

An advanced bispectrum features for EEG-based motor imagery classification

Lei Sun*, Zuren Feng, Na Lu, Beichen Wang, Wenjun Zhang

School of the Electronic and Information Engineering, Xi'an Jiao Tong University, Xi'an 710049, China



ARTICLE INFO

Article history:

Received 26 August 2018
Revised 27 March 2019
Accepted 9 April 2019
Available online 18 April 2019

Keywords:

Motor imagery
Event-related potentials
Variations
Bispectrum
Features extraction
Classification

ABSTRACT

Motor imagery (MI)-related brain activities can be effectively described by frequency analysis. Bispectrum is developed to overcome the drawback of power spectrum that the estimation of power spectrum discards the phase relationship among frequency components. However, the widely used bispectral features extraction method adds up all bispectral values as one feature, which could lead to the loss of effective information and increase of the sensitivity to non-linear and non-Gaussian noises. Thus, the representative bispectral features extraction method may be inefficient for MI classification. In addition, recent research suggests that the variations of EEG signals could provide more useful underlying information of event-related brain responses. This paper presents an advanced variations based bispectral feature extraction method to improve the performance of MI classification. The proposed method calculates the variations of MI-related EEG signals as input to bispectrum estimation. Besides, a new segmented bispectral sum features are developed to reduce the influence of non-linear and non-Gaussian noises and emphasize the valuable information for MI classification. The dataset collected in our laboratory and BCI Competition IV dataset 2b were adopted to validate the proposed method. The results indicate that the proposed method outperforms the power spectrum based methods and the representative bispectral features based methods. Moreover, compared to other state-of-the-art works, our approach also achieves the greater performance for MI classification.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Brain-computer interface (BCI) technology is dedicated to help patients with brain diseases, such as Amyotrophic lateral sclerosis (ALS), cerebral palsy, and motor neurons disease (MND). Thanks to the advantages of non-invasion, convenience, low cost, and low power consumption, Electroencephalogram (EEG) is universally applied as a collection method for BCI system (Allison, Wolpaw, & Wolpaw, 2007; Wolpaw, Birbaumer, Mcfarland, Pfurtscheller, & Vaughan, 2002). In practice, based on event-related synchronization (ERS) and event-related desynchronization (ERD) phenomena, researchers can analyze and recognize motor imagery (MI)-related EEG signals. ERD/ERS is induced by the execution or imagination of movement, and both phenomena represent as the changes in oscillatory EEG power. Hence, ERD/ERS can be characterized with temporal or frequency analysis thereby classifying MI tasks (Neuper & Pfurtscheller, 2001).

Common Spatial Pattern (CSP) algorithm constructs an optimal spatial filters and extracts the temporal features of spatial patterns in different MI tasks (Ang, Chin, Zhang, & Guan, 2008). Zhang et al. (2018) applies CSP to extract temporal features and combines a multi-kernel Extreme learning machine (MKELM)-based method to classify MI tasks. Alvarez-Meza, Velasquez-Martinez, and Castellanos-Dominguez (2015) makes use of CSP and motor imagery discrimination using feature relevance analysis (MIDFR) to improve the performance of MI classification. CSP algorithm is relied heavily on its operational frequency band. Therefore, using CSP features may lead to poor results when MI-related EEG signals are filtered inappropriately (Ang, Chin, Wang, Guan, & Zhang, 2012). As another temporal features extraction method, Hjorth algorithm is also used for MI classification frequently (Gandhi, Prasad, Coyle, Behera, & McGinnity, 2015). However, ERD/ERS phenomena may be insufficiently represented with temporal analysis, but could be detected by frequency analysis (Pfurtscheller & Lopes da Silva, 1999). Power spectrum based frequency methods are commonly adopted to extract MI-related features. Specifically, Herman, Prasad, McGinnity, and Coyle (2008) extracts power spectral density (PSD) features and uses linear and nonlinear classifier to classify the MI tasks of left and right hands. Saa and Çetin (2012) utilizes Burg's method to

* Corresponding author.

E-mail addresses: solayi@stu.xjtu.edu.cn (L. Sun), fzr9910@mail.xjtu.edu.cn (Z. Feng), lvna2009@mail.xjtu.edu.cn (N. Lu), ywawang1700@163.com (B. Wang), 876196430@qq.com (W. Zhang).

estimate power spectrum features, and employs a hidden conditional random field method to enhance MI classification accuracy (CA). However, by the reason of the non-linearity and non-Gaussianity of EEG signals, traditional power spectrum technique is limited for the analysis of MI-related EEG signals. Since power spectrum discards the information of the phase relationship among frequency components, it cannot extract the useful non-linear and non-Gaussian information for MI classification. Bispectrum is proposed to solve the problem of power spectrum. Bispectrum can quantify the interaction of two frequency components in a non-linear and non-Gaussian signal, thus it can unearth more non-linear and non-Gaussian features from MI-related EEG signals. Zhou, Gan, and Sepulveda (2008) estimates the sums of power spectrum and sum of logarithmic amplitudes of bispectrum for two classes MI classification. Shahid and Prasad (2011) normalizes the sum of logarithmic amplitudes of bispectrum and adopts Fisher's Linear Discriminant Analysis (LDA) classifier to discriminate the MI tasks of right and left hands. However, the representative bispectral feature extraction method used in above works adds up all bispectral values as one feature, so it could reduce the validity of a classifier and deteriorate the performance of MI classification. Moreover, since the sensitivity to non-linear and non-Gaussian noises could be highly enhanced within one bispectral feature value (Chella, Marzetti, Pizzella, Zappasodi, & Nolte, 2014), the representative bispectral feature extraction method could be seriously contaminated by non-linear and non-Gaussian noises. On the other hand, although event-related potentials (ERPs) show the differences in peak and amplitudes, the differences in undergoing temporal superimposed undulations and additional local changes of ERPs may more accurately reflect underlying brain activities (Klein & Skrandies, 2017). Hence, compared to ERPs themselves, the variations of ERPs could provide more valuable information regarding MI-related brain activities.

In light of Klein and Skrandies (2017), we present an advanced features extraction method, called variations based segmented bispectrum sum (VSBS), to enhance the separability of bispectral features. The proposed method calculates the variations of raw EEG signals to extract more underlying temporal information for MI classification, and then estimates the sums of segmented bispectrum to reduce the influence of non-linear and non-Gaussian noises. Next, an optimal bispectral segment length is learned from labeled training data with a fivefold cross-validation to improve the performance of MI classification. The proposed VSBS method was applied on the dataset collected in our laboratory and BCI competition IV dataset 2b. The results demonstrate surpassing performance compared to the PSD based methods and representative bispectral feature based methods. Furthermore, our method also yields the best results compared to other state-of-the-art works.

The rest of this paper is organized as follows: Section 2 introduces the related researches and Section 3 describes the proposed VSBS method. Experiments materials and data processing are detailed in Sections 4 and 5, respectively. Section 6 illustrates the experimental results. Finally, the discussion and conclusions are outlined in Sections 7 and 8.

2. Related research

2.1. The definition of the cumulants and bispectrum

Higher order spectra (HOS) serves to reveal the interaction between the frequency components of a non-linear and non-Gaussian signal, and it includes higher order cumulants and their spectra.

By assuming a series of n real random variables $\{x_1, x_2, \dots, x_n\}$, the order $r = r_1 + r_2 + \dots + r_n$ cumulants of this series are

defined as

$$c_{r_1 r_2 \dots r_n} \triangleq (-j)^r \frac{\partial^r \ln \Phi(w_1, w_2 \dots w_n)}{\partial w_1^{r_1} \partial w_2^{r_2} \dots \partial w_n^{r_n}} \Big|_{w_1=w_2=\dots=w_n=0}, \quad (1)$$

where

$$\Phi(w_1, w_2 \dots w_n) = E\{\exp[j(w_1 x_1 + w_2 x_2 + \dots + w_n x_n)]\}. \quad (2)$$

The N th-order spectrum $S_N(w_1, w_2, \dots, w_{N-1})$ is calculated as the Fourier Transform (FT) of the N th-order cumulants $c_N(\tau_1, \tau_2, \dots, \tau_{N-1})$, and the formula is presented as:

$$S_N(w_1, w_2, \dots, w_{N-1}) = \sum_{\tau_1=-\infty}^{+\infty} \dots \sum_{\tau_{N-1}=-\infty}^{+\infty} c_N(\tau_1, \tau_2, \dots, \tau_{N-1}) \cdot \exp\{-j(w_1 \tau_1 + w_2 \tau_2 + \dots + w_{N-1} \tau_{N-1})\}. \quad (3)$$

When $N=2$, the second-order spectrum is calculated as below:

$$S_2(w) = \sum_{\tau_1=-\infty}^{+\infty} c_2(\tau) \cdot \exp\{-j(w\tau)\} \quad (4)$$

The second-order spectrum is equal to power spectrum while a signal $x(n)$ is deterministic and zero mean. When $N=3$, the third-order spectrum, called bispectrum, is computed as follow:

$$S_3(w_1, w_2) = \sum_{\tau_1=-\infty}^{+\infty} \sum_{\tau_2=-\infty}^{+\infty} c_3(\tau_1, \tau_2) \cdot \exp\{-j(w_1 \tau_1 + w_2 \tau_2)\}. \quad (5)$$

For a deterministic, zero mean signal $x(n)$, its bispectrum can be expanded as:

$$S_3(w_1, w_2) = \sum_{\tau_1=-\infty}^{+\infty} \sum_{\tau_2=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} x(n)x(n+\tau_1)x(n+\tau_2) \cdot \exp\{-j(w_1 \tau_1 + w_2 \tau_2)\}. \quad (6)$$

When $n + \tau_1 = m$ and $n + \tau_2 = k$, then

$$S_3(w_1, w_2) = \left\{ \sum_{m=-\infty}^{+\infty} x(m)e^{-jw_1 m} \right\} \left\{ \sum_{k=-\infty}^{+\infty} x(k)e^{-jw_2 k} \right\} \times \left\{ \sum_{n=-\infty}^{+\infty} x(n)e^{j(w_1+w_2)n} \right\}, \quad (7)$$

and the bispectrum can be estimated by

$$B(w_1, w_2) = X(w_1)X(w_2)X^*(w_1 + w_2), \quad (8)$$

where $B(w_1, w_2)$ is the bispectrum in the bi-frequency (w_1, w_2) , $X(w)$ is the discrete time FT of the signal $x(n)$, $(*)$ is complex conjugate.

In practice, bispectrum estimation includes two technique, non-parametric model based on Fourier transformation and parametric model based on autoregressive model (AR), moving average (MA), autoregressive and moving average (ARMA) or Volterra model. Nonparametric technique can be employed with direct and indirect methods (Nikias & Raghuever, 1987). Since direct method is easier to implement and less computational cost (Feng, Si, & Zhang, 2011), we utilize the direct method of nonparametric technique to estimate bispectrum in this work.

The third-order cumulants are identified as skewness of time series, and it can segregate Gaussian and non-Gaussian signals for the reason that the third-order cumulants of Gaussian distribution is zero (Feng et al., 2011; Nikias & Raghuever, 1987). Therefore, as the FT of the third-order cumulants, the bispectrum based method could increase signal-to-noise ratio when a signal is contaminated by Gaussian noises. Moreover, bispectrum technique is able to detect and quantify the quadratic phase coupling (QPC), which is the interaction of phases among two harmonic components for a non-linear and non-Gaussian process (Schwab, Eiselt, Schelenz, & Witte, 2005). Hence, compared to power spectrum, bispectrum is more appropriate to process non-linear and non-Gaussian signals, especially MI-related EEG signals.

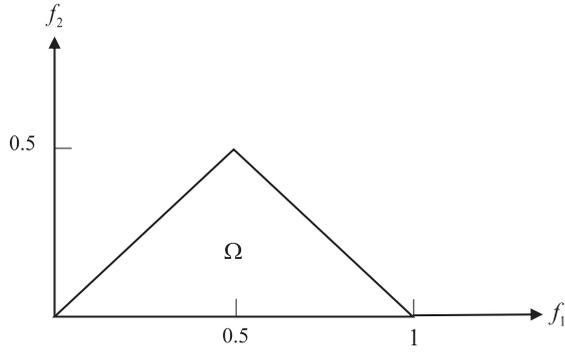


Fig. 1. Principal domain (Ω) of bispectrum estimation.

2.2. Bispectrum feature extraction

The power spectrum of a real-valued signal holds symmetry due to the conjugate symmetry of FT. Similarly, the bispectrum of a real-valued signal also has symmetrical properties as follow:

$$\begin{aligned} B(f_1, f_2) &= B(f_2, f_1) = B^*(-f_1, -f_2) \\ &= B^*(-f_2, -f_1) = B(-f_1 - f_2, f_2) \\ &= B(f_1, -f_1 - f_2) = B(-f_1 - f_2, f_1), \\ &= B(f_2, -f_1 - f_2) \end{aligned} \quad (9)$$

where the f is normalized by the Nyquist frequency. Hence, bispectrum estimation only needs to calculate the bispectral values on the principal domain (non-redundant area). The principal domain is uniquely defined with a triangle $0 \leq f_2 \leq f_1 \leq f_1 + f_2 \leq 1$ as shown in Fig. 1 and denoted as Ω (Collis, White, & Hammond, 1998; Nikias & Raghuveer, 1987).

Considering the bispectral amplitudes on the principal domain can represent phase relationship between two frequency components f_1 and f_2 , and between $f_1 + f_2$ or $f_1 - f_2$ (Helbig, Schwab, Leistritz, Eiselt, & Witte, 2006), many works propose several bispectral features derived from the moments to detect non-linear and non-Gaussian information (Sahayadhas, Sundaraj, Murugappan, & Palaniappan, 2015). The widely adopted derived bispectral feature is the sum of logarithmic amplitudes of bispectrum, and its calculation follows below formula (Shahid & Prasad, 2011):

$$H_1 = \sum_{f_1, f_2 \in \Omega} \log(|B(f_1, f_2)|), \quad (10)$$

where the $B(f_1, f_2)$ is the 2D bispectrum on the principal domain. Since the representative bispectral features is a bispectrum sum, we name its features extraction method as a bispectrum sum (ABS) method in this paper.

3. Proposed method

ABS method is proved effective in machine faults detection and classification (Feng et al., 2011), but it may be inefficient for MI classification. Since ABS method produces only one bispectral feature value, it could deteriorate the performance of a classifier. Furthermore, the sensitivity to non-linear and non-Gaussian noises could be increased within one bispectral feature value, so the valuable information for MI classification may be fully covered by noises with ABS method. On the other hand, Klein and Skrandies (2017) states that the variations of ERPs contain more underlying important information of event-related brain dynamics than ERPs, thus the temporal variations of EEG signal could further improve the performance of MI classification.

Considering above issues, we propose an advanced bispectral feature extraction method called variations based segmented bispectrum sums (VSBS). In order to extract more useful underlying

information for MI classification, the VSBS method uses the variations of MI-related EEG signals as the input to bispectrum analysis. In addition, the sums of logarithmic amplitudes of bispectral segments are computed to reduce the influence of non-linear and non-Gaussian noises, and a weight is assigned to every bispectral segment to emphasize valuable features for MI classification. The proposed method is performed as follows steps:

- (1) Let $x(n)$, $n = 1, 2, \dots, N$ is a discrete EEG signal, so the variations of the EEG signal can be calculated as:

$$v(i) = |x(i+1) - x(i)|, i = 1, \dots, N - 1. \quad (11)$$

- (2) The variations $v(i)$ is sent to bispectrum estimation as below:

$$VB(f_1, f_2) = V(f_1)V(f_2)V^*(f_1 + f_2), \quad (12)$$

where $VB(f_1, f_2)$ is the bispectrum of the variations $v(i)$, $V(f)$ is the discrete FT of $v(i)$, and f is normalized by the Nyquist frequency.

- (3) The 1D horizontal slice of bispectrum on the principal domain is estimated and divided into segments. Then, the variations based segmented bispectrum sum of the m th segment is calculated as:

$$B_m = u \sum_{\lambda=1}^L \sum_{f_2 \in \Omega} \log(|VB(\lambda, f_2)|), \quad (13)$$

$$u = \frac{\sum_{\lambda=1}^L \sum_{f_2 \in \Omega} \log(|B(\lambda, f_2)|)}{\sum_{f_1, f_2 \in \Omega} \log(|B(f_1, f_2)|)}. \quad (14)$$

where L is the length of each bispectral segment, u is the weight of each bispectral segment.

- (4) Through the training data, an optimal bispectral segment length is learned. After all the sums of bispectral segments are calculated, the final VSBS feature is presented as a series $[VB_1, VB_2, \dots, VB_M]$, where M is the number of bispectral segments.

4. Materials and experiment

4.1. Dataset 1

4.1.1. Subjects and equipment

Five subjects (22–26 years old males) participated in this study, and all five subjects are healthy and right-handed. G.tec (Guger Technologies OEG Austria) BCI device was utilized to collect MI-related EEG signals. The EEG signals were recorded from two bipolar electrodes (C3 and C4), and the electrodes scheme were arranged according to the international 10–20 system. The sampling frequency was set to 256 Hz with a notch filter at 50 Hz. The electrode position Fz served as EEG ground.

4.1.2. Experiment protocol

The subject were guided to perform two MI tasks (the imagination of left hand and right hand) and asked to relax their arms and hands during the experiment (see Fig. 2a). The collected experimental data for each subject consisted of two sessions, and the first session was selected as training data while the second session was used as test data. The time scheme of experimental paradigm is described in Fig. 2b. Each trial began with a fixation green cross presented in the center of the screen. After 1 s, a visual cue (an black arrow pointing to the left or right) was appeared for 4 s. During this 4 s, the subject needed to imagine a movement of left or right hand by following the direction of the black arrow. A short break (less than 4 s) was added after MI period to avoid adaptation.

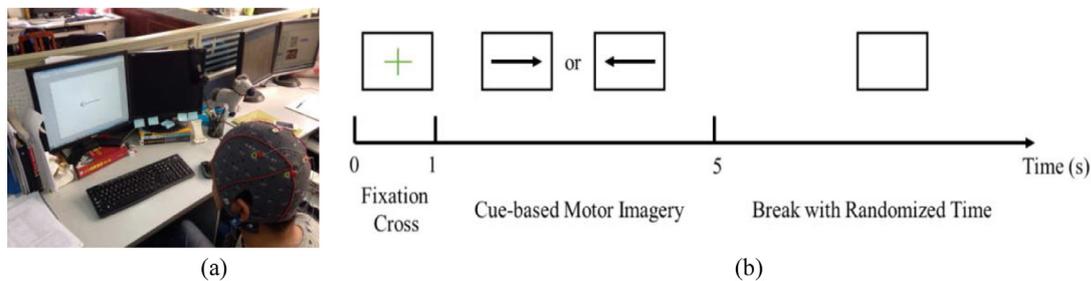


Fig. 2. (a) Experimental setup. (b) Time scheme of the experimental paradigm.

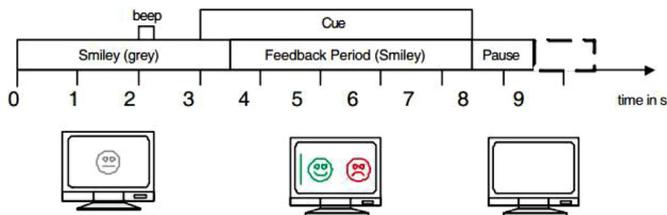


Fig. 3. Time scheme of the paradigm with smiley feedback.

4.2. Dataset 2

BCI Competition IV dataset 2b is universally used for the research of MI classification. Dataset 2b is provided by the Technical University of Graz (TUG) and includes two classes MI tasks involving left and right hands (Tangermann et al., 2012). In this dataset, the MI-related EEG signals were collected in nine subjects with three bipolar electrodes (C3, Cz and C4) and were bandpass filtered between 0.5 Hz and 100 Hz with a sampling frequency of 250 Hz. The EEG electrode Fz served as EEG ground. The data of each subject contained five sessions, two session without feedback and three session with feedback. In this paper, the three sessions with feedback were adopted, and the first session was selected as training data while the other two evaluation sessions were used as test data. Moreover, since the EEG electrode Cz is normally utilized to classify the MI task of foot or tongue (Morash, Bai, Furlani, Lin, & Hallett, 2008), only the electrodes C3 and C4 were used in our experiment.

The time scheme of experimental paradigm is showed in Fig. 3 (Tangermann et al., 2012). Each trial began with a fixation cross and a short beep warning. At second 3, a visual cue (an arrow of the left or right direction) was showed on the screen. Then, the subject began to perform MI task over a period of 4 s. A more than 2.5 s break followed after the MI period to avoid adaptation.

5. Data processing

5.1. Pre-processing

In pre-processing stage, raw EEG data was windowed and band-pass filtered. In order to achieve piecewise stationarity and capture more rich information for MI classification, we applied a 1 s sliding window with overlap on raw EEG data. Then, every temporal window was fed to bandpass filters.

ERD/ERS phenomena normally can be detected in alpha (8–14 Hz) and beta (14–27 Hz) band over the EEG electrodes C3 and C4. However, study (Leocani, Toro, Manganotti, Zhuang, & Hallett, 1997) states that ERD/ERS oscillations in gamma (around 40 Hz) band represent a level of event-related information processing during the preparation and execution of MI tasks. Thus, we band-pass filtered each temporal window at alpha, beta and gamma bands, which are 8–14 Hz, 14–27 Hz and 27–45 Hz. Moreover, on the grounds that IIR filter can obtain better response specifications

with much lower order than FIR filter, we adopted a fourth-order Butterworth bandpass filter to process each temporal window of raw EEG data and performed filtering in both forward and reverse directions to ensure the zero-phase distortion. Fig. 4 illustrates the whole processing steps of raw EEG data. Our experiments were carried out in MATLAB (R2017b) running on a PC with an Intel Xeon E5-2643 @3.40 GHz processor and 128GB RAM.

5.2. Feature extraction and classification

In the interest of extracting more significant underlying information for MI classification, the absolute values of the variations of each bandpass filtered temporal window were computed (see Eq. (11)) for each subject. Then, the variations of each temporal window were sent to bispectrum estimation (see Eq. (12)).

In training phase, an optimal bispectral segment length was learned with training data. First, the 1D horizontal slice of bispectrum on the principal domain was estimated, and VSBS features was extracted by selecting a segment length in a search space (see Eqs. (13) and (14)). The search space is specified from 2 to half of the length of the 1D bispectral horizontal slice. Second, the VSBS features of each temporal windows with each bispectral segment length were fed to a classifier, and CA was obtained by comparing the estimated labels with the true labels through a fivefold cross-validation. In order to improve the generalization capability of the classifier, the average normalized CA with each bispectral segment length was computed by averaging the normalized CA of each temporal window. By repeating above steps, the average normalized CA with each bispectral segment length for every subject was yielded. Finally, the average and sum of the average normalized CA of all subjects with each bispectral segment length were figured, and the optimal bispectral segment length was determined by comparing the average and sum values.

Support vector machine (SVM) solves the binary classification problem by maximizing the margin and finding an optimal hyper plane between two classes, and it is broadly employed owing to its high flexibility and robustness, and powerful theoretical foundation (Sun, Feng, Chen, & Lu, 2018). Therefore, we implemented SVM classifier for MI classification in this paper.

In test phase, the VSBS features of the training and test data was extracted with the optimal bispectral segment length. Then, the VSBS features of training data were adopted to find the optimal hyper plane of SVM classifier, and the final results of all the temporal windows could be achieved.

6. Results

6.1. Performance of segment selection

To find the optimal bispectral segment length, we employed the fivefold cross-validation over the training data of the dataset 1 and dataset 2 (see Section 4), and the average normalized CA of all the temporal windows for each subject was obtained with each bispectral segment length. The best 7 comparison results of the bispectral

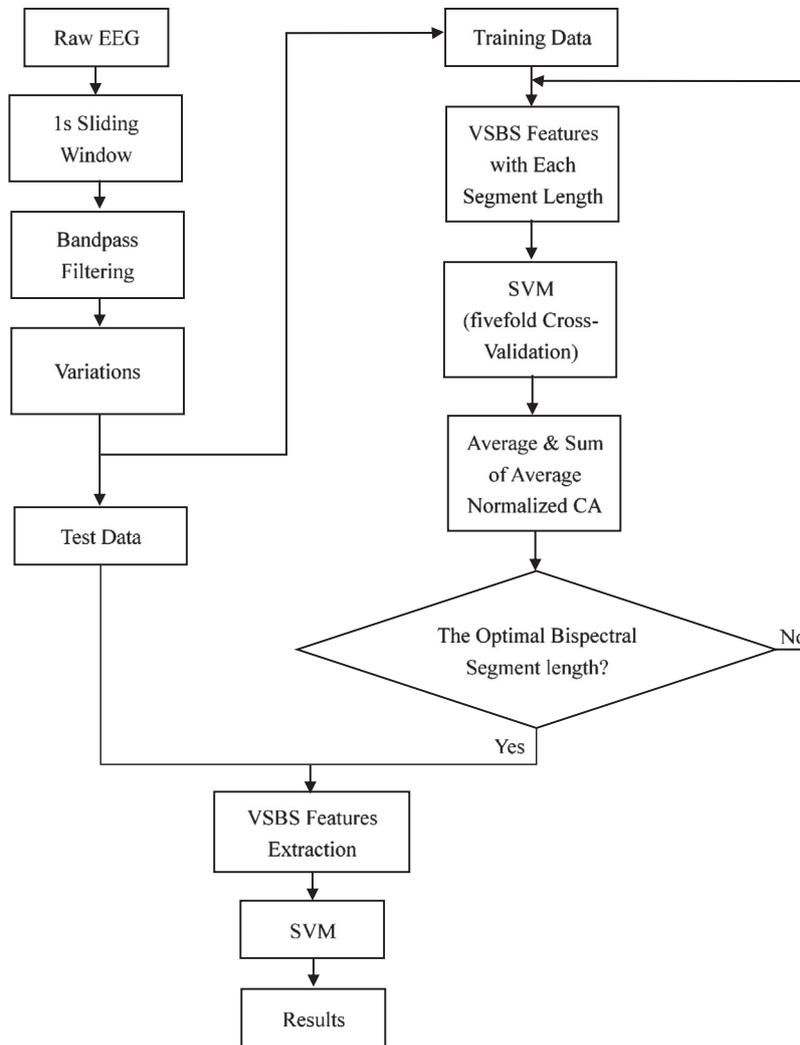


Fig. 4. Flowchart of the proposed VSBS method.

segment lengths for the dataset 1 and dataset 2 are illustrated in Figs. 5 and 6, respectively.

Figs. 5a and 6a show the average normalized CA of all the temporal windows for each subject in the dataset 1 and dataset 2, respectively. In Fig. 5a, the average normalized CA with the bispectral segment length 3 is relatively higher for most subjects, and the gaps of the average normalized CA between each subject with the bispectral segment length 3 is obviously smaller than with other bispectral segment lengths, especially the bispectral segment length 2, 7 and 8. Moreover, for the dataset 2, we can observe in Fig. 6a that the gaps of the average normalized CA between each subject with the bispectral segment length 3 and 4 are narrower than with other bispectral segment lengths. However, for most subjects, the average normalized CA with the bispectral segment length 3 is higher than with the bispectral segment length 4.

In Figs. 5b and 6b, the average and sum of the average normalized CA of all the subjects in the dataset 1 and dataset 2 are plotted, respectively. In Fig. 5b, the average and sum of the average normalized CA with the bispectral segment length 3 are 0.3 and 1.49, which are superior compared with other bispectral segment lengths for the dataset 1. Besides, in Fig. 6b, the average and sum values with the bispectral segment length 3 are, 0.49 and 4.4, greater than the rest bispectral segment lengths for the dataset 2.

Based to the surpassing performance of the bispectral segment length 3 over the training data of the dataset 1 and dataset 2, we selected 3 as the optimal bispectral segment length.

6.2. Performance of the proposed VSBS method on the dataset 1

In this paper, we utilized the CA and kappa value as the statistical measure for comparison (Saa & Çetin, 2012). The kappa value can be computed as below:

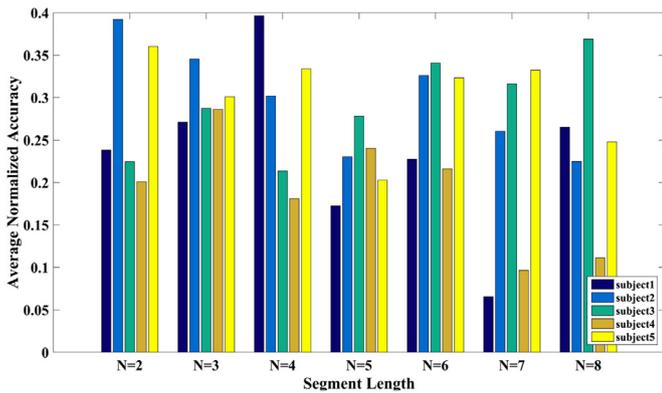
$$k = \frac{C \times P_{cc} - 1}{C - 1}, \quad (15)$$

where C is the number of classes and P_{cc} is the probability of correct classification. The larger kappa value is, the better performance is.

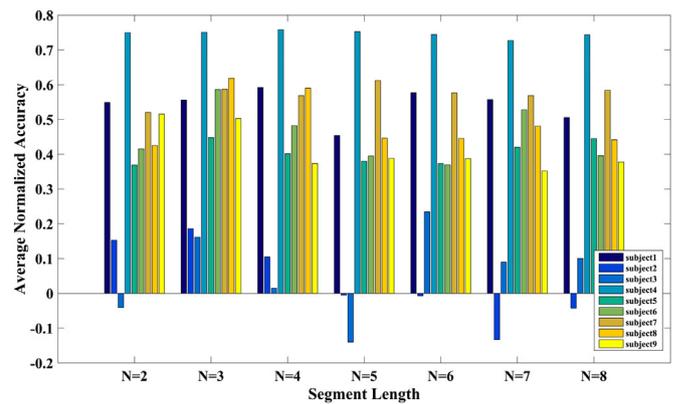
In order to verify the effectiveness of the proposed method, we also employed PSD, ABS, variations based PSD (VPSD) and variations based ABS (VABS) methods over the dataset 1 and dataset 2 for comparison, and all the comparison methods were pre-processed in the same steps as VSBS method. Moreover, the comparison methods adopted the same SVM classifier as in the proposed method. Table 1 and Fig. 7 indicate the comparison performance of PSD, VPSD, ABS, VABS methods and VSBS method for the dataset 1.

Table 1
Classification accuracy (%) and kappa values of PSD, VPSD, ABS, VABS methods and VSBS approach for the dataset 1. The bold shows the best average results.

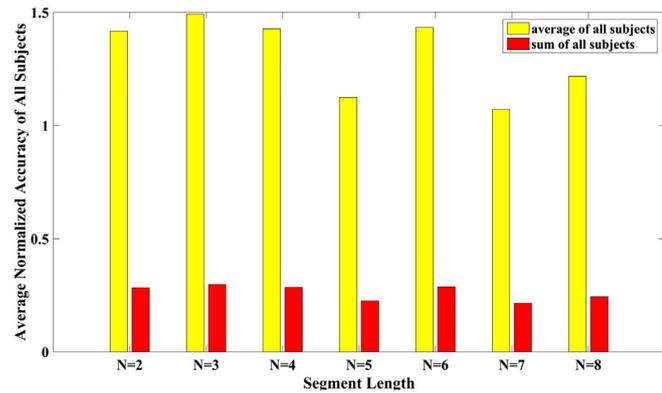
Subject	Accuracy (%)					Kappa				
	PSD	VPSD	ABS	VABS	VSBS	PSD	VPSD	ABS	VABS	VSBS
B1	67.50	70.00	65.00	75.00	82.50	0.35	0.4	0.3	0.5	0.65
B2	70.00	80.23	70.00	68.00	70.69	0.40	0.60	0.40	0.36	0.41
B3	72.50	67.50	68.00	65.00	75.00	0.45	0.35	0.36	0.30	0.50
B4	80.95	76.19	79.00	69.00	81.00	0.62	0.52	0.58	0.38	0.62
B5	71.43	71.43	74.00	69.00	81.00	0.43	0.43	0.48	0.38	0.62
Average	72.48	73.02	71.20	69.20	78.04	0.45	0.46	0.42	0.38	0.56



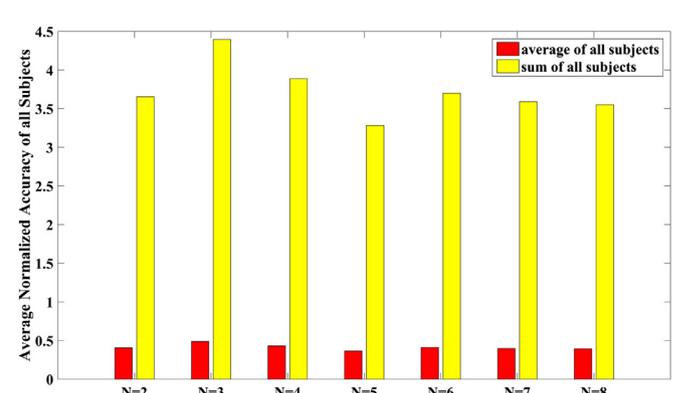
(a)



(a)



(b)



(b)

Fig. 5. Performance of each bispectral segment length over the training data of the dataset 1. (a) The average normalized CA for each subject with the bispectral segment length 2–8. (b) The average and sum of the average normalized CA of all the subjects with the bispectral segment length 2–8.

Fig. 6. Performance of each bispectral segment length over the training data of the dataset 2. (a) The average normalized CA for each subject with bispectral segment length 2–8. (b) The average and sum of the average normalized CA of all the subjects with the bispectral segment length 2–8.

Fig. 7 gives the comparison of the kappa values of PSD, VPSD, ABS, VABS and VSBS methods, and we can observe that the kappa performance of our VSBS method outperforms the other methods for most subjects. Specifically, the results of the proposed method are outstandingly better than the other methods for subject B1, B3 and B5. Table 1 gives the CA and kappa values of PSD, VPSD, ABS, VABS and VSBS methods. It can be seen that VSBS algorithm obtains surpassing results compared to the other features extraction methods. VSBS method achieves an average CA of 78.04% as well as an average kappa of 0.56. Compared to PSD, the proposed method increases 5.56% in average CA and 0.11 in average kappa. For the average enhancement, VSBS method achieves 7.84% in average CA and 0.14 in average kappa against ABS method. More importantly, our method yields greater results than PSD, ABS and VABS methods for every subject in the dataset 1.

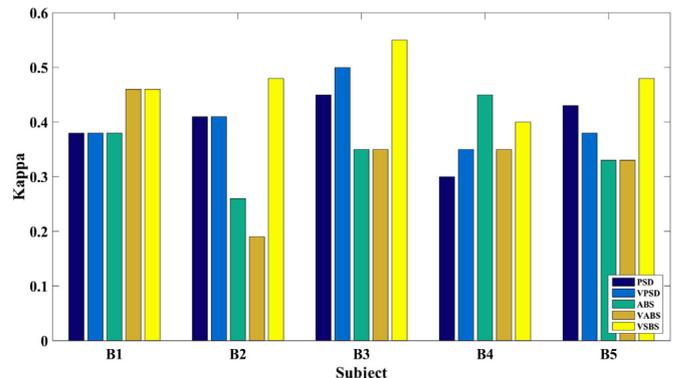


Fig. 7. Comparison of the Kappa values for each subject in the dataset 1.

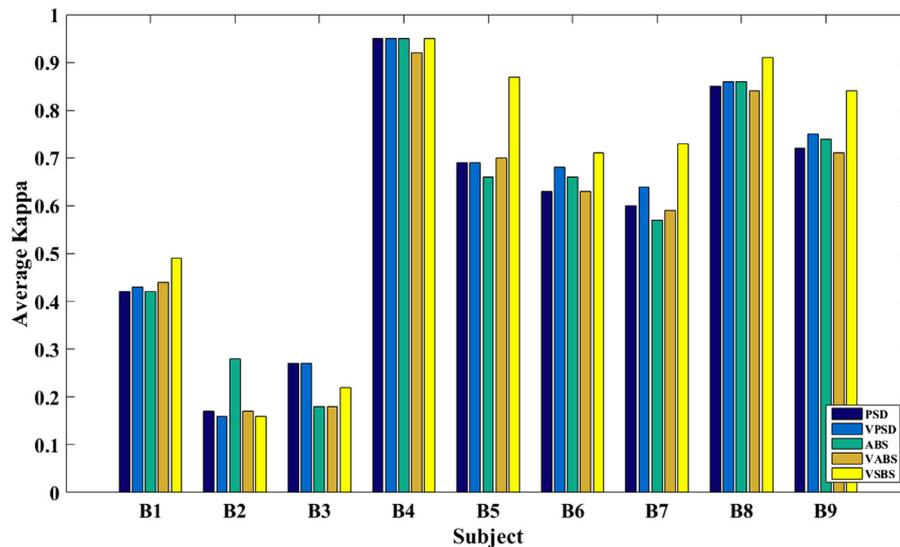


Fig. 8. Comparison of the average kappa values for each subject in the dataset 2.

Table 2

Classification accuracy (%) of the evaluation session (04E) from the dataset 2 with PSD, VPSD, ABS, VABS methods and VSBS approach. The bold indicates the best average result.

Subject	Evaluation (04E)				
	PSD	VPSD	ABS	VABS	VSBS
B1	80.36	81.25	81.25	76.79	83.93
B2	61.76	61.76	64.71	57.84	60.78
B3	65.81	66.67	57.26	61.54	59.83
B4	99.35	99.35	99.35	98.04	99.35
B5	81.41	82.69	80.77	80.77	91.67
B6	81.02	82.48	84.67	78.83	83.21
B7	74.77	76.64	71.96	77.57	84.11
B8	90.48	90.48	92.06	88.89	93.65
B9	93.04	94.78	91.30	93.04	95.65
Average	80.89	81.79	80.37	79.26	83.58

Table 3

Classification accuracy (%) of the evaluation session (05E) from the dataset 2 with PSD, VPSD, ABS, VABS methods and VSBS approach. The bold indicates the best average result.

Subject	Evaluation (05E)				
	PSD	VPSD	ABS	VABS	VSBS
B1	61.21	62.07	61.21	67.24	65.52
B2	55.24	54.55	63.64	59.44	55.24
B3	61.06	60.18	61.06	56.64	61.95
B4	96.10	96.10	96.10	94.16	95.45
B5	88.03	86.32	85.47	88.89	95.73
B6	82.46	85.96	81.58	84.21	87.72
B7	84.80	87.20	84.80	81.60	88.80
B8	94.23	95.19	94.23	95.19	97.12
B9	79.23	80.00	82.31	77.69	88.46
Average	78.04	78.62	78.93	78.34	81.78

Table 4

Maximum kappa of the two evaluation sessions from the dataset 2 with PSD, VPSD, ABS, VABS methods and VSBS approach. The bold indicates the best average rate.

Subject	Max. kappa				
	PSD	VPSD	ABS	VABS	VSBS
B1	0.61	0.63	0.63	0.54	0.68
B2	0.24	0.24	0.29	0.19	0.22
B3	0.32	0.33	0.22	0.23	0.24
B4	0.99	0.99	0.99	0.96	0.99
B5	0.76	0.73	0.71	0.78	0.91
B6	0.65	0.72	0.69	0.68	0.75
B7	0.70	0.74	0.70	0.63	0.78
B8	0.88	0.90	0.88	0.90	0.94
B9	0.86	0.90	0.83	0.86	0.91
Average	0.67	0.69	0.66	0.64	0.71

Table 5

Average kappa of the two evaluation sessions from the dataset 2 with PSD, VPSD, ABS, VABS methods and VSBS approach. The bold indicates the best average rate.

Subject	Average kappa				
	PSD	VPSD	ABS	VABS	VSBS
B1	0.42	0.43	0.42	0.44	0.49
B2	0.17	0.16	0.28	0.17	0.16
B3	0.27	0.27	0.18	0.18	0.22
B4	0.95	0.95	0.95	0.92	0.95
B5	0.69	0.69	0.66	0.70	0.87
B6	0.63	0.68	0.66	0.63	0.71
B7	0.60	0.64	0.57	0.59	0.73
B8	0.85	0.86	0.86	0.84	0.91
B9	0.72	0.75	0.74	0.71	0.84
Average	0.59	0.60	0.59	0.58	0.65

6.3. Performance of the proposed VSBS method on the dataset 2

For the dataset 2, Tables 2–5 and Fig. 8 show the comparison performance of PSD, VPSD, ABS, VABS methods and the proposed method. Tables 2 and 3 show the comparison CA of two evaluation sessions (04E and 05E), Tables 4 and 5 indicate the comparison maximum and average kappa values between the two evaluation sessions, and Fig. 8 illustrates the comparison average kappa val-

ues between the two evaluation sessions. In Tables 2 and 3, our VSBS approach obtains 83.58% and 81.78% in term of an average CA for the two evaluation sessions, respectively. In Tables 4 and 5, the proposed method reaches the highest average results of the nine subjects with 0.71 and 0.65 in term of maximum and average kappa. Compared to PSD method, the average improvement of maximum and average kappa with our VSBS approach is 0.04 and 0.06, respectively. With regard to the maximum and average kappa,

VSBS method obtains an average increase of 0.02 and 0.05 against VPSD method. Compared to ABS method, the average improvement of the maximum and the average kappa with VSBS approach are 0.05 and 0.06, respectively. Regarding the maximum and average kappa, VSBS method yields an average increase of 0.07 against VABS method. In Fig. 8, we can see that the results of VSBS method is prominent for most subjects against PSD and ABS based methods. Specifically, the performance of our method remarkably higher than the other methods for subject B1, B5, B7 and B9. Tables 2–5 and Fig. 8 further illustrate the validity of our VSBS method for MI classification.

For the sake of assessing the statistical significance of improvement with the proposed method, a two-way analysis of variance (ANOVA2) test was employed (Shahid & Prasad, 2011). For VSBS method and PSD method, the p -values is 0.0075. With regard to our method and VPSD method, the p -values is 0.00399. For VSBS and ABS method, the p -values is 0.0103. Regarding VSBS and VABS method, the p -values is 0.0002. Since all the p -values are less than 0.05, the improvement with VSBS method is statistically significant.

7. Discussion

By the reason of the sensitivity of one bispectral feature value to non-linear and non-Gaussian noises, the representative ABS method could deteriorate the performance of MI classification. Therefore, we propose an advanced bispectral features extraction method to reduce the influence of non-linear and non-Gaussian noises as well as improve the performance of MI classification. We verified our proposed VSBS method on the dataset collected in our laboratory and BCI competition IV dataset 2b, and employed the PSD based and ABS based methods for comparison. The comparison results demonstrate the validity and superiority of the proposed method for MI classification.

Observing the results of all the Tables and figures in Section 6, the PSD based methods outperform the ABS based methods. This may be because the PSD features still contain partially valuable features for MI classification compared to the contaminated ABS features. On the other hand, the average results of variations based PSD method are better than PSD method. This confirms that the variations of MI-related EEG signals could provide more useful information of MI-related brain activities.

7.1. Comparison with other state-of-the-art works

To verify our proposed method comprehensively, we compare our proposed approach to other state-of-the-art works, includ-

Table 6

The average kappa values of the proposed method and other state-of-the-art works on the dataset 2. The bold is used to illustrate the highest average rate.

Subject	Average kappa						
	SSMM	CNN-SAE	MKELM	PHVDSN	BSP	VPSD	VSBS
B1	0.48	0.52	0.55	0.43	0.54	0.43	0.49
B2	0.10	0.32	0.29	0.31	0.29	0.16	0.16
B3	0.11	0.49	0.09	0.26	0.22	0.27	0.22
B4	0.88	0.91	0.99	0.94	0.93	0.95	0.95
B5	0.74	0.66	0.69	0.73	0.64	0.69	0.87
B6	0.64	0.58	0.39	0.67	0.69	0.68	0.71
B7	0.53	0.49	0.74	0.67	0.50	0.64	0.73
B8	0.84	0.49	0.80	0.89	0.82	0.86	0.91
B9	0.71	0.46	0.67	0.75	0.74	0.75	0.84
Average	0.56	0.55	0.58	0.63	0.60	0.60	0.65

ing sparse support matrix machine (SSMM) (Zheng, Zhu, Qin, Chen, & Heng, 2018), convolutional neural networks with stacked autoencoders (CNN-SAE) (Tabar & Halici, 2017), multi-kernel extreme learning machine (MKELM) (Zhang et al., 2018), PSO optimized hidden-layer visible deep stacking network (PHVDSN) (Tang, Zhang, Zhou, & Liu, 2017) and ABS based method (BSP) (Shahid & Prasad, 2011). All these comparisons evaluated their methods on the same dataset 2 (see Section 4.2). Table 6 and Fig. 9 demonstrate the average kappa values of our method and the other counterparts.

Fig. 9 shows that the classification performance of VSBS method is more excellent than other comparison methods for most subjects. Specifically, the average kappa results of the proposed method are greatly prominent compared to the other methods for subject B5, B6, B8 and B9. Moreover, Table 6 also demonstrates the superiority of VSBS method for MI classification. Compared to SSMM and CNN-SAE method, the average improvement of our method remarkably increases 0.09 and 0.1, respectively.

BSP method extracts the ABS features and uses LDA classifier for MI classification. However, compare to BSP technique, our VSBS technique obtains an average increase of 0.05, especially 0.23 for subject B5 and B7, and 0.1 for subject B8 and B9. Therefore, the comparison results between BSP method and our method further prove the limitation of ABS features for MI classification. On the other hand, although BSP method yields a same average result with VPSD method, the performance of VPSD method for most subjects (B3, B4, B5, B7, B8, B9) is superior to BSP method. This observation also supports the conclusion that the ABS features could be impacted by non-linear and non-Gaussian noises so seriously that the separability of ABS features are greatly deteriorated.

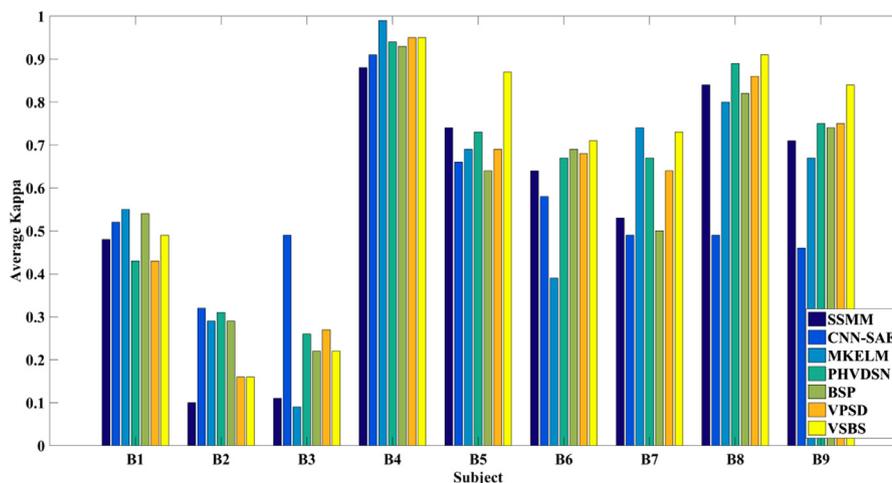


Fig. 9. Comparison results of VSBS method and other state-of-the-art works on the dataset 2.

Table 7

The two-way analysis of variance (ANOVA2) test results for evaluating the statistical significance of the difference between the proposed method and the other approaches.

	SSMM	CNN-SAE	MKELM	PHVDSN	BSP
p-value	0.0009	0.0605	0.0005	0.0003	0.0001

Table 8

Classification accuracy (%) of the proposed method and other state-of-the-art works on BCI competition III dataset IVa. The bold indicates the best average result.

Subject	Accuracy (%)			
	MKELM	MSPCA	SS-CFIS	VSBS
aa	83.3	96	82.1	94.7
al	98.5	92.3	100	99.1
av	71.4	88.9	63.3	82.3
aw	91.3	95.4	83.0	95.0
ay	93.1	91.4	60.3	94.3
Average	87.5	92.8	77.8	93.1

rated. The statistical significance of the difference between the proposed method and other state-of-the-art works is also evaluated with ANOVA2 test, and the results are showed in Table 7.

BCI competition III dataset IVa (Blankertz et al., 2006) was also utilized to evaluate our VSBS method. This dataset contained five healthy subjects, who need to accomplish two MI tasks regarding right hand and foot. The data of each subject was included 280 trials, and each trial was recorded from 118 electrodes with sampling rate 100 Hz. In each trial, different visual cues were appeared for 3.5 s, and more than 2 s relaxing interval was followed. For this dataset, all 280 trials were applied for training and test, and a ten-fold cross-validation is employed to select channels and features (Pfurtscheller, Brunner, Schlögl, & Lopes da Silva, 2006).

Table 8 indicates the comparison results of VSBS method and other state-of-the-art works, including spatio spectral cognitive fuzzy inference System (SS-CFIS) (Das, Suresh, & Sundararajan, 2016), multiscale principal component analysis based (MSPCA) (Kevric & Subasi, 2017) method and MKELM. We can see that VSBS method yields better results than other state-of-the-art works in Table 8. Feet imagery movements can be detected in midcentral foot representation area of scalp while hands imagery movements can be differently revealed in contralateral area of scalp. Moreover, hands and feet imagery movements can trigger different beta rebounds (Pfurtscheller, Neuper, Brunner, & Lopes, 2005). Therefore, the classification results of hand and foot MI tasks could be relatively higher than the results of right and left hands MI tasks.

7.2. Comparison of classifiers

To further highlight the effectiveness of the proposed features extraction method, VSBS method with five different classifiers, SVM, k-nearest neighbor (KNN), random forest (RF), back propagation (BP) network and naïve Bayesian (NB), were employed on the dataset 2. Table 8 and Fig. 10 indicate the average kappa values of the five different classifiers.

Experimental results of RF and BP classifiers are relatively close to SVM classifier as shown in Fig. 10. On the other hand, NB classifier yields the worse classification performance due to the strong independent assumptions. In Table 8, we can observe that SVM, RF and BP classifiers achieve an average kappa of 0.65 and 0.66, respectively. Yet, the classification effectiveness of KNN is unsatisfactory. This may be because KNN may be relatively harder to capture more useful features compared to the other classifiers (Zhu et al., 2019). AUC (Area Under Curve) (Du, Dua, Acharya, & Chua, 2012) and F1-Score (Gerla et al., 2017) are used to evaluate the perfor-

Table 9

The average kappa values of the proposed method with the different classifiers on the dataset 2. The bold is used to illustrate the highest average rate.

Subject	Average Kappa				
	SVM	NB	RF	KNN	BP
B1	0.49	0.45	0.37	0.37	0.55
B2	0.16	0.18	0.22	0.15	0.21
B3	0.22	0.10	0.19	0.12	0.21
B4	0.95	0.94	0.96	0.94	0.96
B5	0.87	0.71	0.86	0.82	0.82
B6	0.71	0.68	0.74	0.77	0.70
B7	0.73	0.61	0.73	0.68	0.75
B8	0.91	0.87	0.92	0.88	0.92
B9	0.84	0.77	0.83	0.85	0.82
Average	0.65	0.59	0.65	0.62	0.66

Table 10

The AUC values of the proposed method with the different classifiers on the dataset 2. The bold is used to illustrate the highest average rate.

Subject	AUC				
	SVM	NB	RF	KNN	BP
B1	77.16	75.63	72.74	68.77	78.70
B2	67.34	61.67	65.40	62.21	65.46
B3	63.89	61.55	67.24	62.67	66.22
B4	98.54	97.13	97.11	96.79	98.12
B5	91.08	87.31	92.91	90.97	91.00
B6	88.48	83.80	85.90	85.50	85.29
B7	87.57	82.02	89.50	87.46	86.71
B8	96.29	92.74	95.73	94.44	96.33
B9	96.44	91.36	91.66	92.92	92.36
Average	85.20	81.47	84.24	82.41	84.46

Table 11

The F1-Score of the proposed method with the different classifiers on the dataset 2. The bold is used to illustrate the highest average rate.

Subject	F1-score				
	SVM	NB	RF	KNN	BP
B1	76.82	74.94	72.37	71.30	79.90
B2	68.58	64.15	68.62	65.00	67.62
B3	63.61	62.38	67.00	62.83	66.69
B4	98.04	96.73	97.71	96.73	98.04
B5	91.34	86.53	92.63	91.66	91.02
B6	90.50	85.02	87.22	88.68	85.41
B7	86.45	81.78	87.35	86.91	86.91
B8	95.64	93.26	95.64	94.05	95.64
B9	93.47	89.12	90.87	92.61	91.29
Average	84.94	81.55	84.38	83.31	84.72

mance of classifiers, and Tables 10 and 11 give the comparison AUC values and F1-Score of the five different classifiers. It also can be seen that SVM, RF and BP classifiers yield higher results in Tables 9 and 10. Therefore, we can draw the conclusion that our VSBS method could produce robustly better performance for MI classification with most classifiers.

7.3. Limitation

Since VSBS method selects the optimal bispectral segment length by comparing the average and sum of the average normalized CA of all the subjects, this global optimal bispectral segment length may be ineffective for some subjects who are affected by EOG artifacts seriously (Saa & Çetin, 2012). Therefore, the further work will focus on finding the sub-optimal bispectral segment length for each subject and further improving the separability of the bispectral features for MI classification.

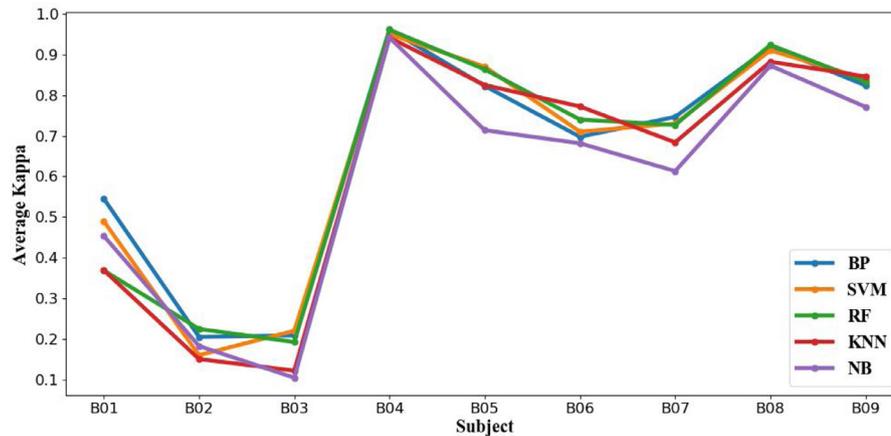


Fig. 10. Comparison results of VSBS method with the different classifiers on the dataset 2.

8. Conclusion

In this paper, an advanced features extraction method is proposed to improve the performance of MI classification. The proposed method utilizes the variations of MI-related EEG signals as input to bispectrum analysis, thus more valuable information of underlying MI-related brain activities could be revealed. Moreover, a new segmented bispectral features are developed to reduce the impact of non-linear and non-Gaussian noises, and an optimal bispectral segment length is learn from training data to highlight the useful features for MI classification. The dataset collected in our laboratory and BCI competition IV dataset 2b were used to verify the effectiveness of the proposed method. The results prove that our method improves the separability of bispectral features for MI classification. Moreover, the comparison results with other state-of-the-art works also indicates the superiority of the proposed method for MI classification.

In conclusion, several advantages of the proposed method could be summarized: (1) the proposed method verifies that the variations of MI-related EEG signals could provide more valuable information of MI-related brain dynamics; (2) our method enhances the discriminability of bispectral features and reduces the impact of non-linear and non-Gaussian noises; (3) the proposed method has validated practical for the sake of the superior performance for more than one dataset.

Conflicts of interest

The authors declare that no conflicts of interest exist.

Credit Author Statement

We have read and have abided by the statement of ethical standards for manuscripts submitted to the Journal of Expert Systems With Applications.

We declare that there are no known conflicts of interest associated with this publication, An Advanced Bispectrum Features for EEG-based Motor Imagery Classification, and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

Acknowledgments

This work is supported by National Natural Science Foundation of China grant 61673312, 91648208, Beijing Advanced Innovation Center for Intelligent Robots and Systems grant 2016IRS19, Research Fund for the Doctoral Program of Higher Education of China grant 20100201120040, China Postdoctoral Science Foundation grant 20110491662, 2012T50805.

References

- Allison, B. Z., Wolpaw, E. W., & Wolpaw, J. R. (2007). Brain-computer interface systems: progress and prospects. *Expert Review of Medical Devices*, 4(4), 463–474.
- Alvarez-Meza, A. M., Velasquez-Martinez, L. F., & Castellanos-Dominguez, G. (2015). Time-series discrimination using feature relevance analysis in motor imagery classification. *Neurocomputing*, 151(P1), 122–129.
- Ang, K., Chin, Z. Y., Zhang, H., & Guan, C. (2008). Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)* (pp. 2390–2397).
- Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., & Zhang, H. (2012). Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Frontiers in Neuroscience*, 6(39), 1–9.
- Blankertz, B., Müller, K., Krusienski, D., Schalk, G., Wolpaw, J. R., Schlögl, A., et al. (2006). The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2), 153–159.
- Chella, F., Marzetti, L., Pizzella, V., Zappasodi, F., & Nolte, G. (2014). Third order spectral analysis robust to mixing artifacts for mapping cross-frequency interactions in EEG/MEG. *NeuroImage*, 91, 146–161.
- Collis, W. B., White, P. R., & Hammond, J. K. (1998). Higher-order spectra: The bispectrum and trispectrum. *Mechanical Systems and Signal Processing*, 12(3), 375–394.
- Das, A. K., Suresh, S., & Sundararajan, N. (2016). A discriminative subject-specific spatio-spectral filter selection approach for EEG based motor-imagery task classification. *Expert Systems with Applications*, 64, 375–384.
- Du, X., Dua, S., Acharya, R. U., & Chua, C. K. (2012). Classification of epilepsy using high-order spectra features and principle component analysis. *Journal of Medical Systems*, 36(3), 1731–1743.
- Feng, F., Si, A., & Zhang, H. (2011). Research on fault diagnosis of diesel engine based on bispectrum analysis and genetic neural network. *Procedia Engineering*, 15, 2454–2458.
- Gandhi, V., Prasad, G., Coyle, D., Behera, L., & McGinnity, T. M. (2015). Evaluating Quantum Neural Network filtered motor imagery brain-computer interface using multiple classification techniques. *Neurocomputing*, 170, 161–167.
- Gerla, V., Kremen, V., Covassin, N., Lhotska, L., Saifutdinova, E. A., Bukartyk, J., et al. (2017). Automatic identification of artifacts and unwanted physiologic signals in EEG and EOG during wakefulness. *Biomedical Signal Processing and Control*, 31, 381–390.
- Helbig, M., Schwab, K., Leistriz, L., Eiselt, M., & Witte, H. (2006). Analysis of time-variant quadratic phase couplings in the trac?? alternant EEG by recursive estimation of 3rd-order time-frequency distributions. *Journal of Neuroscience Methods*, 157(1), 168–177.
- Herman, P., Prasad, G., McGinnity, T. M., & Coyle, D. (2008). Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(4), 317–326.

- Kevric, J., & Subasi, A. (2017). Biomedical signal processing and control comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system. *Biomedical Signal Processing and Control*, 31, 398–406.
- Klein, A., & Skrandies, W. (2017). Total variation for the analysis of event-related potentials. *Journal of Neuroscience Methods*, 275, 33–44.
- Leocani, L., Toro, C., Manganotti, P., Zhuang, P., & Hallett, M. (1997). Event-related coherence and event-related desynchronization/synchronization in the 10Hz and 20Hz EEG during self-paced movements. *Electroencephalography and Clinical Neurophysiology—Evoked Potentials*, 104(3), 199–206.
- Morash, V., Bai, O., Furlani, S., Lin, P., & Hallett, M. (2008). Classifying EEG signals preceding right hand, left hand, tongue, and right foot movements and motor imageries. *Clinical Neurophysiology*, 119(11), 2570–2578.
- Neuper, C., & Pfurtscheller, G. (2001). Event-related dynamics of cortical rhythms: Frequency-specific features and functional correlates. *International journal of psychophysiology*, 43(1), 41–58.
- Nikias, C. L., & Raghuveer, M. R. (1987). Bispectrum estimation: A digital signal processing framework. *Proceedings of the IEEE*, 75(7), 869–891.
- Pfurtscheller, G., Brunner, C., Schlögl, A., & Lopes da Silva, F. H. (2006). Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage*, 31(1), 153–159.
- Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology*, 110(11), 1842–1857.
- Pfurtscheller, G., Neuper, C., Brunner, C., & Lopes, F. (2005). Beta rebound after different types of motor imagery in man, 378, 156–159.
- Saa, J. F. D., & Çetin, M. (2012). A latent discriminative model-based approach for classification of imaginary motor tasks from EEG data. *Journal of Neural Engineering*, 9(2). doi:10.1088/1741-2560/9/2/026020.
- Sahayadhas, A., Sundaraj, K., Murugappan, M., & Palaniappan, R. (2015). Physiological signal based detection of driver hypovigilance using higher order spectra. *Expert Systems With Applications*, 42(22), 8669–8677.
- Schwab, K., Eiselt, M., Schelenz, C., & Witte, H. (2005). Time-variant parametric estimation of transient quadratic phase couplings during electroencephalographic burst activity. *Methods of Information in Medicine*, 44(3), 374–383.
- Shahid, S., & Prasad, G. (2011). Bispectrum-based feature extraction technique for devising a practical brain-computer interface. *Journal of Neural Engineering*, 8(2), 025014.
- Sun, L., Feng, Z., Chen, B., & Lu, N. (2018). A contralateral channel guided model for EEG based motor imagery classification. *Biomedical Signal Processing and Control*, 41, 1–9.
- Tabar, Y. R., & Halici, U. (2017). A novel deep learning approach for classification of EEG motor imagery signals. *Journal of Neural Engineering*, 14(1), 016003.
- Tang, X., Zhang, N., Zhou, J., & Liu, Q. (2017). Neurocomputing Hidden-layer visible deep stacking network optimized by PSO for motor imagery EEG recognition. *Neurocomputing*, 234(April 2016), 1–10.
- Tangermann, M., Müller, K., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C. et al. (2012). Review of the BCI competition IV, 6(July), 1–31.
- Wolpaw, J. R., Birbaumer, N., Mcfarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain – computer interfaces for communication and control, 113, 767–791.
- Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X., et al. (2018). Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces. *Expert Systems With Applications*, 96, 302–310.
- Zheng, Q., Zhu, F., Qin, J., Chen, B., & Heng, P. (2018). Sparse support matrix machine. *Pattern Recognition*, 76, 715–726.
- Zhou, S. M., Gan, J. Q., & Sepulveda, F. (2008). Classifying mental tasks based on features of higher-order statistics from EEG signals in brain-computer interface. *Information Sciences*, 178(6), 1629–1640.
- Zhu, X., Li, P., Li, C., Yao, D., Zhang, R., & Xu, P. (2019). Biomedical Signal Processing and Control Separated channel convolutional neural network to realize the training free motor imagery BCI systems. *Biomedical Signal Processing and Control*, 49, 396–403.