Computers in Biology and Medicine 91 (2017) 231–242

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

CSP-TSM: Optimizing the performance of Riemannian tangent space mapping using common spatial pattern for MI-BCI

Shiu Kumar [a,b,*], Kabir Mamun [b], Alok Sharma [b,c,d]

[a] Department of Electronics, Instrumentation and Control, School of Electrical & Electronics Engineering, College of Engineering, Science and Technology, Fiji National University, Suva, Fiji
[b] School of Engineering and Physics, Faculty of Science, Technology and Environment, The University of the South Pacific, Suva, Fiji
[c] Institute for Integrated and Intelligent Systems (IIIS), Griffith University, Brisbane, Australia
[d] RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan

ARTICLE INFO

Keywords:
Brain computer interface (BCI)
Common spatial pattern (CSP)
Electroencephalography (EEG)
Motor imagery (MI)
Riemannian distance
Tangent space mapping (TSM)

ABSTRACT

Background: Classification of electroencephalography (EEG) signals for motor imagery based brain computer interface (MI-BCI) is an exigent task and common spatial pattern (CSP) has been extensively explored for this purpose. In this work, we focused on developing a new framework for classification of EEG signals for MI-BCI.
Method: We propose a single band CSP framework for MI-BCI that utilizes the concept of tangent space mapping (TSM) in the manifold of covariance matrices. The proposed method is named CSP-TSM. Spatial filtering is performed on the bandpass filtered MI EEG signal. Riemannian tangent space is utilized for extracting features from the spatial filtered signal. The TSM features are then fused with the CSP variance based features and feature selection is performed using Lasso. Linear discriminant analysis (LDA) is then applied to the selected features and finally classification is done using support vector machine (SVM) classifier.
Results: The proposed framework gives improved performance for MI EEG signal classification in comparison with several competing methods. Experiments conducted shows that the proposed framework reduces the overall classification error rate for MI-BCI by 3.16%, 5.10% and 1.70% (for BCI Competition III dataset IVa, BCI Competition IV Dataset I and BCI Competition IV Dataset IIb, respectively) compared to the conventional CSP method under the same experimental settings.
Conclusion: The proposed CSP-TSM method produces promising results when compared with several competing methods in this paper. In addition, the computational complexity is less compared to that of TSM method. Our proposed CSP-TSM framework can be potentially used for developing improved MI-BCI systems.

1. Introduction

In a brain computer interface (BCI) system, direct communication takes place between the brain and the external devices without the involvement of peripheral nerves and muscles in order to allow humans to interact with their surroundings [1]. The electrical control signals generated by the brain activity are measured using the electroencephalogram (EEG) method. Motivated by the aim of re-establishing independence and reducing social exclusion for people with disabilities, BCI research has gained a vast range of interest in this field. BCI has numerous emerging applications such as applications for communication control [2,3], environment control [4], movement control [5,6] and neuro-rehabilitation [7–9]. Researchers have also concentrated on developing BCI systems such as P300-based word typing system [10], EEG controlled mouse [11], brain wave controlled robots [12] and brain wave controlled wheelchair for the disabled [13]. EEG signal has also been used for biometric identification [14]. The major focus and emphasis of BCI research is in the field of biomedical engineering [15–19].

Motor imagery (MI) activity detection is based on the sensory-motor rhythm's modulation that is the event-related synchronization and desynchronization (ERS and ERD). Due to major lateralization of pre-movement and post-movement ERD and ERS activities, distinct brain patterns are produced for brief imagery movements that are used for enhancing the MI pattern recognition [20]. Actual movement and MI of same task activate the motor cortex almost to the same extent as the

* Corresponding author. Department of Electronics, Instrumentation and Control, School of Electrical & Electronics Engineering, College of Engineering, Science and Technology, Fiji National University, Suva, Fiji.
E-mail addresses: shiu.kumar@fnu.ac.fj (S. Kumar), kabir.mamun@usp.ac.fj (K. Mamun), alok.sharma@griffith.edu.au (A. Sharma).

https://doi.org/10.1016/j.compbiomed.2017.10.025
Received 19 June 2017; Received in revised form 8 October 2017; Accepted 23 October 2017

autonomic nervous system is engaged by a MI as if an actual movement was under way. In Ref. [21], a MI-BCI system uses EEG signal for on-line measure of MI activity which provides neurofeedback to enable the users to better focus on the task. The study was designed for providing feedback to stroke patients during MI practice. MI has attracted substantial attention as a possible neuro-rehabilitation technique for improving motor recovery following stroke [7,9,22–26] and is also used for entertainment [27–29].

Generally, in the realization of a MI-BCI system the following stages are involved: signal acquisition, preprocessing, feature extraction, signal classification, and control interface. Several methods are used for measuring and studying the brain waves such as EEG, magnetoencephalogram (MEG), positron emission tomography (PET) and single photon emission computed tomography (SPECT). However, due to the fact that EEG is non-invasive, has relatively low cost, supports real time analysis and is portable makes it the most suitable method for BCI systems [30,31]. With technological advancements, devices having low cost and reduced complexity such as Neurosky Mindwave [32] and Emotiv EPOC/EPOC + headset [33] have been developed, which are used in BCI research and applications.

Several techniques have been proposed with respect to different aspects of the problem, the three major aspects being filter band selection, spatial filter estimation and feature extraction for optimal BCI performance. The conventional common spatial pattern (CSP) method utilized a single filter band. However, various methods such as common spatio-spectral pattern (CSSP) [34], common sparse spectral spatial pattern (CSSSP) [35], sub-band CSP (SBCSP) [36], filter band CSP (FBCSP) [37], discriminative filter band CSP (DFBCSP) [38], sparse filter band CSP (SFBCSP) [39] and filter band selection using binary particle swarm optimization (BPSO) [40] have been proposed. These methods have been proposed due to the fact that the filter band that gives maximum discrimination between different tasks for optimal classification varies from subject to subject i.e. dependent on the users. CSP is applied to the signals combined with its time delayed signals to obtain the finite impulse response filter (FIR) coefficients in CSSP (limited to single delay) and CSSSP. For each channel, a different spectral pattern is found by CSSP whereas CSSSP finds a spectral pattern that is common to all channels. In SBCSP and FBCSP methods, the signal is decomposed into sub-bands using multiple filter banks having different frequency passband. CSP is then applied to each of the sub-bands separately. Recursive band elimination using support vector machine (SVM) is performed for SBCSP method in order to obtain the top ten sub-bands from the 24 sub-bands (each having bandwidth of 4 Hz) with better discriminating power. The selected sub-bands are then used for feature extraction and classification (using SVM). On the other hand, the FBCSP method utilizes all the sub-bands (having a bandwidth of 4 Hz in the frequency range of 4–40 Hz). However, feature selection using various techniques has been employed to obtain the features that give optimal discrimination resulting in better classification. DFBCSP selects the best four sub-bands from 12 filter banks (each with bandwidth of 4 Hz) using the fisher ratio calculated using single channel C3, which provides a measure of the discrimination between the two MI tasks. CSP features are then extracted separately from these sub-bands and SVM classifier is used for classification.

Recently, a SFBCSP [39] method that also uses multiple filter bands is proposed, which optimizes the sparse patterns. CSP on multiple overlapping frequency bands is performed, and significant CSP features are then selected using supervised technique. SVM classifier is then used for MI classification using the selected features. Over the years, Bayesian learning has also gained increased attention and it has been used for feature selection in various applications [41]. In Ref. [42], a sparse Bayesian learning of filter banks (SBLFB) approach has been proposed for obtaining sparse features that are used for classification using the SVM classifier. The EEG signals were decomposed into multiple sub-bands, CSP features were extracted and significant features were selected using Bayesian learning. Furthermore, in Ref. [40], a frequency band selection using BPSO is presented. Unlike the DFBCSP method, the BPSO selects either one or two best sub-bands producing optimal results. However, the BPSO algorithm is computationally expensive and takes a huge amount of time due to which the authors have only used selected (25 and 14) channels of data for processing. These methods recommend that feature extraction using sub-band decomposition can improve the classification accuracy of MI-BCI.

The measured EEG signals contain measurement noise and spontaneous components related to other brain activities that degrade the signal-to-noise ratio (SNR). Therefore, spatial filtering is usually performed on the EEG signal in order to discriminate between the overlapping activities and improve the SNR. In spatial filtering, EEG signals from multiple channels are linearly combined so that the unwanted sources are obscured and those sources that are of interest are enhanced. To this end, CSP has been widely explored [43] and used in BCI research. Several methods have been proposed to optimize the temporal filter parameters for better classification in order to prevent time consuming manual tuning. In Ref. [44], an iterative spatio-spectral patterns learning (ISSPL) method is proposed. In this method, the classifier and spectral filters are parameterized simultaneously to enhance the classification performance. A weighted averaging method [45] has also been proposed for estimating the covariance matrix for CSP. The number of training trials available for MI tasks is usually small and can result in the deterioration of the classification accuracy. To tackle this problem, a regularized CSP (R-CSP) with aggregation has been proposed by Lu et al. [46]. In this method, two parameters (β, γ) regularize the estimation of the covariance matrix with the aim of reducing the estimation variance and bias. In the regularization process, EEG trials from other subjects (other than the subject of interest) are utilized. A divergence based framework for CSP that also utilizes information from other subjects is presented in Ref. [47].

Barachant et al. [48] proposed multiclass BCI classification by Riemannian geometry, showing that it improves the MI classification accuracy. The authors proposed to extract the spatial information directly from the bandpass filtered EEG signal. However, as the number of channels increases, it requires higher computational complexity since it requires eigenvalue decomposition of the input signal. Thus, the improved performance was achieved at the expense of an increased computation time. Since the computation time is high, it is not practical for real time application. However, it should be noted that it outperformed CSP in terms of classification performance. On the other hand, other methods have mostly focused on either optimizing the spatial filter or improving the classification performance by optimizing the filter bands. For improving classification performance by optimizing filter bands, use of multiple sub-bands with either band selection (using different approaches) or feature selection has been employed. Use of multiple sub-bands also increases the computation time. Therefore, to address these issues, we propose to optimize the performance of Riemannian tangent space mapping (TSM) by combining it with CSP. We propose to extract Riemannian tangent space features from the spatial filtered EEG signal. Using TSM to extract features from spatial filtered signal addresses the issue of the high computation time required by TSM, while still retaining its improved classification performance. To keep the computation time low, we proposed to use a single filter band. Furthermore, CSP on its own has shown to have good performance for MI signal classification. Therefore, to take further advantage of CSP, Riemannian tangent space features are combined with the CSP variance based features and feature selection is performed to obtain significant features. Linear discriminant analysis (LDA) [49] reduces the dimensionality of the selected features and SVM classifier with linear kernel is then used for classification. The proposed method is validated on three publically available EEG datasets (BCI Competition III dataset IVa, BCI Competition IV dataset I and BCI Competition IV dataset IIb), and compared with other related literature such as CSP, FBCSP, DFBCSP, SFBCSP and SBLFB. Experimental results show that the proposed method enhances the MI classification accuracy.

The rest of the paper is organized as follows. Section 2 presents the CSP method followed by the proposed CSP-TSM framework for MI classification. The results and discussion are presented in Sections 3 and 4, respectively, while Section 5 draws the conclusion and gives insight of some future works.

2. CSP-TSM for MI classification

Because existing competing methods such as FBCSP, DFBCSP, SFBCSP and SBLFB uses multiple frequency bands, the computational complexity of the system is higher in comparison with that of if a single frequency band is utilized. The proposed CSP-TSM method utilizes a single CSP based filter such as the CSSP, CSSSP, and SPEC-CSP [50] methods. This paper shows that a single frequency band relating to the brain activities associated with MI-BCI tasks can be sufficient to obtain optimal results.

In this section, the CSP-TSM algorithm for MI classification is presented. The single bandpass filtered EEG signal is transformed using the spatial filter, obtaining a lower dimensional signal. Features are then extracted from the spatial filtered signal using tangent space mapping of the Riemannian distance. The features obtained after performing TSM are then fused with the CSP variance based features. Regression analysis using least absolute shrinkage and selection operator (Lasso) is used for feature selection in order to obtain the significant features. LDA is applied to the selected features to reduce the dimensionality of the features. SVM with linear kernel is then implemented for EEG signal classification of MI tasks.

2.1. Feature extraction using CSP

The CSP technique that was first used for detection of abnormalities using EEG signal [51] has gained a wide attention in the field of MI-BCI for feature extraction. CSP is commonly used for transforming the signal to a new time series that has maximum discrimination between different MI tasks. Consider the two class problem having EEG trials $x_{n,1}$ and $x_{n,2} \in R^{C \times T}$, where n denotes the n-th trial, C is the number of channels and T is the number of sample points. The trials are bandpass filtered and it is assumed that the trials have been centered. The normalized spatial covariance matrix Σ of class l is given by (1), where N_l is the number of trials in class l.

$$\sum_l = \frac{1}{N_l} \sum_{n=1}^{N_l} \frac{X_{n,l} X_{n,l}^T}{trace\left(X_{n,l} X_{n,l}^T\right)} \tag{1}$$

In CSP, the function in (2) is maximized in order to get maximum separability between the variance of the two classes, where $w \in R^C$ is the spatial filter and $||\cdot||_2$ denotes the l_2 norm.

$$\max_w J(w) = \frac{w^T \sum_1 w}{w^T \sum_2 w} \qquad s.t. \quad \|w\|_2 = 1 \tag{2}$$

Solving the generalized eigenvalue problem in (3) will result in the maximization of the Rayleigh quotient $J(w)$.

$$\sum\nolimits_1 w = \lambda \sum\nolimits_2 w \tag{3}$$

The spatial filter matrix W is then obtained using (4), where P is the whitening transformation matrix and U is matrix containing the eigenvectors. In this paper, it is assumed that the eigenvalues and their corresponding eigenvectors are arranged in descending order.

$$W = U^T P \tag{4}$$

The CSP spatial filter W_{csp} is then formed by selecting the first and last m columns of W. The transformation Z of a given EEG trial X is thus obtained using (5).

$$Z = W_{CSP}^T X \tag{5}$$

The variance based CSP feature vector is then formed as $F_{csp} = [F_1, F_2, \ldots, F_{2m}]$, where F_i is given by (6) with var (Z_m) denoting the variance of m-th row of Z.

$$F_i = \log\left(\frac{\text{var}(Z_i)}{\sum_{n=1}^{2m} \text{var}(Z_n)}\right) \tag{6}$$

2.2. Single band CSP with TSM (CSP-TSM)

The CSP-TSM framework utilizes the theory of Riemannian geometry in the manifold of covariance matrices [48] and uses the tangent space vectors as features. Fig. 1 illustrates the proposed CSP-TSM framework that utilizes a single frequency band. The projected data $\widehat{X} \in R^{2m \times T \times N}$ obtained after spatial filtering using (5) is utilized for further processing, where N is the total number of trials. The normalized covariance matrix $\Sigma_i \in R^{2m \times 2m}$ for each projected trial $\widehat{X}_i$ is computed using (7), where Σ_i represents the trial covariance matrix (TCM) of the spatial filtered data for the i-th trial.

$$\sum_i = \frac{\widehat{X}_i \widehat{X}_i^T}{trace\left(\widehat{X}_i \widehat{X}_i^T\right)} \tag{7}$$

Consider $\Gamma(t) : [0,1] \rightarrow \Sigma(n) \in R^{2m \times 2m}$ to be any differentiable path between $\Gamma(0) = \Sigma_1$ and $\Gamma(1) = \Sigma_2$, where $\Sigma(n)$ is symmetric positive-definite (SPD) matrices (the space of which is a differentiable Riemannian manifold (M)). Equation (8) gives the length of $\Gamma(t)$. The geodesic is the curve with minimum length that connects the two points on the manifold. The length of this curve gives the Riemannian distance between the two points.

$$L(\Gamma(t)) = \int_0^1 \left|\left|\Gamma(t)\right|\right|_{\Gamma(t)} dt \tag{8}$$

Classification algorithms such as SVM, LDA and Neural Networks cannot be implemented directly in the Riemannian manifold as they are based on projections into hyper-planes. Thus each of the TCM is mapped into the tangent space located at the Riemannian mean Σ of the whole set of TCM matrices. The Riemannian mean is given by (9), where R $(\cdot)$ operator denotes computation of Riemannian mean and $i = 1, \ldots, N$. Due to the fact that there is no closed-form expression for computing the Riemannian mean, an efficient iterative algorithm given in Ref. [52] has been adopted.

$$\Sigma = \mathscr{R}(\Sigma_i) = \arg\min_{\Sigma \in \Sigma(n)} \sum_{i=1}^{N} \delta_R^2(\Sigma, \Sigma_i) \tag{9}$$

The Riemannian distance δ_R can be obtained using (10), where upper $(\cdot)$ means taking the upper triangular part of the symmetric matrix, vectorizing it and applying $\sqrt{2}$weight to the out-of-diagonal elements [53], the logarithmic mapping $\text{Log}_\Sigma(\Sigma_i)$ is given by (11) and s_i is a vector of the normalized tangent space.

$$\delta_R(\Sigma, \Sigma_i) = \left|\left|\text{Log}_\Sigma(\Sigma_i)\right|\right|_\Sigma = \left|\left|\text{upper}\left(\Sigma^{-1/2} \text{Log}_\Sigma(\Sigma_i) \Sigma^{-1/2}\right)\right|\right|_2 = ||s_i||_2 \tag{10}$$

$$\text{Log}_\Sigma(\Sigma_i) = \Sigma^{1/2} \log\left(\Sigma^{-1/2} \Sigma_i \Sigma^{-1/2}\right) \Sigma^{1/2} \tag{11}$$

Therefore, each of the TCM Σ_i is mapped into the tangent space resulting in a set of $2m(2m+1)/2$ dimensional vectors s_i. These vectors s_i are used as the tangent space features in our proposed framework.

2.3. Feature selection and classification

The variance based CSP features F_{csp} are concatenated with the tangent space features s_i to form a $2m^2 + 3m$ dimensional feature vector

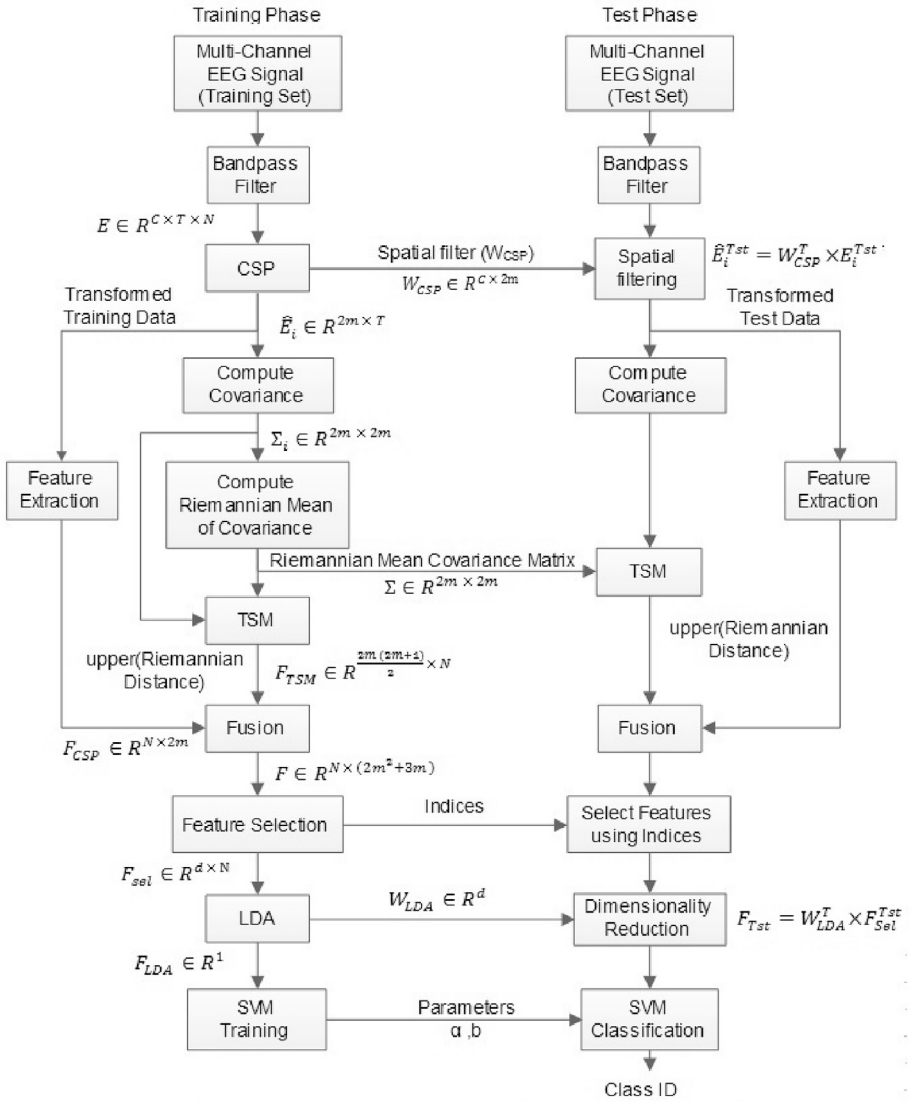

Fig. 1. Illustration of the proposed single band CSP-TSM framework.

F. However, the success of any MI-BCI depends largely on the extracted features. A good selection of features will result in an improvement in the classification performance of MI tasks for BCI applications. Thus we employ feature selection using sparse regression model known as Lasso estimate [54] given by (12) in order to obtain significant features, where u is the sparse vector that is to be learned, λ is the nonnegative regularization parameter that controls the sparsity of u, and $y \in R^N$ is the vector consisting of the target class labels.

$$u = \underset{u}{\operatorname{argmin}} \frac{1}{2} \left|\left| Fu - y \right|\right|_2^2 + \lambda \left|\left| u \right|\right|_1 \tag{12}$$

The expression in (12) is an optimization problem and for solving this the coordinate descent algorithm [55] has been adopted. The optimized sparse vector u obtained is used for selecting the features. Top d features in *F*corresponding to the largest d values in u are selected to form an optimized feature set F_{sel} containing the significant features. LDA is applied to further reduce the dimensionality of the optimized feature set using (13), where w_{LDA} is the LDA transformation matrix obtained using (14), in which μ_l is the mean of class l, and S_B and S_W are the between-class and within-class scatter matrices, respectively.

$$F_{LDA} = w_{LDA}^T F_{sel} \tag{13}$$

$$w_{FDA} = \arg\max_w \left(\frac{w^T S_B w}{w^T S_W w} \right) = S_W^{-1}(\mu_1 - \mu_2) \tag{14}$$

LDA maximizes the between class scatter while minimizing the within class scatter i.e. it finds an orientation such that the separation between classes is maximized in the reduced dimensional space following Fisher's criterion function. SVM is a discriminative classifier that is defined by the separating hyperplane. It finds the hyperplane that maximizes the margin of the support vectors i.e. the training data closest to the hyperplane. The features F_{LDA} obtained after applying LDA are used for training the SVM classifier using a linear kernel. The trained SVM classifier is then used to classify the test trials.

3. Results

3.1. Description of data

The BCI Competition III dataset IVa [56], BCI Competition IV dataset I [57] and BCI Competition IV dataset IIb, available publicly have been used for evaluating the proposed method. The BCI Competition III dataset IVa contains the EEG signals recorded from five subjects labeled *aa*, *al*, *av*, *aw*, and *ay* performing two different (right hand and left foot) MI tasks. 118 channels at positions of the extended international 10/20

system [58] were used for measuring the EEG signal, sampled at a rate of 1000 Hz. The dataset contains a total of 280 trials for each subject with equal number of trials for each task. The data provided was bandpass filtered with passband of 0.05–200 Hz. The down-sampled data at 100 Hz is used. A detail description of the dataset can be found at http://www.bbci.de/competition/iii/.

The BCI Competition IV dataset I contains 59 channels of real long-term EEG signals recorded for left hand, and right hand MI tasks acquired from 7 healthy subjects (named *a* to *g*). The down sampled signal at 100 Hz has been used, which contains 200 trials for each subject with equal number of each type of MI tasks. A detail description of the dataset can be found online at http://www.bbci.de/competition/iv/.

BCI Competition IV dataset IIb comprises of 3 channels (C3, Cz, and C4) data for right hand and left hand motor imagery tasks. The data was recorded from nine subjects at a sampling rate of 250 Hz. As in Ref. [39], only the third session data is used for evaluation. A total of 160 trials of motor imagery EEG measurements are available each of the subjects (there are equal number of trials for each motor imagery task). A detail description of the dataset can be found online at http://www.bbci.de/competition/iv/.

3.2. Experimental setup

In this study, data between 0.5 s and 2.5 s (i.e. 200 sample points for BCI Competition III dataset IVa and BCI Competition IV dataset I and 500 sample points for BCI Competition IV dataset IIb) after the visual cue has been extracted for processing [59]. The data was further bandpass filtered using a butterworth bandpass filter having passband of 7–30 Hz. The parameter m used was also chosen out of {1, …,10} in advance. The value of $m = 3$ as been adopted as it gave the maximum classification accuracy for all subjects.

The parameter λ in (12) was chosen out of {0.01, …,0.99} using a 10-fold cross validation. Fig. 2 presents an example of the effect of varying λ on the mean squared error for subject *aa*. The value of λ selected can be any value between the points circled in blue and green. Similar result was obtained for all other subjects and thus the value of $\lambda = 0.15$ have been used.

3.3. Analysis of the CSP-TSM framework with TSM and CSP methods

Sparse regression has been exploited in this study and a properly determined regularization parameter λ leads to an optimal learned sparse vector u from (12). The sparse vector u is utilized for selecting the significant features from the variance based CSP features and TSM features obtained from spatially filtered data.

An example of the sparse vectors learned by the proposed CSP-TSM framework for each subject in the BCI Competition III dataset IVa and BCI Competition IV dataset I are presented in Fig. 3 and Fig. 4, respectively. The feature indices from 1 to 6 corresponds to the variance based CSP features while the indices from 7 to 27 corresponds to the 21 TSM features obtained from the spatially filtered data. The significant features appeared at both the types of features used in this study i.e. the CSP variance based features and TSM features. Similar results were obtained for BCI Competition IV dataset IIb. It can be noted from Fig. 3 that the

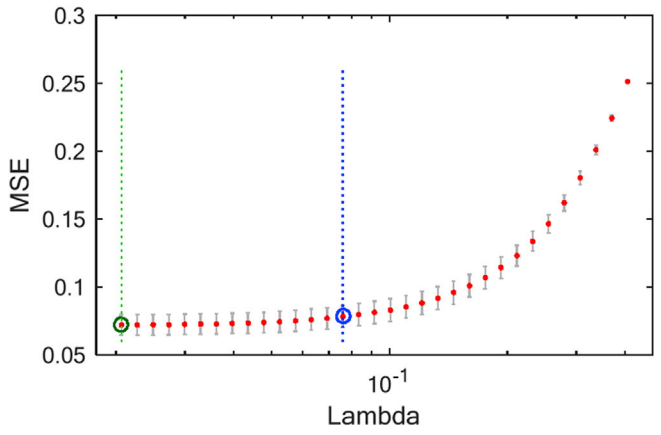

Fig. 2. Cross-validated MSE of Lasso fit for one of the experimental runs for subject *aa*.

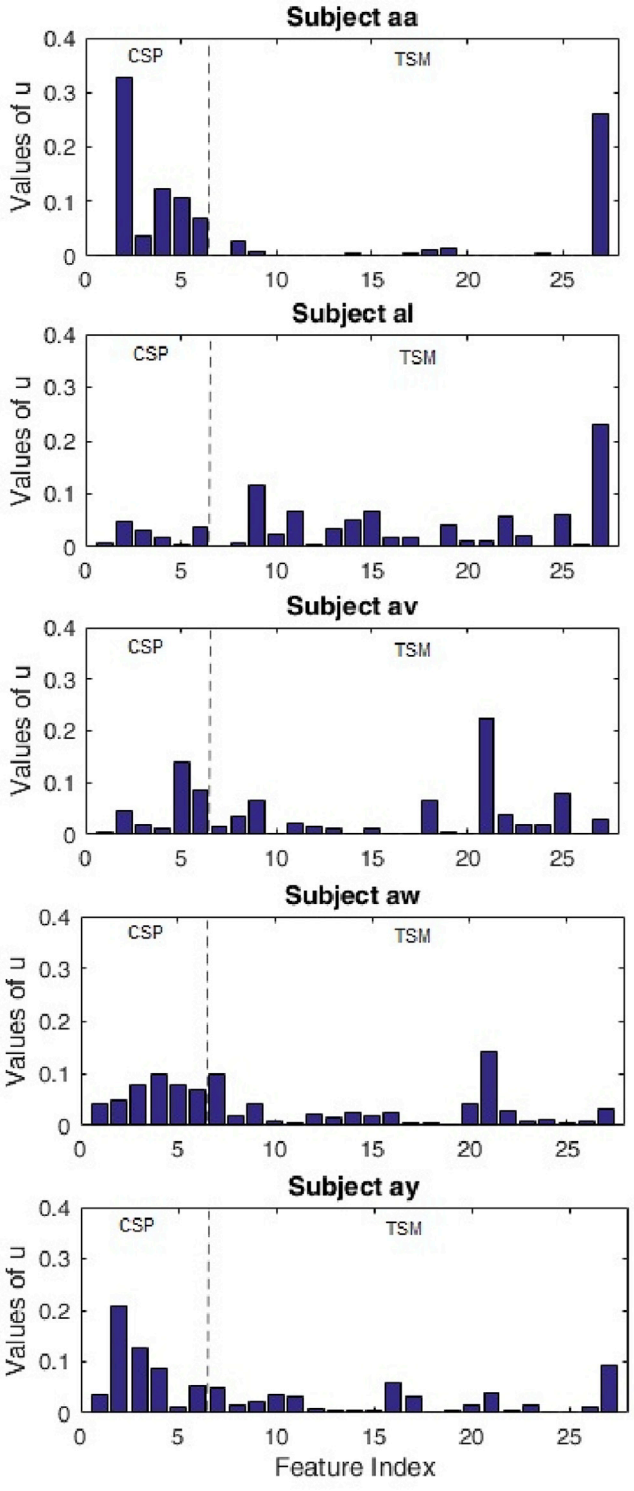

Fig. 3. Example of sparse vectors learned by CSP-TSM method for each of the five subjects in BCI Competition III dataset IVa. On the left of the dotted line are the CSP sparse vectors and on the right are the TSM sparse vectors.

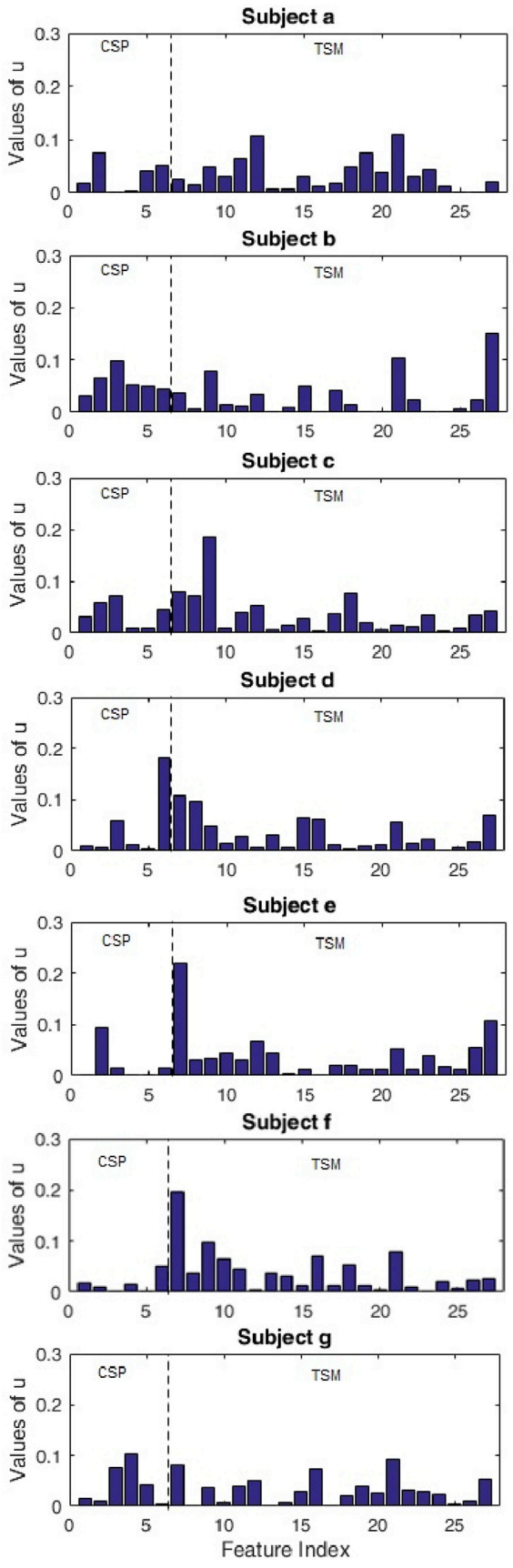

Fig. 4. Example of sparse vectors learned by CSP-TSM method for each of the seven subjects in BCI Competition IV dataset I. On the left of the dotted line are the CSP sparse vectors and on the right are the TSM sparse vectors.

most significant feature for subjects *aa* and *ay* were from the variance based CSP features.

On the other hand, for subjects *al, av* and *aw* the most significant feature was from the TSM features. In Fig. 4, the most significant feature for subjects *d* and *g* was from variance based CSP features while for subjects *a*, *b*, *c*, *e* and *f* the most significant feature was from TSM features. The significant features selected were different for each of the subjects, which is in agreement with the related literature [38,40,42,60] (can be due to different skull size, skin thickness and due to the fact that the way subjects think about the same task differs from subject to subject). This indicates the reason for using both types of features and explains the need for feature selection in our proposed framework for enhanced MI classification accuracy.

To compare the performance of the CSP-TSM framework with the TSM and conventional CSP algorithm, we have evaluated the classification performance of these methods using selected 25 channels of data (for BCI Competition III dataset IVa). This is because the TSM algorithm could not be run on the desktop personal computer (Matlab® on windows 7 with 6 GB Ram operating at 3.3 GHz) using all 118 channels of EEG signal as Matlab becomes irresponsive, the computer freezes and runs out of memory. The channels selected was adopted from Ref. [40] and includes channels that are important for the neurophysiological discrimination between two mental tasks. The three algorithms (CSP, TSM, and CSP-TSM) were run simultaneously so that all the classifiers were built using the same set of training data and the results evaluated exactly on the same set of test data for all the methods. The results obtained are depicted in Fig. 5. For all subjects except for subject *aw*, the CSP-TSM method performs better while for subject *aw*, the best performance was obtained using TSM method. It should be noted that on average the CSP-TSM method performs better than both the conventional CSP and TSM methods by 1.73% and 1.17%, respectively.

3.4. Analysis of the CSP-TSM framework with other competing methods

In this section, the CSP-TSM framework is evaluated against other competing methods. Table 1, Table 2 and Table 3 shows the 10 × 10-fold cross validation classification error rates of the proposed CSP-TSM framework along with the conventional CSP, FBCSP, DFBCSP, SFBCSP and SBLFB methods for the two datasets. All the methods are evaluated under the same experimental conditions using all channels of EEG data. For the conventional CSP, a wide filter band of 6–40 Hz was adopted while for FBCSP, DFBCSP and SFBCSP 17 filter bands (sub-bands) with a bandwidth of 4 Hz overlapping each other at a rate of 2 Hz was adopted from Ref. [38]. The regularization parameter λ for SFBCSP was determined using the same procedure as mentioned earlier for CSP-TSM framework. All methods employ SVM as the classifier. In comparison with other competing methods; conventional CSP, FBCSP, DFBCSP, SFBCSP and SBLFB, the proposed CSP-TSM framework offers average error rate reductions of 3.16%, 1.76%, 0.63%, 4.10% and 3.42%, respectively for BCI Competition III dataset IVa, 5.10%, 4.10%, 3.80%,

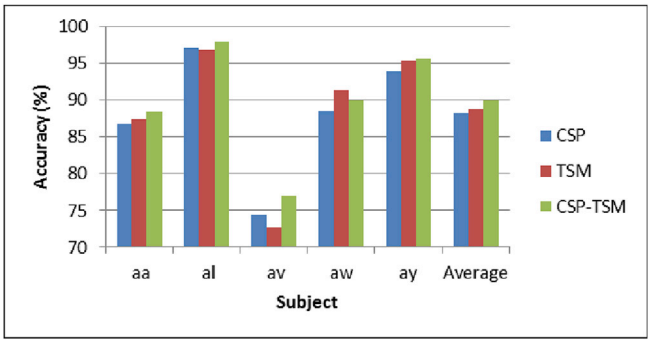

Fig. 5. 10 × 10-fold cross validation accuracies for conventional CSP, TSM and CSP-TSM methods using selected (25) channels for BCI Competition III dataset IVa.

Table 1
10 × 10-fold cross validation classification error rates (%) of CSP, FBCSP, DFBCSP, SFBCSP, SBLFB and CSP-TSM methods, respectively, using BCI Competition III Dataset IVa. For each subject the lowest error is marked in boldface.

Method	Subject					Average
	aa	*al*	*av*	*aw*	*ay*	
CSP	21.00 ± 5.31	3.86 ± 3.93	28.29 ± 7.46	10.36 ± 5.10	3.86 ± 4.11	13.47 ± 5.18
FBCSP	17.14 ± 8.19	1.29 ± 1.88	30.36 ± 8.23	6.50 ± 4.55	5.07 ± 4.68	12.07 ± 5.51
DFBCSP	**9.64 ± 5.01**	**1.00 ± 1.91**	31.21 ± 8.92	4.64 ± 4.75	8.21 ± 5.06	10.94 ± 5.13
SFBCSP	18.43 ± 8.26	1.64 ± 1.36	29.93 ± 6.44	9.29 ± 8.58	12.79 ± 5.96	14.41 ± 5.57
SBLFB	16.79 ± 8.93	1.36 ± 2.03	28.07 ± 8.45	5.57 ± 4.90	11.00 ± 6.03	12.56 ± 6.07
CSP-TSM (Proposed)	16.79 ± 6.29	2.14 ± 3.53	**24.90 ± 9.10**	**4.53 ± 2.80**	**3.21 ± 2.53**	**10.31 ± 4.85**

Table 2
10 × 10-fold cross validation classification error rates (%) of CSP, FBCSP, DFBCSP, SFBCSP, SBLFB and CSP-TSM methods, respectively, using BCI Competition IV Dataset I. For each subject the lowest error is marked in boldface.

Method	Subject							Average
	a	*b*	*c*	*d*	*e*	*f*	*g*	
CSP	13.2 ± 8.1	42.8 ± 12.3	43.7 ± 11.3	22.4 ± 8.8	18.0 ± 9.7	22.5 ± 10.8	7.10 ± 5.1	24.2 ± 9.4
FBCSP	19.1 ± 9.4	44.7 ± 11.3	35.7 ± 9.6	22.2 ± 9.0	14.0 ± 9.1	19.6 ± 8.6	6.9 ± 6.6	23.2 ± 9.1
DFBCSP	16.8 ± 7.8	42.9 ± 9.7	35.2 ± 8.5	23.5 ± 8.4	18.3 ± 8.8	14.3 ± 8.6	9.0 ± 5.1	22.9 ± 8.1
SFBCSP	17.4 ± 5.9	45.3 ± 6.6	43.0 ± 11.6	29.5 ± 10.1	24.7 ± 10.3	20.9 ± 6.5	9.7 ± 5.0	27.2 ± 8.0
SBLFB	19.1 ± 9.7	41.5 ± 11.1	33.2 ± 12.5	**11.5 ± 7.9**	11.6 ± 6.9	21.2 ± 9.0	**5.9 ± 5.4**	20.6 ± 9.4
CSP-TSM (Proposed)	**11.9 ± 6.6**	**40.9 ± 10.3**	**32.1 ± 8.7**	15.7 ± 7.9	**9.8 ± 5.8**	**14.1 ± 7.1**	7.8 ± 5.7	**19.1 ± 7.5**

8.10% and 1.50%, respectively for BCI Competition IV dataset I, and 1.70%, 3.55%, 1.07%, 1.37% and 0.92%, respectively for BCI Competition IV dataset IIb. Our method outperforms all competing methods on average while also obtaining the lowest classification error rate for 3 out of 5 subjects (*av*, *aw* and *ay*) for BCI Competition III dataset IVa, 5 out of 7 subjects (*a*, *b*, *c*, *e* and *f*) for BCI Competition IV dataset I, and 3 out of 9 subjects (*B0103T*, *B0203T* and *B0503T*) for BCI Competition IV dataset IIb. Our proposed method performed best for subject *al* of BCI Competition III dataset IVa, subject *g* of BCI Competition IV dataset I and subject *B0403T* of BCI Competition IV dataset IIb, having error rate of 2.14%, 7.8% and 0.75%, respectively. Moreover, as expected, the worst performance was noted for subject *av*, subject *b* and subject *B0303T* with error rate of 24.90%, 40.90% and 49.06%, respectively (consistent with results of other methods as it contains low quality signals).

Cohen's kappa coefficient (κ), which is a statistical method of measuring the consistency of the agreement between two variables or classes, is used to further validate the reliability of the results obtained. The coefficient κ takes into account the possibility of the agreement occurring by chance, thus making it a more robust measure of performance. A larger value of κ indicates a greater strength of agreement while a lower value indicates that the strength of agreement is weak [61]. The κ values obtained for BCI Competition III dataset IVa, BCI Competition IV dataset I and BCI Competition IV dataset IIb are given in Table 4, Table 5 and Table 6, respectively. Our proposed method achieved the highest average κ values on all the three datasets obtaining average κ values of 0.795, 0.630 and 0.560 for BCI Competition III dataset IVa, BCI Competition IV dataset I and BCI Competition IV dataset IIb, respectively. This shows that the results obtained using the proposed CSP-TSM method is comparatively reliable.

Furthermore, paired *t*-test with 5% significance level has been conducted by comparing the 2nd best results obtained from the existing technique to show the significance of the proposed method. The p-value obtained was 0.028, which showed that the improvement achieved was significant.

4. Discussion

All the body activities including motor and muscle movement are controlled by the brain, which is an essential part of the human body. The message is firstly constructed in the brain for every communication that is initiated. The human brain contains over 100 billion neurons [62]. These neurons communicate with each other producing different pattern of electrical signals (generated due to electromagnetic activities inside the brain) for different thoughts [63]. The function of a BCI system is to capture the EEG signal and decode them for different brain activities thus enabling a direct channel of communication between the brain and the external devices without the need for any muscular movement [64]. Neural network have been widely used in various applications [65,66] and has also been used for EEG signal classification [67–69]. However, methods that use CSP [35–38,60,70–72] have shown to be effective in MI EEG signal classification. Although EEG signals of larger fragments are highly non-Gaussian, up to 90% of short fragments of 4 s can be considered as Gaussian [73,74]. In this research, we have used 2 s segments and therefore the EEG signals are assumed to be Gaussian. This is shown by the fact that the CSP method is able to perform well in MI EEG signal classification. Thus, we have combined the power of CSP and TSM to improve the performance of MI EEG signal classification. We have employed Lasso estimate for feature selection, however, other methods of feature selection [75] could also be employed.

Research is also being carried out to improve the CSP algorithm. In Ref. [76], a neural network-based optimal spatial filter design method has been proposed. The spatial filter and the classifier are trained simultaneously using a neural network named spatial filter network (SFN). Methods have also been proposed to optimize the classifier performance [68]. The performance of CSP largely depends on the frequency bands that are selected and as such research still continues in finding the best frequency bands. In Ref. [42], a sparse Bayesian learning approach for filter band selection (SBLFB) is proposed to find significant features. Binary particle swarm optimization has been proposed for selecting the filter bands that contain the most significant features [40], while a subject based feature extraction method has been proposed that uses wavelet packet decomposition (WPD) together with CSP [77]. Therefore, implementing the proposed method with multiple selected filter banks may further improve the classification performance of the system at the cost of higher computational time.

Table 3
10 × 10-fold cross validation classification error rates (%) of CSP, FBCSP, DFBCSP, SFBCSP, SBLFB and CSP-TSM methods, respectively, using BCI Competition IV Dataset IIb. For each subject the lowest error is marked in boldface.

Method	Subject									Average
	B0103T	*B0203T*	*B0303T*	*B0403T*	*B0503T*	*B0603T*	*B0703T*	*B0803T*	*B0903T*	
CSP	23.19 ± 10.14	41.94 ± 11.04	46.69 ± 9.38	0.75 ± 2.04	17.85 ± 8.70	35.19 ± 11.05	14.50 ± 8.56	13.06 ± 8.43	**19.13** ± 9.96	23.59 ± 8.81
FBCSP	25.31 ± 9.99	42.94 ± 11.74	48.44 ± 10.82	**0.63** ± 1.88	42.25 ± 16.34	23.81 ± 10.94	13.81 ± 8.11	14.50 ± 8.56	17.25 ± 8.66	25.44 ± 9.67
DFBCSP	23.25 ± 11.23	40.75 ± 12.45	50.50 ± 12.87	0.75 ± 2.05	25.00 ± 10.71	20.88 ± 10.38	12.13 ± 9.05	11.23 ± 6.95	22.25 ± 10.80	22.96 ± 9.61
SFBCSP	26.50 ± 9.24	42.75 ± 12.84	**44.97** ± 11.65	0.38 ± 1.50	25.02 ± 7.38	**20.06** ± 10.70	12.25 ± 7.47	12.38 ± 7.63	25.00 ± 9.62	23.26 ± 8.67
SBLFB	25.25 ± 10.33	40.75 ± 11.99	50.68 ± 13.34	0.88 ± 2.53	20.21 ± 10.26	25.12 ± 12.32	**11.88** ± 9.39	**11.13** ± 8.95	19.38 ± 10.58	22.81 ± 9.97
CSP-TSM (Proposed)	**22.50** ± 9.44	**39.50** ± 10.43	49.06 ± 11.09	0.75 ± 2.23	**17.18** ± 10.25	23.01 ± 9.35	13.81 ± 7.89	11.44 ± 7.95	19.75 ± 9.47	**21.89** ± 8.67

Table 4
Cohen's kappa coefficient values obtained for the different methods using BCI Competition III Dataset IVa. For each subject the highest κ value is marked in boldface.

Method	Subject					Average
	aa	*al*	*av*	*aw*	*ay*	
CSP	0.613	0.927	0.426	0.800	0.903	0.734
FBCSP	0.601	0.970	0.384	0.837	0.881	0.735
DFBCSP	**0.816**	**0.976**	0.329	0.890	0.847	0.771
SFBCSP	0.394	0.917	0.389	0.743	0.763	0.641
SBLFB	0.664	0.973	0.439	0.889	0.780	0.749
CSP-TSM (Proposed)	0.636	0.950	**0.543**	**0.896**	**0.950**	**0.795**

4.1. Comparison of feature distribution

To further compare the performance of the CSP-TSM framework with the TSM and conventional CSP algorithm, we have analyzed the distribution of the extracted features for each of the methods. The distributions of the two most significant features obtained by CSP, TSM and CSP-TSM methods respectively from subject *al* is depicted in Fig. 6. If we draw an imaginary non-linear line separating the two clusters, the minimum number of misclassified trials obtained for CSP, TSM, CSP-TSM and SBLFB methods from Fig. 6 would be 15, 8, 7 and 14, respectively. Thus it can be seen that CSP-TSM method achieved the maximum discriminability of features. This explains the fact that the CSP-TSM method outperforms the CSP and TSM methods. The dataset used is from BCI Competition III and thus it may or may not contain outliers. However, the overlap in the feature points of CSP and TSM are in agreement with the related literature. It is shown in Ref. [48] that using Riemannian distance to Riemannian mean as used by TSM is more effective in providing relevant information about the class membership than using Euclidean distance to its mean as used by CSP. This explains the fact that the CSP-TSM approach is able to attain a better separation as it uses the properties of both CSP and TSM. The feature space of CSP can be observed as a filtered and approximated depiction of the Riemannian manifold [78]. Similarly, tangent space is an approximated depiction of the Riemannian manifold [48]. However, in CSP the spatial filters are learned from the training data using supervised technique while TSM is an unsupervised technique. Therefore, CSP can be subject to over fitting whereas TSM is not subject to over-fitting. Thus, TSM features may have shown to be more significant than CSP features in cases where over fitting occurs in CSP, which makes our proposed CSP-TSM method robust. This may also explain the reason why some subjects show TSM features are better than CSP features and other subjects show otherwise. It can be noted from Fig. 6 that some of the feature points from one class appear to completely resemble that of another class. This can be due to poor quality of the signal recorded or possibility of an error made during recording, for example the subject was actually performing class 1 MI task during recording of class 2 MI task. Supervised methods of training adopted in this research cannot overcome this problem. Therefore, use of unsupervised methods such as cluster methods [79–82] can be utilized to overcome this problem and may further improve the performance of the system.

4.2. Feature selection

The proposed CSP-TSM algorithm utilizes feature selection of the fused CSP variance based features and TSM (from spatially filtered data - CSP) features. We have employed Lasso for selecting the significant features as it has performed well in Ref. [39]. However, it requires the selection of the number of features to be selected. We have conducted experiments to select features in the range of 1–27 and the classification accuracies obtained for BCI Competition III dataset IVa and BCI Competition IV dataset I are shown in Fig. 7 (similar results were obtained for BCI Competition IV dataset IIb). From Fig. 7, it can be noted that using 10 selected features produced optimal results and therefore, the top 10 significant features have been used. Increasing the number of

Table 5
Cohen's kappa coefficient values obtained for the different methods using BCI Competition IV Dataset I. For each subject the highest κ value is marked in boldface.

Method	Subject							Average
	a	*b*	*c*	*d*	*e*	*f*	*g*	
CSP	0.736	0.144	0.126	0.552	0.640	0.550	0.858	0.515
FBCSP	0.618	0.106	0.286	0.556	0.720	0.608	0.862	0.537
DFBCSP	0.664	0.142	0.290	0.530	0.634	0.714	0.820	0.542
SFBCSP	0.652	0.094	0.140	0.410	0.506	0.582	0.806	0.456
SBLFB	0.618	0.170	0.336	**0.770**	0.768	0.576	**0.882**	0.589
CSP-TSM (Proposed)	**0.757**	**0.206**	**0.359**	0.696	**0.826**	**0.717**	0.851	**0.630**

Table 6
Cohen's kappa coefficient values obtained for the different methods using BCI Competition IV Dataset IIb. For each subject the highest κ value is marked in boldface.

Method	Subject									Average
	B0103T	*B0203T*	*B0303T*	*B0403T*	*B0503T*	*B0603T*	*B0703T*	*B0803T*	*B0903T*	
CSP	0.536	0.161	**0.106**	0.983	0.650	0.296	0.710	0.739	**0.618**	0.534
FBCSP	**0.620**	0.088	0.018	0.965	0.430	0.513	0.690	0.623	0.583	0.503
DFBCSP	0.535	0.185	0.010	0.984	0.500	0.583	0.758	**0.778**	0.555	0.543
SFBCSP	0.470	0.145	0.100	0.973	0.499	**0.598**	0.755	0.753	0.500	0.533
SBLFB	0.495	0.185	0.014	0.983	0.595	0.457	**0.763**	0.776	0.613	0.542
CSP-TSM (Proposed)	0.550	**0.210**	0.010	**0.985**	**0.655**	0.530	0.724	0.771	0.605	**0.560**

selected features to more than 10 either did not have an impact on the error rate or increased the error rate. This shows that there are features that are redundant and explains the need for feature selection. LDA have been employed after feature selection as in our previous work [83], we have shown that using LDA improves the classification performance of the system. Other approaches such as using only the TSM features obtained from the spatially filtered data and applying principal component analysis (PCA) or LDA have also been evaluated. However, the proposed framework shown in Fig. 1 produced optimal results and thus has been adopted. It is left as future study to evaluate other feature selection methods such as mutual information based feature selection [37] and sparse Bayesian learning for feature selection [42], which have also performed well in selecting the most significant features.

4.3. Comparison of computational complexity

The authors in Ref. [48] proposed to use TSM technique, which directly depends on the covariance matrices of the EEG signal. The TSM method is an unsupervised process and has the ability to extract spatial information that is comparable to the state-of-the-art CSP. Because the TSM method directly depends on the covariance matrices, the computational complexity of the system increases drastically when signal from large number of EEG channels is used. In this case, a channel selection algorithm will be required in order to select the channels with most significant information. In this work, we have proposed to use CSP to reduce the dimensionality of the EEG signal and then apply TSM. In this way, the computational complexity of the system is reduced drastically.

To evaluate the computational complexity of our proposed method with that of TSM, the Big O Notation has been used. Big O Notation has been widely used for evaluating the computational complexity of an algorithm. The worst case scenario is specifically defined by Big O. It can be used to evaluate the execution time required by an algorithm. Table 7 shows the computational complexity of CSP, TSM, and the proposed CSP-TSM framework. It can be noted that the computational complexity of the TSM method will increase as a cubic function with respect to the number of channels of EEG signal used. On the other hand, for the CSP and proposed CSP-TSM method, the computational complexity would only increase linearly. Thus, it is evident that the proposed CSP-TSM method is computationally less expensive compared to the TSM method. We did not take into account the training phase for evaluating the computational complexity of the methods as it can be performed offline.

To further evaluate and compare the computational complexity of the TSM and CSP-TSM algorithms, the training and test time taken by the algorithms were computed. Fig. 8 and Fig. 9 shows the training and test time required by the TSM and proposed CSP-TSM algorithms, respectively. The test time given is the time required to process a single test trial. The number of channels used varies from 3 to 63. The training and test time using 3 to 30 channels for TSM and CSP-TSM were less than 2.0 s and 1.0 ms, respectively. We could not evaluate the TSM method using more than 63 channels as the computer freezes and runs out of memory. The training time for the CSP-TSM algorithm was in the range of 1.80 s (using 3 channels) to 2.21 s (using 118 channels) while the test time varied between 230 μs (using 3 channels) to 290 μs (using 118 channels). However, it can be noted from Figs. 8 and 9 that the training time and test time for the TSM algorithm starts to increase rapidly when the number of channels used increases to more than 55 channels while no significant increase in the training and test time for CSP-TSM is noted. Moreover, the test time for the TSM algorithm using 63 channels was 217 ms. Increasing the number of channels used to more than 63 will cause a rapid increase in training and test time (as can be seen from the trends in Figs. 8 and 9) and require higher computational resources. This shows that the implementation of the TSM algorithm is not practical for real time applications with large (greater than 63) number of channels data. On the other hand, the proposed CSP-TSM method can easily be implemented in real time as the training time and test time are low even when higher channels of data is used. This shows that the performance of our proposed system is better in terms of misclassification rate and computational complexity when compared with TSM. Thus, the problem of high computational complexity of TSM algorithm when large number of channel data is used has been addressed.

Moreover, the state-of-the-art methods such as FBCSP, DFBCSP, SFBCSP and SBLFB utilize multiple frequency bands, whereas our method uses only a single frequency band for achieving optimal results. The use of a single frequency band compared to using multiple frequency bands makes our system computationally less expensive.

5. Conclusion

This paper presents a new framework for MI classification by utilizing two types of features; the variance based CSP features and the TSM features of the spatially filtered data. The proposed method selects the most significant features from these two types of features. LDA is applied to the selected features, which is then classified using the trained SVM classifier. The proposed method has enhanced the classification accuracy

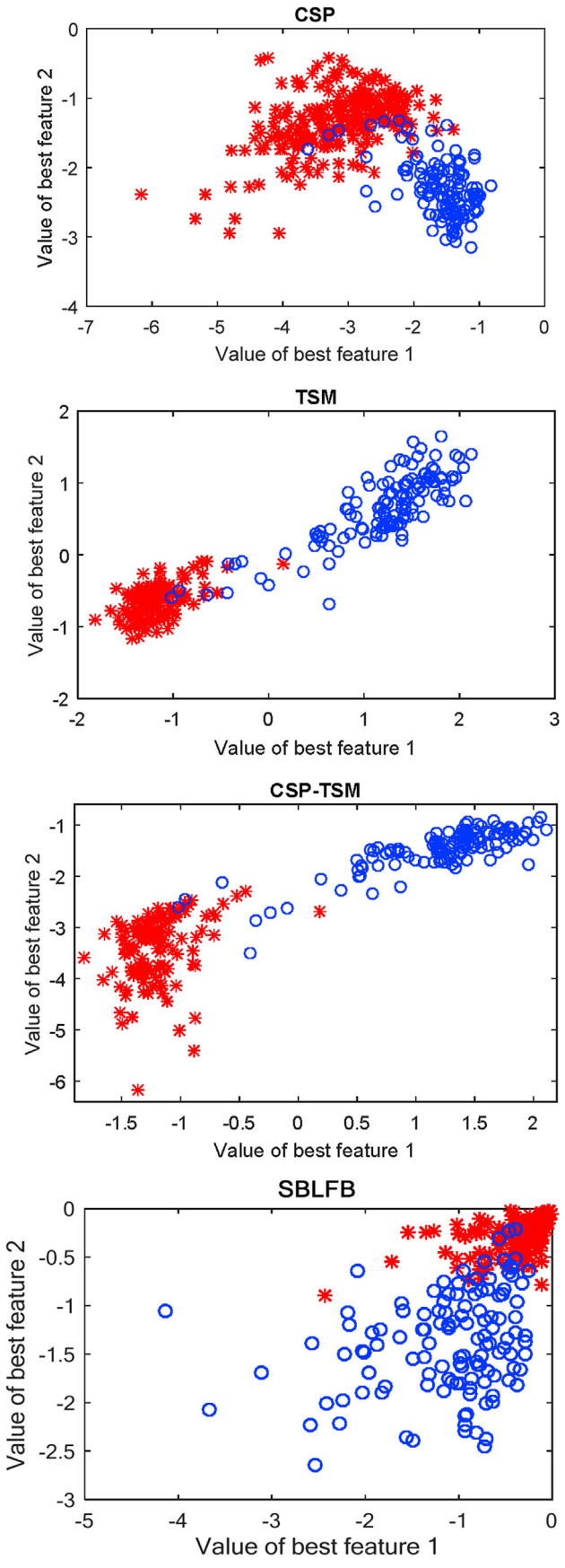

Fig. 6. Distribution of two most significant features obtained by CSP, TSM, CSP-TSM and SBLFB, respectively, from subject *al* of BCI Competition III dataset IVa.

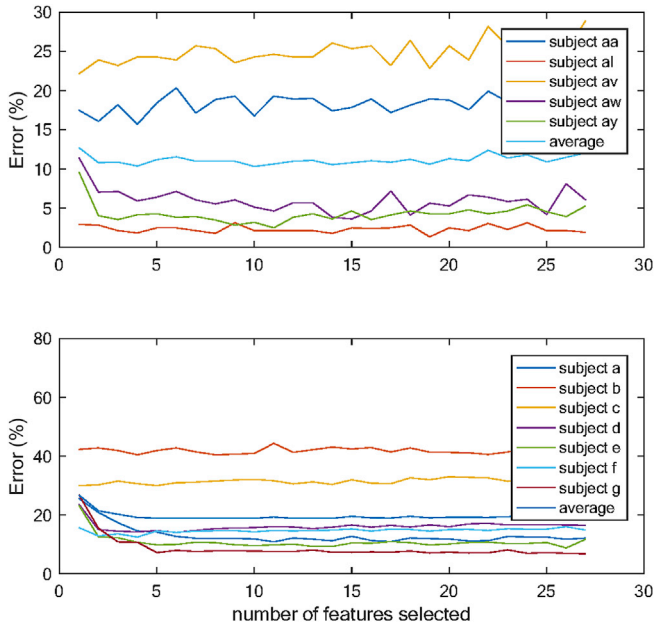

Fig. 7. Misclassification error rates using different number of selected features. Top: BCI Competition III dataset IVa. Bottom: BCI Competition IV dataset I).

Table 7
Complexity of TSM and CSP-TSM methods (where n is the number of channels).

Method	Complexity
CSP	O(n)
TSM	O (n^3)
CSP-TSM	O(n)

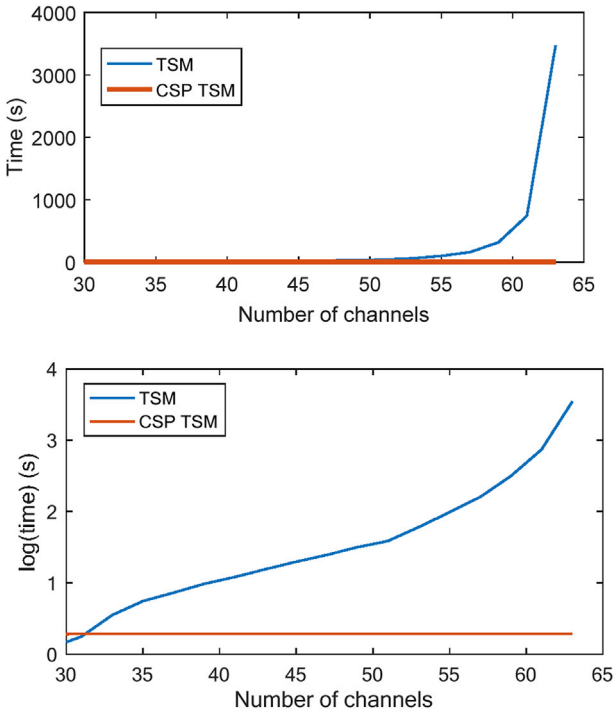

Fig. 8. The training time required by TSM and CSP-TSM algorithms using different number of channels data.

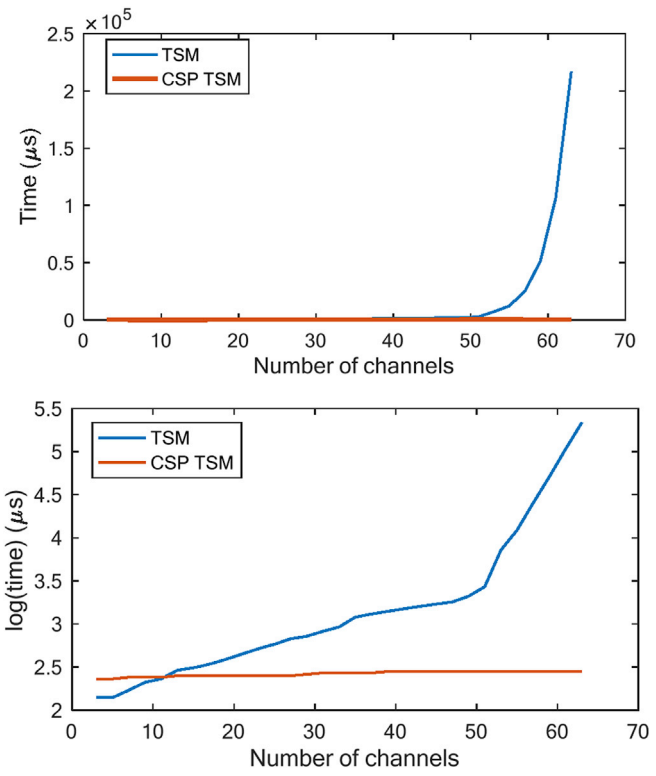

Fig. 9. The test time required by TSM and CSP-TSM algorithms using different number of channels data.

of BCI competition III dataset IVa, BCI competition IV dataset I and BCI competition IV dataset IIb, and has outperformed all other competing methods. We have also shown that the CSP-TSM framework enhances the performance of TSM method while also reducing the computational complexity of the TSM method. Future research should aim to test the proposed framework on a larger set of experimental data. Use of multiple bands with appropriate sub-bands such as that of FBCSP, DFBCSP, SFBCSP and SBLFB might further improve the performance of the proposed CSP-TSM framework. As future works, we will also investigate the use of weighted voting scheme for MI EEG signal classification using features from multiple filter banks.

Conflict of interest

None Declared.

Acknowledgement

The research was partly supported by College Research Committee (CRC) of Fiji National University, and the University of the South Pacific, Fiji (6C482-1351-Acct-00).

References

[1] L.F. Nicolas-Alonso, J. Gomez-Gil, Brain computer interfaces, a review, Sensors 12 (2012) 1211.
[2] A. Kübler, N. Neumann, J. Kaiser, B. Kotchoubey, T. Hinterberger, N.P. Birbaumer, Brain-computer communication: self-regulation of slow cortical potentials for verbal communication, Archives Phys. Med. Rehab. 82 (November 2001) 1533–1539.
[3] S.C. Kleih, T. Kuafmann, C. Zickler, S. Halder, F. Leotta, F. Cincotti, et al., Out of the frying pan into the fire–the P300-based BCI faces real-world challenges, Prog. Brain Res. 194 (2011) 27–46.
[4] F. Cincotti, D. Mattia, F. Aloise, S. Bufalari, G. Schalk, G. Oriolo, et al., Non-invasive brain–computer interface system: towards its application as assistive technology, Brain Res. Bull. 75 (April 2008) 796–803.
[5] D.J. McFarland, W.A. Sarnacki, J.R. Wolpaw, Electroencephalographic (EEG) control of three-dimensional movement, J. Neural Eng. 7 (May 2010) 036007.
[6] J.R. Wolpaw, D.J. McFarland, Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 17849–17854.
[7] A. Ramos-Murguialday, D. Broetz, M. Rea, L. Läer, Ö. Yilmaz, F.L. Brasil, et al., Brain–machine interface in chronic stroke rehabilitation: a controlled study, Ann. Neurol. 74 (2013) 100–108.
[8] M.D. Serruya, Bottlenecks to clinical translation of direct brain-computer interfaces, Front. Syst. Neurosci. 8 (2014) 226.
[9] S. Silvoni, A. Ramos-Murguialday, M. Cavinato, C. Volpato, G. Cisotto, A. Turolla, et al., Brain-computer interface in stroke: a review of progress, Clin. EEG Neurosci. 42 (October 2011) 245–252.
[10] F. Akram, M.K. Metwally, H. Hee-Sok, J. Hyun-Jae, K. Tae-Seong, A novel P300-based BCI system for words typing, in: Brain-computer Interface (BCI), 2013, pp. 24–25. International Winter Workshop on, 2013.
[11] M.H. Alomari, A. AbuBaker, A. Turani, A.M. Baniyounes, A. Manasreh, EEG mouse: a machine learning-based brain computer interface, Int. J. Adv. Comput. Sci. Appl. 5 (2014) 193–198.
[12] S. Ramesh, M.G. Krishna, M. Nakirekanti, Brain computer interface system for mind controlled robot using bluetooth, Int. J. Comput. Appl. 104 (October 2014) 20–23.
[13] R.S. Naveen, A. Julian, Brain computing interface for wheel chair control, in: Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 2013, pp. 1–5.
[14] D. La Rocca, P. Campisi, J. Sole-Casals, EEG based user recognition using BUMP modelling, in: International Conference of the Biometrics Special Interest Group (BIOSIG), 2013, pp. 1–12.
[15] S. Jirayucharoensak, S. Pan-Ngum, P. Israsena, EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation, Sci. World J. 2014 (2014) 10.
[16] P.P. Acharjee, R. Phlypo, L. Wu, V.D. Calhoun, T. Adali, Independent vector analysis for gradient artifact removal in concurrent EEG-fMRI data, IEEE Trans. Biomed. Eng. 62 (2015) 1750–1758.
[17] A. Sohrabpour, Y. Lu, P. Kankirawatana, J. Blount, H. Kim, B. He, Effect of EEG electrode number on epileptic source localization in pediatric patients, Clin. Neurophysiol. 126 (2015) 472–480.
[18] H. Woehrle, M.M. Krell, S. Straube, S.K. Kim, E.A. Kirchner, F. Kirchner, An adaptive spatial filter for user-independent single trial detection of event-related potentials, IEEE Trans. Biomed. Eng. 62 (2015) 1696–1705.
[19] T. Yu, J. Xiao, F. Wang, R. Zhang, Z. Gu, A. Cichocki, et al., Enhanced motor imagery training using a hybrid BCI with feedback, IEEE Trans. Biomed. Eng. 62 (2015) 1706–1717.
[20] E. Thomas, J. Fruitet, M. Clerc, Combining ERD and ERS features to create a system-paced BCI, J. Neurosci. methods 216 (June 2013) 96–103.
[21] G. Prasad, P. Herman, D. Coyle, Using motor imagery based brain-computer interface for post-stroke rehabilitation, in: 4th International IEEE/EMBS Conference on Neural Engineering, Antalya, Turkey, 2009.
[22] P.L. Jackson, M.F. Lafleur, F. Malouin, C. Richards, J. Doyon, Potential role of mental practice using motor imagery in neurologic rehabilitation, Archives Phys. Med. Rehab. 82 (August 2001) 1133–1141.
[23] Y. Liu, M. Li, H. Zhang, H. Wang, J. Li, J. Jia, et al., A tensor-based scheme for stroke patients' motor imagery EEG analysis in BCI-FES rehabilitation training, J. Neurosci. Methods 222 (January 2014) 238–249.
[24] R. Ortner, D.C. Irimia, J. Scharinger, C. Guger, A motor imagery based brain-computer interface for stroke rehabilitation, Stud. Health Technol. Inf. 181 (2012) 319–323.
[25] K.K. Ang, K.S. Chua, K.S. Phua, C. Wang, Z.Y. Chin, C.W. Kuah, et al., A randomized controlled trial of EEG-based motor imagery brain-computer interface robotic rehabilitation for stroke, Clin. EEG Neurosci. 46 (October 2015) 310–320.
[26] T. Wei-Peng, C. Effie, Is motor-imagery brain-computer interface feasible in stroke rehabilitation? PM&R 6 (August 2014) 723–728.
[27] B. S, S. U, H. S. M, Motor imagery based BCI for a maze game, in: Presented at the 4th International Conference on Intelligent Human Computer Interaction (IHCI), Kharagpur, 2012.
[28] T. Li, J. Zhang, T. Xue, B. Wang, Development of a novel motor imagery control technique and application in a gaming environment, Comput. Intell. Neurosci. (2017) 16.
[29] J. Asensio-Cubero, J.Q. Gan, R. Palaniappan, Multiresolution analysis over graphs for a motor imagery based online BCI game, Comput. Biol. Med. 68 (2016) 21–26.
[30] Z. Cheng, Y. Kimura, H. Higashi, T. Tanaka, A simple platform of brain-controlled mobile robot and its implementation by SSVEP, in: International Joint Conference on Neural Networks (IJCNN), 2012, pp. 1–7.
[31] D.J. McFarland, J.R. Wolpaw, Brain-computer interface operation of robotic and prosthetic devices, Computer 41 (2008) 52–56.
[32] (27 January). Neurosky: EEG Headsets. Available: http://store.neurosky.com/collections/eeg-headsets.
[33] (27 January). Emotiv eStore: EPOC Headset. Available: https://emotiv.com/store/.
[34] S. Lemm, B. Blankertz, G. Curio, K. Muller, Spatio-spectral filters for improving the classification of single trial EEG, IEEE Trans. Biomed. Eng. 52 (2005) 1541–1548.
[35] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, K.R. Muller, Combined optimization of spatial and temporal filters for improving brain-computer interfacing, IEEE Trans. Biomed. Eng. 53 (2006) 2274–2281.
[36] Q. Novi, G. Cuntai, T.H. Dat, X. Ping, Sub-band common spatial pattern (SBCSP) for brain-computer interface, in: 3rd International IEEE/EMBS Conference on Neural Engineering, 2007, pp. 204–207.

[37] K.K. Ang, Z.Y. Chin, H. Zhang, C. Guan, filter bank common spatial pattern (FBCSP) in brain-computer interface, in: IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, 2008, pp. 2390–2397.
[38] K.P. Thomas, G. Cuntai, C.T. Lau, A.P. Vinod, A. Kai Keng, A new discriminative common spatial pattern method for motor imagery brain computer interfaces, IEEE Trans. Biomed. Eng. 56 (2009) 2730–2733.
[39] Y. Zhang, G. Zhou, J. Jin, X. Wang, A. Cichocki, Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface, J. Neurosci. Methods 255 (November 2015) 85–91.
[40] Q. Wei, Z. Wei, Binary particle swarm optimization for frequency band selection in motor imagery based brain-computer interfaces, Bio Med. Mater. Eng. 26 (2015) S1523–S1532.
[41] H.I. Suk, S.W. Lee, A novel bayesian framework for discriminative feature extraction in brain-computer interfaces, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 286–299.
[42] Y. Zhang, Y. Wang, J. Jin, X. Wang, Sparse bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification, Int. J. Neural Syst. 27 (2017) 1650032.
[43] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.R. Muller, Optimizing spatial filters for robust EEG single-trial analysis, Signal Process. Mag. IEEE 25 (2008) 41–56.
[44] W. Wei, G. Xiaorong, H. Bo, G. Shangkai, Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL), IEEE Trans. Biomed. Eng. 55 (2008) 1733–1743.
[45] N. Tomida, T. Tanaka, S. Ono, M. Yamagishi, H. Higashi, Active data selection for motor imagery EEG classification, IEEE Trans. Biomed. Eng. 62 (2015) 458–467.
[46] L. Haiping, E. How-Lung, G. Cuntai, K.N. Plataniotis, A.N. Venetsanopoulos, Regularized common spatial pattern with aggregation for EEG classification in small-sample setting, IEEE Trans. Biomed. Eng. 57 (2010) 2936–2946.
[47] W. Samek, M. Kawanabe, K.R. Muller, Divergence-based framework for common spatial patterns algorithms, IEEE Rev. Biomed. Eng. 7 (2014) 50–72.
[48] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Multiclass brain–computer interface classification by riemannian geometry, IEEE Trans. Biomed. Eng. 59 (April 2012) 920–928.
[49] A. Sharma, K.K. Paliwal, A deterministic approach to regularized linear discriminant analysis, Neurocomputing 151 (Part 1) (March 2015) 207–214.
[50] A. Ashok, A.K. Bharathan, V.R. Soujya, P. Nandakumar, Tikhonov regularized spectrally weighted common spatial patterns, in: International Conference on Control Communication and Computing (ICCC), 2013, pp. 315–318.
[51] Z.J. Koles, The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG, Electroencephalogr. Clin. Neurophysiol. 79 (December 1991) 440–447.
[52] P.T. Fletcher, S. Joshi, Principal geodesic analysis on symmetric spaces: statistics of diffusion tensors, in: Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 87–98.
[53] O. Tuzel, F. Porikli, P. Meer, Pedestrian detection via classification on riemannian manifolds, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2008) 1713–1727.
[54] R. Tibshirani, Regression shrinkage and selection via the Lasso, J. R. Stat. Soc. Ser. B 58 (1996) 267–288.
[55] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J. Stat. Softw. 33 (2010).
[56] G. Dornhege, B. Blankertz, G. Curio, K. Muller, Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms, IEEE Trans. Biomed. Eng. 51 (2004) 993–1002.
[57] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, G. Curio, The non-invasive Berlin Brain–Computer Interface: fast acquisition of effective performance in untrained subjects, NeuroImage 37 (2007) 539–550.
[58] T. C. T. Limited, 10/20 System Positioning, 2012. Hong Kong.
[59] L. Song, J. Epps, Classifying EEG for brain-computer interfaces: learning optimal filters for dynamical system features, in: Presented at the Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, Pennsylvania, USA, 2006.
[60] A.S. Aghaei, M.S. Mahanta, K.N. Plataniotis, Separable common spatio-spectral patterns for motor imagery BCI systems, IEEE Trans. Biomed. Eng. 63 (2016) 15–29.
[61] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (March 1977) 159–174.
[62] S. Herculano-Houzel, The human brain in numbers: a linearly scaled-up primate brain, Front. Hum. Neurosci. 3 (September 2009) 31.
[63] M.A. Jatoi, N. Kamel, A.S. Malik, I. Faye, T. Begum, A survey of methods used for source localization using EEG signals, Biomed. Signal Process. Control 11 (May 2014) 42–52.
[64] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (2002) 767–791.
[65] J.S. Chiu, Y.C. Li, F.C. Yu, Y.F. Wang, Applying an artificial neural network to predict osteoporosis in the elderly, Stud. Health Technol. Inf. 124 (2006) 609–614.
[66] J.S. Chiu, C.F. Chong, Y.F. Lin, C.C. Wu, Y.F. Wang, Y.C. Li, Applying an artificial neural network to predict total body water in hemodialysis patients, Am. J. Nephrol. 25 (2005) 507–513.
[67] M.M. El Bahy, M. Hosny, W.A. Mohamed, S. Ibrahim, EEG signal classification using neural network and support vector machine in brain computer interface, in: A.E. Hassanien, K. Shaalan, T. Gaber, A.T. Azar, M.F. Tolba, ed Cham (Eds.), Proceedings of the International Conference on Advanced Intelligent Systems and Informatics, Springer International Publishing, 2017, pp. 246–256.
[68] Y. Ma, X. Ding, Q. She, Z. Luo, T. Potter, Y. Zhang, Classification of motor imagery EEG signals with support vector machines and particle swarm optimization, Comput. Math. Methods Med. 2016 (2016) 8.
[69] S. Kumar, A. Sharma, K. Mamun, T. Tsunoda, A deep learning approach for motor imagery EEG signal classification, in: Presented at the 3rd Asia-Pacific World Congress on Computer Science and Engineering, Denarau Island, Fiji, 2016.
[70] W. Wu, Z. Chen, X. Gao, Y. Li, E.N. Brown, S. Gao, Probabilistic common spatial patterns for multichannel EEG analysis, Pattern Anal. Mach. Intell. IEEE Trans. 37 (2015) 639–653.
[71] D. Enzeng, L. Liting, C. Chao, Improved common spatial pattern for brain-computer interfacing, in: IEEE International Conference on Mechatronics and Automation (ICMA), 2015, pp. 2112–2116.
[72] H. Higashi, T. Tanaka, Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification, IEEE Trans. Biomed. Eng. 60 (2013) 1100–1110.
[73] F.F. Gonen, G.V. Tcheslavski, Techniques to assess stationarity and gaussianity of EEG: an overview, Int. J. Bioautomotion 16 (2012) 135–142.
[74] J.A. McEwen, G.B. Anderson, Modeling the stationarity and gaussianity of spontaneous electroencephalographic activity, IEEE Trans. Biomed. Eng. BME-22 (1975) 361–369.
[75] A. Sharma, K.K. Paliwal, S. Imoto, S. Miyano, A feature selection method using improved regularized linear discriminant analysis, Mach. Vis. Appl. 25 (2014) 775–786.
[76] A. Yuksel, T. Olmez, A neural network-based optimal spatial filter design method for motor imagery classification, PLoS One 10 (2015) e0125039.
[77] B. Yang, H. Li, Q. Wang, Y. Zhang, Subject-based feature extraction by using fisher WPD-CSP in brain–computer interfaces, Comput. Methods Programs Biomed. 129 (June 2016) 21–28.
[78] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Common spatial pattern revisited by riemannian geometry, in: IEEE International Workshop on Multimedia Signal Processing (MMSP), 2010, pp. 472–476.
[79] A. Sharma, D. Shigemizu, K.A. Boroevich, Y. López, Y. Kamatani, M. Kubo, et al., Stepwise iterative maximum likelihood clustering approach, BMC Bioinforma. 17 (2016) 1–14.
[80] D. Charalampidis, A modified K-Means algorithm for circular invariant clustering, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1856–1865.
[81] T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, A.Y. Wu, An efficient k-means clustering algorithm: analysis and implementation, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 881–892.
[82] A. Sharma, K. Boroevich, D. Shigemizu, Y. Kamatani, M. Kubo, T. Tsunoda, Hierarchical maximum likelihood clustering approach, IEEE Trans. Biomed. Eng. 64 (2017) 112–122.
[83] S. Kumar, R. Sharma, A. Sharma, T. Tsunoda, Decimation filter with common spatial pattern and fishers discriminant analysis for motor imagery classification, in: Presented at the IEEE World Congress on Computational Intelligence, Vancouver, Canada, 2016.