)^2}{\tilde{B}_{Sla}^1 + \tilde{B}_{Sla}^2 + \tilde{B}_{Fosm}^1 + \tilde{B}_{Fosm}^2} \quad (4)$$

with

$$\overline{BIS}_b^\chi = \frac{1}{N^\chi} \sum_{i=1}^{N^\chi} BIS_b^\chi(i) \quad (5)$$

$$\widetilde{BIS}_b^\chi = \frac{1}{N^\chi - 1} \sum_{i=1}^{N^\chi} (BIS_b^\chi(i) - \overline{BIS}_b^\chi)^2 \quad (6)$$

where $\overline{BIS}_b^\chi(i)$ and $\widetilde{BIS}_b^\chi(i)$ are the estimated means and variances of the bispectrum features (e.g., SLA and FOSM), and $N^\chi$ is the number of training trials belong to class $\chi \in \{1, 2\}$. It should be noted that this study only discusses the issue of channel selection with two classes of EEG data.

Let $X \in \mathrm{R}^{N_C \times N_T \times N_S}$ denote the recorded EEG signals, where $N_C, N_T$ and $N_S$ denote the number of channels, the number of temporal points, and the trial number of EEG signals, respectively. To be noted, $N_s = \sum_{\chi=1}^{2} N^\chi$. The diagram and steps of the proposed method are shown in Fig. 2 and the following table, respectively.

## F. MI Feature Extraction and Classification

The common spatial pattern (CSP) algorithm is used to extract features in this study. CSP can identify a spatial filter

**Algorithm 1** Bispectrum-based channel selection algorithm (BCS)

1 Extract data $X \in \mathrm{R}^{N_C \times N_T \times N_S}$ from a fixed time window of raw EEG data, $N_C$ = number of channels, $N_T$ = number of temporal points, $N_S$ = trial number of EEG signals
2 For $i = 1:N_S$
3    For $j = 1:N_C$
4       Estimate the bispectrum of each channel in each trial of $X$ according to Eq. (1), $H_{Bis}(:,:,j,i) = B(X(j,:,i))$
5       Calculate the SLA features $B_{Sla}(j,i)$ of $H_{Bis}^\chi(:,:,j,i)$ according Eq. (2)
6       Calculate the FOSM features $B_{Fosm}(j,i)$ of $H_{Bis}^\chi(:,:,j,i)$ according Eq. (3)
7    End
8 End
9 SLA and FOSM feature matrices $B_{Sla}, B_{Fosm} \in R^{N_C \times N_S}$ are obtained. Then, $B_{Sla}, B_{Fosm}$ can be split into $B_{Sla}^1, B_{Sla}^2, B_{Fosm}^1 and B_{Fosm}^2$ according to the label information of each trial.
10 Compute the $F$ score for each channel according to Eqs. (4), (5), and (6)
11 Arrange $N_C$ channels in descending order according to the $F$ score, and acquire the first $N_{Ch}$ channels.

to maximize the variance for one class and minimize the variance for the other at the same time [30]–[32]. Suppose $X_{i,l} \in \mathrm{R}^{N_C \times N_T}$ denotes EEG temporal data in the $i$ th trial

of class $l$ ($l \in \{1, 2\}$), where $N_C$ and $N_T$ are the number of channels and the number of samples, respectively. The spatial filter can be obtained by the following steps.

First, the raw EEG signals should be bandpass filtered at a specified frequency band and the mean should be subtracted. Then, the spatial covariance matrix of class $\chi$ can be calculated as:

$$\Sigma_\chi = \frac{1}{N^\chi} \sum_{i=1}^{N^\chi} X_{i,l} X_{i,l}^T \tag{7}$$

where $N^\chi$ and $T$ denote the number of trials belong to $\chi$ and the transpose operator, respectively.

Then, the way to find the spatial filter can be expressed as the following Rayleigh quotient [19]:

$$\arg\min_w \frac{w^T \Sigma_1 w}{w^T \Sigma_2 w} s.t. ||w||_2 = 1 \tag{8}$$

To solve Eq. (8), it can be transformed into a generalized eigenvalue problem:

$$\Sigma_1 w = \lambda \Sigma_2 w \tag{9}$$

Hence, the $M$ largest and smallest generalized eigenvalues can form a spatial filter $W = [w_1, \cdots, w_{2M}] \in R^{N_C \times 2M}$. In this study, $M$ is set as 2. Finally, the features of a single trial $X_i$ can be computed as:

$$z = \log(var(W^T X_i)) \tag{10}$$

After feature extraction by the CSP algorithm, we fed the features into a support vector machine (SVM) classifier with a radial basis function (RBF) kernel for the classification. In this study, we use the LIBSVM [33] toolbox in MATLAB to implement the SVM classifier. Many methods can be used to classify. The reason for using SVM as a classifier is that it performs well in BCI systems. The main idea of SVM is to find a hyperplane with the largest possible margin to separate the data from the two classes [34], [35].

$$H : x \mapsto sign(w^T x + b) \tag{11}$$

where $w$ and $b$ are the weight vector and the offset of the hyperplane, respectively.

## III. MATH MATERIALS AND EXPERIMENT

### A. Data Descriptions

DS1: We denote the dataset 1 from the BCI Competition IV [36] as DS1 in this study. DS1 contains MI EEG data recorded from 7 participants. However, we only use the data for participants 'a', 'b', 'f' and 'g' because the data for participants 'c', 'd' and 'e' are artificially generated. At the beginning of each run, a fixation cross was displayed at the center of the computer screen for 2 s. Then, an arrow with the direction of left, right or down was displayed for 4 s to instruct participants to perform the corresponding MI task. Next, participants were asked to rest for 2 s while looking at a blank screen. After the rest time, the next trials occurred. Each of run contained 100 trials. The timeline of each trial is shown in Fig. 3(a). We down-sampled the EEG recorded from 59 channels to
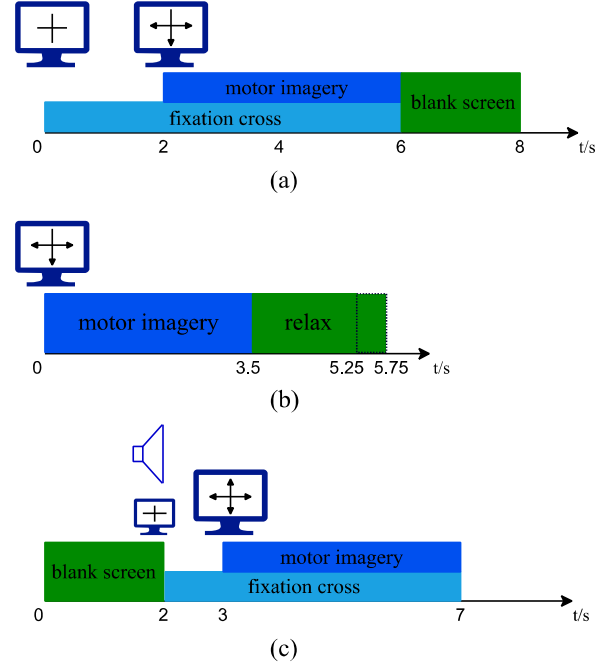


Fig. 3. Timeline of each trial for DS1(a), DS2(b) and DS3(c).

100 Hz. The channels were positioned according international EEG 10-20 systems. More details can be found in the following website: http://www.bbci.de/competition/iv/.

DS2: We denote dataset IVa from the BCI Competition III [37] as DS2. This dataset was recorded from 5 healthy participants. The timeline is shown in Fig. 3(b). It can be seen that visual cues were first displayed for 3.5 s to instruct participants to perform left hand, right hand and foot imagery. Then, participants were instructed to randomly relax for 1.75 to 2.25 s. The total number of trials completed for each participant was 280 and the EEG was recorded from 118 channels that were positioned according to the international 10-20 system for electrode placement. The EEG were down-sampled to 100 Hz. More details can be found in the following website: http://www.bbci.de/competition/iii/.

DS3: Dataset IIIa from BCI Competition III [37] is denoted as DS3. The original data was recorded over 60 channels with a sample rate of 250 Hz from three participants labeled k3b, k6b, l1b. As shown in Fig. 3(c), a blank screen was displayed in the first 2 s of each trial to instruct participants to rest. An acoustic stimulus and fixation cross were presented in the next 1 s. Then, a visual cue of an arrow pointing left, right, up, or down was displayed for 1 s. From t = 3 to 7 s, participants were asked to perform corresponding MI tasks (left hand, right hand, tongue, or foot movement). In this study, we only use the trials of left and right hand MI tasks. More details also can be found in the following website: http://www.bbci.de/competition/iii/.

### B. Whole Framework

The proposed framework is shown in Fig. 4. In preprocessing stage, we extract data from fixed time windows (as shown in Fig. 3, 2-6 s for DS1, 0-3 s for DS2, and 3-6 s for DS3). In addition, the EEG for each trial is bandpass filtered in
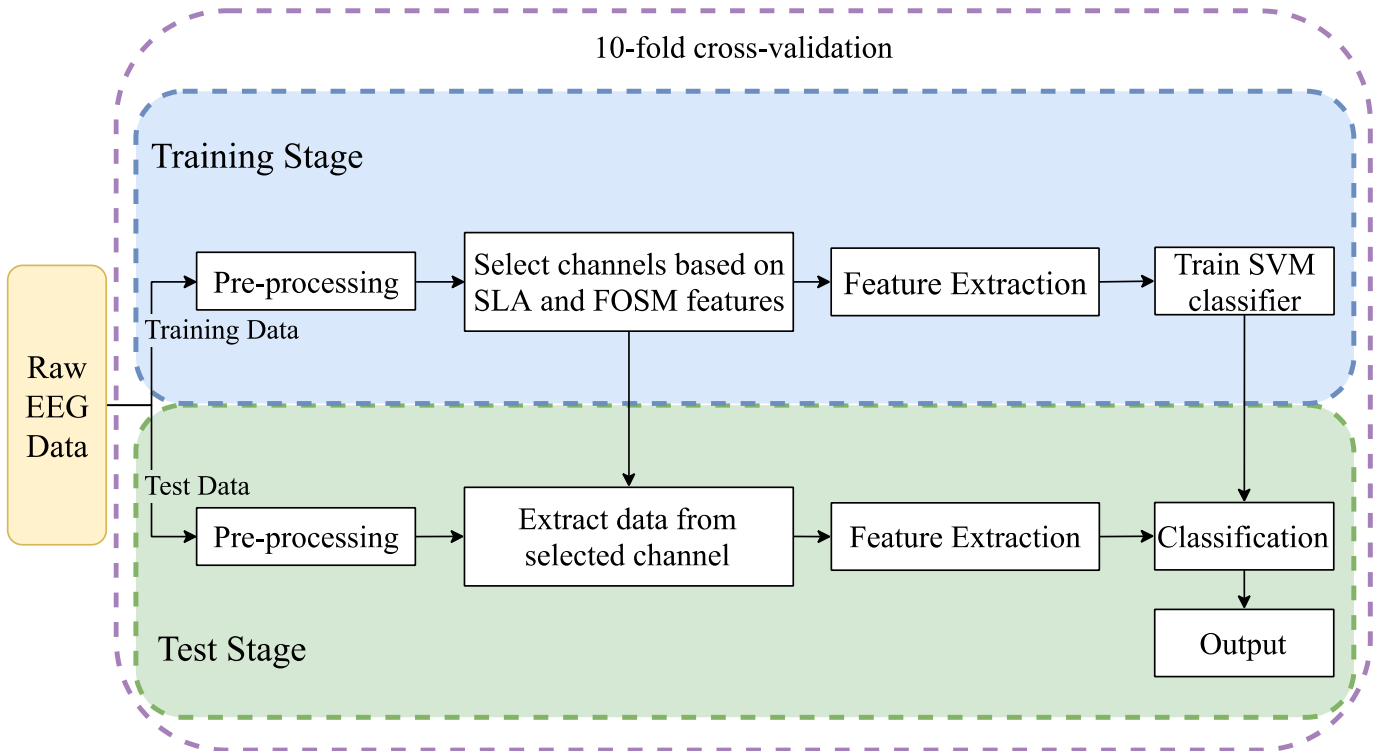
Fig. 4. Block Diagram of the proposed framework. The framework consists of training and test stage. In training stage, training data are used to select channels, train CSP filter and train SVM classifier. In test stage, test data are used to evaluate the performance of the proposed BCS method.

the range of 8 to 30 Hz with a third-order Butterworth filter. Then, we use the bispectrum-based channel selection (BCS) algorithm to select optimal channels. Moreover, CSP and SVM are used for feature extraction and classification of MI tasks, respectively.

## IV. RESULT

### A. Classification Accuracies Comparison

As mentioned above, an experiment is conducted to evaluate the performance of the proposed BCS algorithm. In this study, 10-fold cross validation is applied to avoid overfitting. In addition, 3C-CSP and CSP-rank methods were used to compare with our proposed method.

The 3C-CSP method only uses the data extracted from the Cz, C3, and C4 EEG channels. This method has the advantages of ease of implementation and can be prepared quickly. Nevertheless, it is not appropriate for everyone. For example, for individuals who are not able to control typical BCIs with high levels of accuracy, the 3C-CSP method may cause further deterioration in the performance of the classifier (see Table I'AC-CSP' and '3C-CSP').

Tam *et al.* proposed a channel selection method called CSP-rank in 2011 [38]. This method first calculates the projection matrix $w$ according to Eq. (8). Then, the absolute values of elements of the eigenvectors corresponding to the maximum and minimum eigenvalues were used alternately to select channel. The first channel is selected according to the maximum absolute values of either elements of these two eigenvectors. The index of the maximum absolute member of the eigenvector denotes the channel to select. The second

channel is selected according to the maximum absolute value of the elements of the other eigenvectors. If the channel is already selected, the channel corresponding to the next maximum absolute value will be selected. According this selecting method, all channels will be sorted.

The best classification accuracies for all participants with four different methods are shown in Table I. It may be noted that information on electrode labels and positions is not provided in DS3. Therefore, the 3C-CSP method was not used on DS3.

As can be seen, for all three dataset (e.g., DS1, DS2, and DS3), the proposed BCS method can achieve the highest classification accuracies. Specifically, for four participants from DS1, the mean accuracies are 69.4% (with AC-CSP), 65.4% (with 3C-CSP), 79.1% (with CSP-Rank), and 83.8% (with BCS -CSP), respectively. For five participants from DS2, the mean accuracies are 82.9% (with AC-CSP), 72.1% (with 3C-CSP), 86.3% (with CSP-Rank), and 86.3% (with BCS -CSP), respectively. For three participants from DS3, the mean accuracies are 68.2% (with AC-CSP), 74.5% (with CSP-Rank) and 77.8% (with BCS -CSP), respectively. Among all methods, the performance of the BCS algorithm is superior in ten out of all twelve of the participants. Since the classification accuracies were not meet the normal distribution (p=0.006, Kolmogorov-Smirnov test, BCS-CSP vs. AC-CSP), Wilcoxon signed rank test is performed (using IBM SPSS v.25) to evaluate the significance of this difference. The results confirms the superiority of the BCS method over the AC-CSP and CSP-rank methods (BCS vs. AC-CSP: $Z = -3.059$, $p = 0.002 < 0.05$; BCS vs. CSP-rank: $Z = -2.395$, $p = 0.017 < 0.05$).

TABLE I

ACCURACY (%) AND RESULT OF WILCOXON SIGNED RANK
TESTS COMPARISONS ON DS1, DS2 AND DS3 WITH
FOUR DIFFERENT METHODS

| Dataset | Subjects | Methods | | | |
|---------|----------|---------|---------|----------|---------|
| | | AC-CSP | 3C-CSP | CSP-rank | BCS-CSP |
| DS1 | a | 62.0 | 72.5 | 73.0 | **78.5** |
| | b | 72.0 | 56.5 | 77.0 | **77.5** |
| | f | 87.0 | 68.0 | 89.5 | **92.0** |
| | g | 56.5 | 64.5 | 77.0 | **87.0** |
| | Mean±Std | 69.4±13.4 | 65.4±6.8 | 79.1±7.2 | **83.8±7.0** |
| DS2 | aa | 79.6 | 55.7 | 81.1 | **82.1** |
| | al | 88.6 | 84.3 | 94.3 | **95.0** |
| | av | 68.2 | 60.0 | 71.1 | **72.1** |
| | aw | 90.4 | 76.4 | **92.5** | 90.7 |
| | ay | 87.5 | 83.9 | **92.5** | 91.8 |
| | Mean±Std | 82.9±9.2 | 72.1±13.4 | **86.3±10.0** | **86.3±9.3** |
| DS3 | k3 | 77.2 | - | 81.1 | **85.0** |
| | k6 | 50.8 | - | 61.7 | **65.0** |
| | l1 | 76.7 | - | 80.8 | **83.3** |
| | Mean±Std | 68.2±15.1 | - | 74.5±11.1 | **77.8±11.1** |
| | *p*-value | 0.002 | - | 0.017 | - |

(AC-CSP: using CSP algorithm extract features from all channels)

TABLE II

THE NUMBER OF SELECTED CHANNELS WITH FOUR DIFFERENT
CHANNEL SELECTION METHODS ON DS1, DS2 AND DS3

| Dataset | Subjects | Num | | | |
|---------|----------|--------|--------|----------|---------|
| | | AC-CSP | 3C-CSP | CSP-Rank | BCS-CSP |
| DS1 | a | 59 | 3 | 5 | 12 |
| | b | 59 | 3 | 38 | 40 |
| | f | 59 | 3 | 47 | 33 |
| | g | 59 | 3 | 6 | 13 |
| Average | | 59 | 3 | 24 | 25[*] |
| DS2 | aa | 118 | 3 | 80 | 105 |
| | al | 118 | 3 | 74 | 35 |
| | av | 118 | 3 | 64 | 72 |
| | aw | 118 | 3 | 27 | 68 |
| | ay | 118 | 3 | 68 | 111 |
| Average | | 118 | 3 | 63[*] | 79[*] |
| DS3 | k3 | 60 | - | 45 | 9 |
| | k6 | 60 | - | 13 | 15 |
| | l1 | 60 | - | 23 | 16 |
| Average | | 59 | - | 27 | 14[*] |

* indicates the original number is not an integer. The displayed number is the result of rounding up.

## B. Comparison of the Number of Selected Channels

We also compare the number of selected channels when using the different channel selection methods. Table II shows the number of selected channels using AC-CSP, 3C-CSP, CSP-Rank, and our proposed method. The $*$ indicates that the average selected number of channels is not an integer, and the displayed number is the result of rounding up. The 3C-CSP method was not applied to DS3 because the information about electrode positions is not provided in DS3. As can be seen, to achieve the best classification accuracy, the number of channels selected using the CSP-rank and BCS-CSP methods is similar (24 vs 25 for DS1, 63 vs 79 for DS2, 27 vs 14 for DS3).

## C. Distribution of Selected Channels

For DS1 and DS2, the distribution of the channels selected by the BCS algorithm is discussed. As mentioned before, we are unable to analyze the distribution of selected channel for DS3.

To investigate the distribution of selected channels, we sort all channels based on their F scores, which are calculated according to Eq. (4). The index score (IS) of the channel with the highest F score is the number of all channels (e.g., 59 for DS1 and 118 for DS2). The IS of the channel with the second highest F score is equal to the number of all channels minus
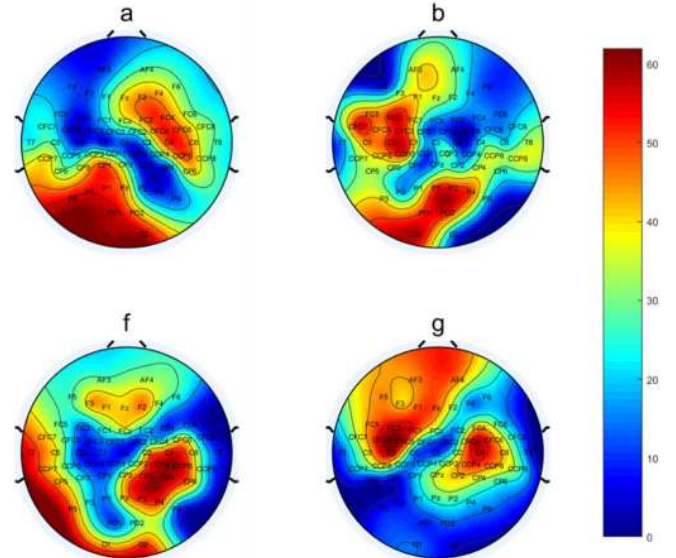


Fig. 5. Topography of index scores of channels selected for DS1 (Participants a, b, f, and g). Color changes from red to blue indicate that the channel is becoming less important.

one, and so on. Hence, for 10-fold validation, all channels can get 10 IS numbers. Then, we sort all channels according to the average IS. The range of average IS values for DS1 and DS2 is 1 to 59 and 1 to 118, respectively. According to the average
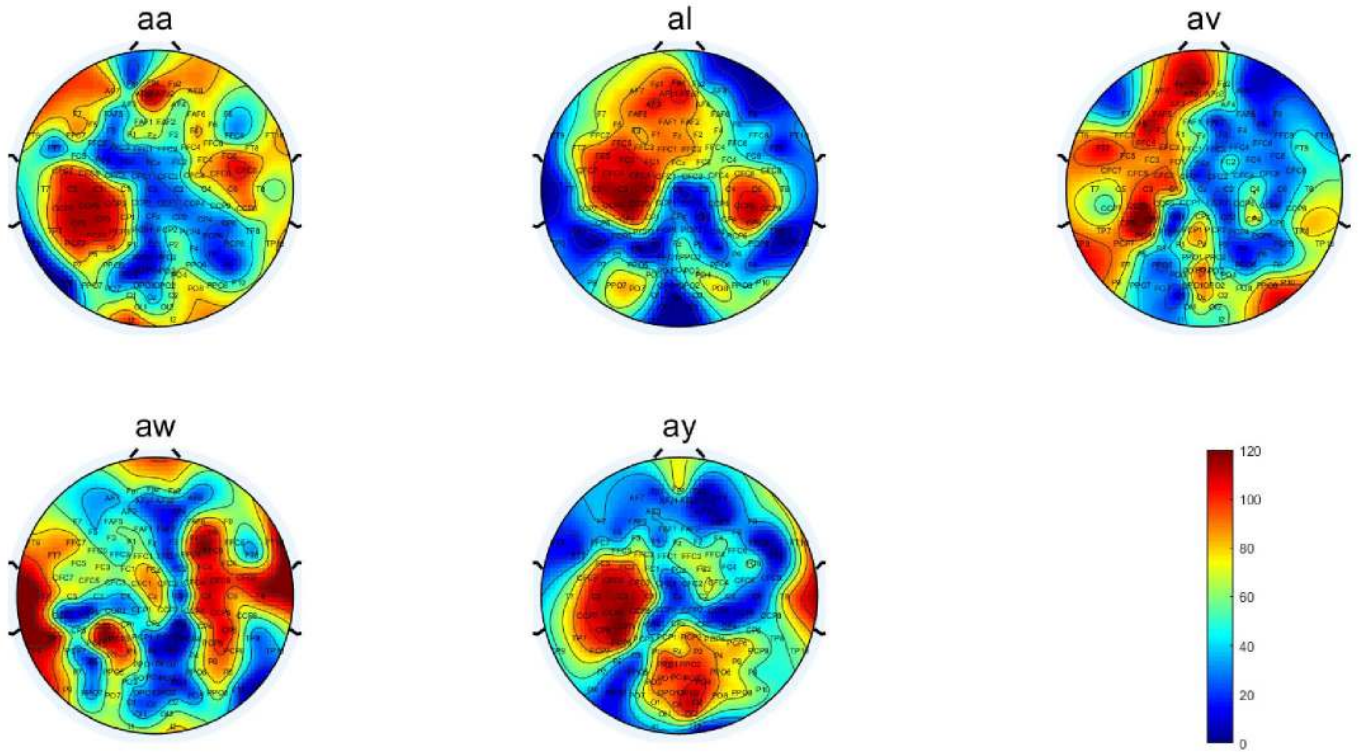
Fig. 6. Topography of index scores of channels in DS2 (Participants aa, al, av, aw, and y). As with Fig. 5, color changes from red to blue indicate that the channel is becoming less important.

IS, we plot the topography for all participants from DS1 and DS2 using MATLAB 2019a with the EEGLAB toolbox [39].

Fig. 5 shows the distribution of selected channels for participants a, b, f, and g from DS1. Red colored areas indicate higher IS values, while blue color areas indicate lower IS values. It can be seen that the channels with high average IS values (red color) for participants a and g are located in the vicinity of both channels CCP3 and CCP4. For participants b and f, the channels with high average IS values are located in the vicinity of channels CCP3 or CCP4.

Fig. 6 shows the distribution of selected channels for participants aa, al, av, aw, and ay from DS2. The meaning of the color in Fig. 6 is the same as in Fig. 5. As can be seen, for participants aa, al, and aw, the channels with high average IS values are also located in the vicinity of channels CCP3 or CCP4. But for participants av and ay, the channels with high average IS values are only located in the vicinity of channel CCP3. Similarities in the distributions of selected channel for DS1 and DS2 suggests that the BCS algorithm could be identifying neurophysiological meaningful patterns in the data.

We also used the mean (83.84) and standard deviation (30.16) of the numbers of the original channels and the mean of the numbers of chosen channels (44.08) to estimate the statistical power [40]. The result shows that we can achieve the statistical power of 0.8696. The relation between the statistical power and the sample size is shown in Fig. 7.

### D. Parameter Sensitivity

As shown in Fig. 8, for each participant, we explore the classification accuracies of the data with different numbers of channels.
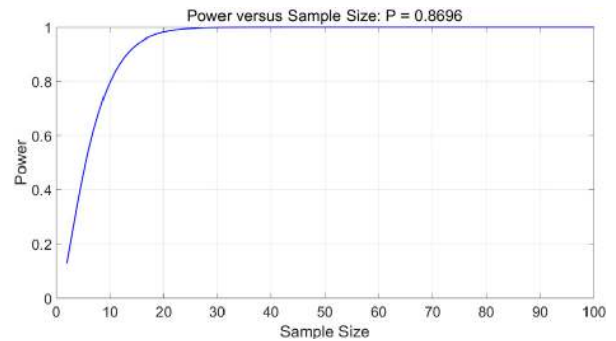


Fig. 7. The relation between the statistical power and the sample size.

For all three datasets, the trend of classification accuracies is the same, increasing first then decreasing. The reason perhaps is the different proportions of useful and redundant information in the selected channels. When using 4 channels for the first time, even if all information in the 4 channels is not redundant, the discriminative feature is not enough for accurate classification. Then, with the number of selected channels increasing, the discriminative features can be extracted from these channels. But redundant information will become dominant if the number of selected channels is too large. For example, the highest classification accuracy for participant g from DS1 is achieved when using around 13 channels. If too many channels are used, the accuracy of the SVM classifier will deteriorate. In addition, participants k3, k6, and l1 from DS3 also show the same phenomenon.

However, the circumstances for participants aa and ay from DS2 are not identical. As can be seen from Table II,
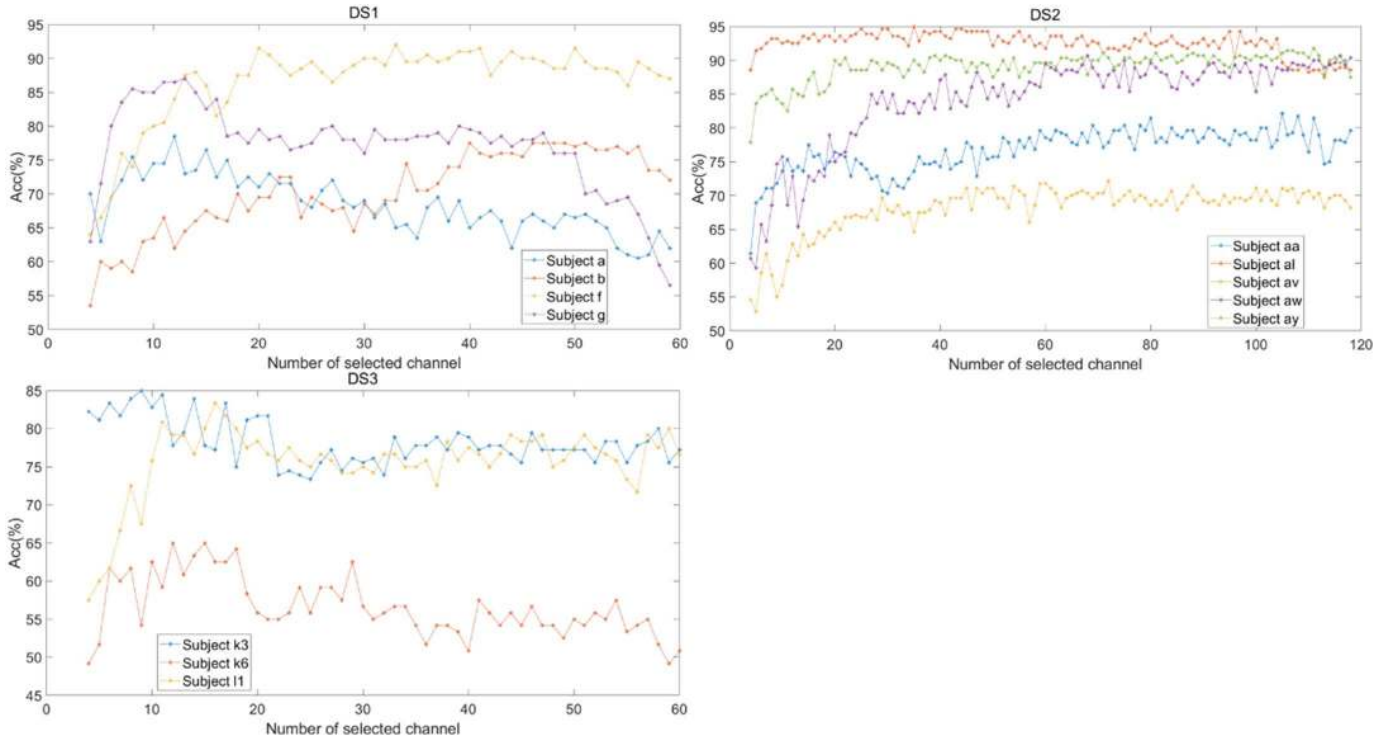
Fig. 8. Classification accuracies of four participants from DS1, DS2 and DS3 with different numbers of selected channels.

the number of channels which can achieve the best classification accuracies for participants aa and ay are 105 and 111, respectively. This may because each channel has only a small amount of useful information. In order to achieve better performance, a larger number of channels must be used to extract discriminative features for the SVM classifier.

### E. Impacts of Different Kinds of Filters and Classifiers

We also used the second-order Chebyshev bandpass filter, the linear discriminate analysis (LDA) and the SVM classifier with linear kernel (SVM-Linear) to investigate the impacts of different kinds of filter and classifiers on BCS algorithm. The bandpass range of the Chebyshev filter is the same as that of the Butterworth filter and the results are shown in Table III and IV.

Table III show the accuracies and the number of selected channels on DS1, DS2 and DS3 using the Butterworth filter, the LDA classifier and the SVM-Linear. As shown in Table III, the BCS algorithm can achieve higher mean accuracies than that of the CSP-rank on DS1 and DS3 for both two classifiers. For DS2, the BCS algorithm has the same mean accuracy (86.9%) when using SVM-Linear classifier whereas when using the LDA classifier, the CSP-rank outperforms the BCS algorithm.

Table IV show the accuracies and the number of selected channels on DS1, DS2 and DS3 using the second-order Chebyshev filter, the LDA classifier, the SVM-Linear and the SVM-RBF. For all three datasets and three classifiers, the mean accuracies achieved by the BCS algorithm are both higher than that achieved by the CSP-rank.

When compared Table III and IV with Table I, it can be seen that different kinds of filters could affect the performance

of MI-BCI (Butterworth filter outperforms Chebyshev filter), however, the impacts of different kinds of classifier will not be great. In addition, in most case, compared with the CSP-rank, the proposed BCS algorithm can achieve higher accuracies.

## V. DISCUSSION

Previous studies have demonstrated that using bispectrum features as classifier inputs in BCI systems can achieve good performance [41]–[43]. Moreover, it has been shown that bispectrum can work especially well for MI BCIs because of the ERD/ERS phenomenon which is most frequently analyzed in the frequency domain [44]. Bispectrum amplitudes can represent the phase relationship between two frequency components [22], whereas sum of logarithmic amplitudes (SLA) and first order spectral moment (FOSM) are two most widely used features.

However, it is still unclear whether bispectrum features can be used to select channels in BCI systems. Thus, in this study, we proposed a channel selection method based on SLA and FOSM features extracted from the bispectrum. The proposed method can directly use the F score calculated by the SLA and FOSM features to sort channels, which is different from the wrapper-based channel selection methods. For example, in [45], authors use the sequential feature selection method to reduce MEG sensors, which is required to use the classifier to evaluate the effect of the combination of each sequential feature whereas the proposed BCS method can sort all channels at once according the F score of each channel.

We have demonstrated the effective of the proposed BCS algorithm in three public BCI competition datasets. As shown in Table I and II, the BCS algorithm can achieve better

| Dataset | Subjects | Butter+LDA | | | | Butter+SVM(Linear) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | CSP-rank | Num | BCS-CSP | Num | CSP-rank | Num | BCS-CSP | Num |
| DS1 | a | 72.5 | 7 | 79.0 | 12 | 72.0 | 7 | 78.5 | 12 |
| | b | 78.5 | 49 | 79.0 | 54 | 80.5 | 49 | 79.5 | 51 |
| | f | 90.0 | 42 | 92.0 | 21 | 89.5 | 42 | 91.5 | 33 |
| | g | 74.5 | 10 | 88.0 | 12 | 74.5 | 6 | 88.0 | 13 |
| | Mean | 78.9 | 27 | **84.5** | 25* | 79.1 | 26 | **84.4** | 28* |
| DS2 | aa | 82.1 | 62 | 81.8 | 99 | 82.9 | 80 | 83.2 | 80 |
| | al | 93.9 | 57 | 95.0 | 51 | 93.9 | 45 | 95.0 | 35 |
| | av | 73.6 | 73 | 72.1 | 70 | 72.5 | 50 | 72.5 | 71 |
| | aw | 93.6 | 95 | 92.1 | 116 | 92.9 | 89 | 91.8 | 114 |
| | ay | 93.2 | 103 | 92.9 | 106 | 92.1 | 30 | 91.8 | 111 |
| | Mean | **87.3** | 78 | 86.8 | 89* | **86.9** | 59* | **86.9** | 83* |
| DS3 | k3 | 79.4 | 30 | 85.0 | 4 | 78.9 | 30 | 85.0 | 11 |
| | k6 | 56.7 | 5 | 70.0 | 12 | 55.0 | 4 | 65.8 | 12 |
| | l1 | 81.7 | 32 | 81.7 | 15 | 81.7 | 23 | 82.5 | 51 |
| | Mean | 72.6 | 23* | **78.9** | 11* | 71.9 | 19 | **77.8** | 25* |

\* indicates the original number is not an integer. The displayed number is the result of rounding up.

| Dataset | Subjects | Che+SVM(RBF) | | | | Che+SVM(Linear) | | | | Che+LDA | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | CSP-rank | Num | BCS-CSP | Num | CSP-rank | Num | BCS-CSP | Num | CSP-rank | Num | BCS-CSP | Num |
| DS1 | a | 62.0 | 5 | 72.5 | 19 | 61.5 | 17 | 72.5 | 12 | 64.0 | 17 | 74.0 | 24 |
| | b | 56.0 | 8 | 55.0 | 37 | 55.5 | 8 | 57.0 | 36 | 57.0 | 8 | 56.5 | 39 |
| | f | 68.5 | 6 | 64.5 | 27 | 67.0 | 6 | 65.0 | 18 | 64.0 | 4 | 65.5 | 19 |
| | g | 73.0 | 10 | 77.0 | 22 | 70.0 | 10 | 77.5 | 20 | 71.0 | 10 | 77.5 | 21 |
| | Mean | 64.9 | 8* | **67.3** | 27* | 63.5 | 11* | **68.0** | 22* | 64.0 | 10* | **68.4** | 26* |
| DS2 | aa | 62.5 | 4 | 73.6 | 14 | 62.9 | 4 | 74.3 | 16 | 60.0 | 4 | 75.0 | 16 |
| | al | 90.7 | 42 | 93.6 | 50 | 91.1 | 42 | 93.6 | 50 | 90.7 | 42 | 93.6 | 36 |
| | av | 61.4 | 4 | 56.4 | 22 | 61.8 | 4 | 56.4 | 12 | 62.5 | 4 | 55.4 | 20 |
| | aw | 78.2 | 35 | 76.1 | 24 | 77.9 | 35 | 75.0 | 24 | 78.6 | 34 | 74.6 | 24 |
| | ay | 56.1 | 6 | 81.4 | 40 | 55.7 | 53 | 82.5 | 40 | 59.6 | 4 | 82.1 | 21 |
| | Mean | 69.8 | 19* | **76.2** | 30 | 69.9 | 28* | **76.4** | 29* | 70.3 | 18* | **76.1** | 24* |
| DS3 | k3 | 63.3 | 4 | 86.1 | 7 | 63.3 | 4 | 85.6 | 7 | 65.0 | 4 | 85.6 | 7 |
| | k6 | 55.8 | 22 | 56.7 | 16 | 55.0 | 22 | 59.2 | 16 | 55.8 | 33 | 57.5 | 9 |
| | l1 | 77.5 | 16 | 81.7 | 31 | 77.5 | 16 | 82.5 | 31 | 77.5 | 55 | 81.6 | 29 |
| | Mean | 65.5 | 14 | **74.8** | 18 | 65.3 | 14 | **75.8** | 18 | 66.1 | 31* | **74.9** | 15 |

\* indicates the original number is not an integer. The displayed number is the result of rounding up.

classification performance with fewer channels. Compare with other state-of-the-art channel selection method such as CSP-rank method, Wilcoxon signed rank test results also show that the BCS algorithm can achieve significantly better classification accuracies ($p = 0.017 < 0.05$). Compared with the traditional 3C-CSP channel selection method, the proposed method aslo shows the superiority in the classification accuracies (83.8% vs. 65.4% on DS1, 86.3% vs. 72.1% on DS2).

The proposed BCS method also can be used in real-time BCI systems. Specifically, before the real-time BCI systems run, the training data of EEG signals need to be used to calculate the bispectrum for channel selection.

Our proposed method also has some limitations. On the one hand, according to the formula (4), SLA and FOSM features had the same contribution to the process of F score calculation. That is to say, the selected channels were equally affected by the SLA and FOSM features. But the SLA and FOSM are two different features extracted from bispectrum (formula (2) and (3)), and using different weights for SLA and FOSM features may have better performance. On the other hand, the effect of different frequency ranges of the filter on the calculation of bispectrum is not to be considered. Although our proposed method achieves good performance, the frequency range of the filter may affect the BCS algorithm.

Xing *et al.* first proposed a metric learning method to improve clustering performance [46]. Recently, Sheoran *et al.* used various distance metric learning algorithms to map features to a discriminant space, improving the accuracy of epileptic seizure detection [47]. Mapping the SLA and FOSM feature to a discriminant space may be helpful for identifying which channels contain redundant information. Therefore, we will try to combine the metric learning methods to further improve the performance of the BCS algorithm in future work. We will also investigate the different impacts of the SLA, FOSM features and the influences of different frequency ranges of the bandpass filter on the channel selection method in MI based BCIs.

## VI. Conclusion

In this paper, we proposed a bispectrum-based channel selection (BCS) method for MI based BCIs. SLA and FOSM features, extracted from bispectrum, were used to calculate the F score for each channel. According to the F score, channels without redundant information are selected. Compared to the 3C-CSP and CSP-rank methods, the proposed BCS algorithm can achieve significantly better results. Overall, the BCS algorithm is an efficient method to improve the performance of MI based BCIs.

## References

[1] R. Ortner, B. Z. Allison, G. Korisek, H. Gaggl, and G. Pfurtscheller, "An SSVEP BCI to control a hand orthosis for persons with tetraplegia," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 1, pp. 1–5, Feb. 2011.

[2] R. Leeb, D. Friedman, G. R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller, "Self-paced (Asynchronous) BCI control of a wheelchair in virtual environments: A case study with a tetraplegic," *Comput. Intell. Neurosci.*, vol. 2007, Sep. 2007, Art. no. 079642.

[3] M. Ahn *et al.*, "Gamma band activity associated with BCI performance: Simultaneous MEG/EEG study," *Frontiers Human Neurosci.*, vol. 7, p. 848, 2013.

[4] J. Jin, E. W. Sellers, S. Zhou, Y. Zhang, X. Wang, and A. Cichocki, "A P300 brain–computer interface based on a modification of the mismatch negativity paradigm," *Int. J. Neural Syst.*, vol. 25, no. 3, 2015, Art. no. 1550011.

[5] J. Jin, H. Zhang, I. Daly, X. Wang, and A. Cichocki, "An improved P300 pattern in BCI to catch user's attention," *J. Neural Eng.*, vol. 14, no. 3, 2017, Art. no. 36001.

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

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

[8] Q. Liu *et al.*, "Efficient representations of EEG signals for SSVEP frequency recognition based on deep multiset CCA," *Neurocomputing*, vol. 378, pp. 36–44, Feb. 2020.

[9] L. Shao *et al.*, "EEG-controlled wall-crawling cleaning robot using SSVEP-based brain-computer interface," *J. Healthcare Eng.*, vol. 2020, pp. 1–11, Jan. 2020.

[10] Y. Zhang, Y. Wang, J. Jin, and X. Wang, "Sparse Bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification," *Int. J. Neural Syst.*, vol. 27, no. 02, Mar. 2017, Art. no. 1650032.

[11] S.-L. Wu *et al.*, "Fuzzy integral with particle swarm optimization for a motor-imagery-based brain–computer interface," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 1, pp. 21–28, Feb. 2017.

[12] C. Chen *et al.*, "G-causality brain connectivity differences of finger movements between motor execution and motor imagery," *J. Healthcare Eng.*, vol. 2019, pp. 1–12, Oct. 2019.

[13] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and dempster-shafer theory," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 24, 2020, doi: 10.1109/TNNLS.2020.3015505.

[14] F. Velasco-Álvarez, R. Ron-Angevin, L. da Silva-Sauer, and S. Sancha-Ros, "Audio-cued motor imagery-based brain–computer interface: Navigation through virtual and real environments," *Neurocomputing*, vol. 121, pp. 89–98, Dec. 2013.

[15] T. Yu, Y. Li, J. Long, and Z. Gu, "Surfing the Internet with a BCI mouse," *J. Neural Eng.*, vol. 9, no. 3, 2012, p. 36012.

[16] G. Pfurtscheller and C. Neuper, "Future prospects of ERD/ERS in the context of brain–computer interface (BCI) developments," *Prog. Brain Res.*, vol. 159, pp. 433–437, 2006.

[17] Z. Qiu, J. Jin, H.-K. Lam, Y. Zhang, X. Wang, and A. Cichocki, "Improved SFFS method for channel selection in motor imagery based BCI," *Neurocomputing*, vol. 207, pp. 519–527, Sep. 2016.

[18] Y. Yang, I. Bloch, S. Chevallier, and J. Wiart, "Subject-specific channel selection using time information for motor imagery brain–computer interfaces," *Cogn. Comput.*, vol. 8, no. 3, pp. 505–518, 2016, doi: 10.1007/s12559-015-9379-z.

[19] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Netw.*, vol. 118, pp. 262–270, Oct. 2019, doi: 10.1016/j.neunet.2019.07.008.

[20] N. Kotoky and S. M. Hazarika, "Bispectrum analysis of EEG for motor imagery classification," in *Proc. Int. Conf. Signal Process. Integr. Netw. (SPIN)*, New Delhi, India, Feb. 2014, pp. 581–586.

[21] R. Alazrai, F. Alqasem, S. Alaarag, K. M. Ahmad Yousef, and M. I. Daoud, "A bispectrum-based approach for detecting deception using EEG signals," in *Proc. IEEE 20th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Sep. 2018, pp. 1–6.

[22] L. Sun, Z. Feng, N. Lu, B. Wang, and W. Zhang, "An advanced bispectrum features for EEG-based motor imagery classification," *Expert Syst. Appl.*, vol. 131, pp. 9–19, Oct. 2019, doi: 10.1016/j.eswa.2019.04.021.

[23] R. J. Gajraj, M. Doi, H. Mantzaridis, and G. N. Kenny, "Analysis of the EEG bispectrum, auditory evoked potentials and the EEG power spectrum during repeated transitions from consciousness to unconsciousness," *Brit. J. Anaesthesia*, vol. 80, no. 1, pp. 46–52, Jan. 1998.

[24] C. L. Nikias and M. R. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *Proc. IEEE*, vol. 75, no. 7, pp. 869–891, Jul. 1987.

[25] F. Feng, A. Si, and H. Zhang, "Research on fault diagnosis of diesel engine based on bispectrum analysis and genetic neural network," *Procedia Eng.*, vol. 15, pp. 2454–2458, 2011.

[26] S.-M. Zhou, J. Q. Gan, and F. Sepulveda, "Classifying mental tasks based on features of higher-order statistics from EEG signals in brain–computer interface," *Inf. Sci.*, vol. 178, no. 6, pp. 1629–1640, 2008.

[27] S. Bordoloi, U. Sharmah, and S. M. Hazarika, "Motor imagery based BCI for a maze game," in *Proc. 4th Int. Conf. Intell. Human Comput. Interact. (IHCI)*, Dec. 2012, pp. 1–6.

[28] W. B. Collis, P. R. White, and J. K. Hammond, "Higher-order spectra: The bispectrum and trispectrum," *Mech. Syst. Signal Process.*, vol. 12, no. 3, pp. 375–394, May 1998.

[29] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. L. Da Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.

[30] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3322–3332, Sep. 2019.

[31] M. A. Ali, D. Ucuncu, P. K. Atas, and S. Ozogur-Akyuz, "Classification of motor imagery task by using novel ensemble pruning approach," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 1, pp. 85–91, Jan. 2020.

[32] D. Wu, J.-T. King, C.-H. Chuang, C.-T. Lin, and T.-P. Jung, "Spatial filtering for EEG-based regression problems in brain–computer interface (BCI)," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 771–781, Apr. 2018.

[33] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27, 2011.

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

[35] C. Chen *et al.*, "The mixed kernel function SVM-based point cloud classification," *Int. J. Precis. Eng. Manuf.*, vol. 20, no. 5, pp. 737–747, May 2019.

[36] M. Tangermann *et al.*, "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, p. 55, 2012, doi: 10.3389/fnins.2012.00055.

[37] B. Blankertz *et al.*, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.

[38] W.-K. Tam, Z. Ke, and K.-Y. Tong, "Performance of common spatial pattern under a smaller set of EEG electrodes in brain-computer interface on chronic stroke patients: A multi-session dataset study," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 6344–6347.

[39] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[40] A. A. Frolov *et al.*, "Post-stroke rehabilitation training with a Motor-Imagery-Based brain-computer interface (BCI)-controlled hand exoskeleton: A randomized controlled multicenter trial," *Frontiers Neurosci.*, vol. 11, p. 400, Jul. 2017.

[41] O. Trigui, W. Zouch, M. Ben Slima, and M. Ben Messaoud, "Bispectral analysis-based approach for steady-state visual evoked potentials detection," *Multimedia Tools Appl.*, vol. 78, no. 10, pp. 12865–12882, May 2019, doi: 10.1007/s11042-018-6029-y.

[42] S. Shahid and G. Prasad, "Bispectrum-based feature extraction technique for devising a practical brain–computer interface," *J. Neural Eng.*, vol. 8, no. 2, p. 25014, 2011.

[43] S. Rimbert, P. Riff, N. Gayraud, D. Schmartz, and L. Bougrain, "Median nerve stimulation based BCI: A new approach to detect intraoperative awareness during general anesthesia," *Frontiers Neurosci.*, vol. 13, p. 622, Jun. 2019.

[44] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, Nov. 1999.

[45] A. N. Belkacem, S. Nishio, T. Suzuki, H. Ishiguro, and M. Hirata, "Neuromagnetic decoding of simultaneous bilateral hand movements for multidimensional brain–machine interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 6, pp. 1301–1310, Jun. 2018.

[46] E. P. Xing, M. I. Jordan, S. J. Russell, and A. Y. Ng, "Distance metric learning with application to clustering with side-information," in *Proc. Adv. Neural Inf. Process. Syst.*, 2003, pp. 521–528.

[47] P. Sheoran, N. Rathee, and J. S. Saini, "Epileptic seizure detection using bidimensional empirical mode decomposition and distance metric learning on scalogram," in *Proc. 7th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2020, pp. 675–680.