



# A Novel Features Selection Approach with Common Spatial Pattern for EEG Based Brain–Computer Interface Implementation

Sujan Ali, Jannatul Ferdous, Ekramul Hamid & Khademul Islam Molla

To cite this article: Sujan Ali, Jannatul Ferdous, Ekramul Hamid & Khademul Islam Molla (2019): A Novel Features Selection Approach with Common Spatial Pattern for EEG Based Brain–Computer Interface Implementation, IETE Journal of Research, DOI: [10.1080/03772063.2019.1670106](https://doi.org/10.1080/03772063.2019.1670106)

To link to this article: <https://doi.org/10.1080/03772063.2019.1670106>



Published online: 09 Oct 2019.



Submit your article to this journal [↗](#)




View related articles [↗](#)



View Crossmark data [↗](#)

# A Novel Features Selection Approach with Common Spatial Pattern for EEG Based Brain–Computer Interface Implementation

Sujan Ali<sup>1</sup>, Jannatul Ferdous <sup>1</sup>, Ekramul Hamid<sup>2</sup> and Khademul Islam Molla<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, Jatiya Kabi Kazi Nazrul Islam University, Mymensingh, Bangladesh; <sup>2</sup>Department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh

## ABSTRACT

This paper presents a novel method for the selection of spatial filters and features in electroencephalography (EEG) based motor imagery classification. The analyzing EEG data are divided into training and test sets. The training set is used to select appropriate spatial filters with dominant features. To accomplish such features, the EEG of training set is segmented again into two subsets termed as training subset and test subset. The features of both subsets are extracted using common spatial pattern. Then features of training subset are ranked using mutual information based approach. Besides, the features of test subset are also ranked according to the order of the training subset features. The initial classification performance using training and test subsets are obtained by using linear discriminant analysis. Then a grid search method is employed to select the effective number of spatial filter pairs as well as the discriminative features. Thus obtained spatial filter and features are used in actual classification accuracy of the test set of EEG. The experimental results show that the proposed approach yields comparatively superior classification performance compared to prevailing methods.

## KEYWORDS

Brain–computer interface (BCI);  
Electroencephalography;  
Common spatial pattern;  
Grid search

## 1. INTRODUCTION

A brain–computer interface (BCI) is an alternative communicating system between a brain and an external device that translates neuronal activities into user commands.

To control communication [1], movement [2], and environment [3] the BCI systems have been used. Moreover, the systems have been employed in neuro-rehabilitation [4–6], P300-based word typing system [7], biometric identification [8], and brain wave-controlled robots and wheelchairs for the disabled people [9,10]. The key concentration and importance of the current BCI research are found in biomedical engineering [4–6,11–15]. The main target of the BCI is to restore or repair useful functions to people disabled by neuromuscular disorders such as Amyotrophic Lateral Sclerosis (ALS), cerebral palsy, stroke, spinal cord injury, and so on. Although people may become totally paralyzed through these types of disorders their minds remain unaffected. Considering this issue BCI system translates human thoughts directly to the external world [16]. The neuronal activities are generally measured by electroencephalography (EEG). The EEG is a widely used non-invasive BCI due to its low expense and high-temporal resolution [17]. In order to command the external devices for various purposes, the user needs to generate different signal patterns with his

brain. The most important function of the BCI system is to discriminate the EEG patterns and make them control commands over the external devices. However, the EEG based BCI need constant high concentration and training. This type of concentration is not essential for tongue movement. The limitations of the BCIs can overcome by glossokinetic potential based tongue machine interface [18].

The neurophysiological event-related desynchronization (ERD) or event-related synchronization (ERS) phenomenon during the imagination of body-part movement is called motor imagery (MI), which reflects sensorimotor brain activity in the alpha and beta bands (8–30 Hz)[19,20]. The ERD/ERS patterns are highly inconsistent across subjects and even trials for the same subject [21]. Hence, the classification of ERD/ERS associated class discriminative spatial patterns is the major-concern in the BCI research. In BCI, before classification feature extraction is very important part. The features of the EEG signals can be extracted with average signal power, root mean square (RMS), standard deviation (SD), variance (VAR), and mean value (MV) [18,22]. To reduce the high dimensional EEG data, the statistical algorithms, principal component analysis (PCA), and independent component analysis (ICA) are used in [23,24].

Due to the topographical organization in the motor and somatosensory cortex, MI tasks can be identified on the basis of their spatial location [25]. EEG signals recorded from a particular area of scalp are the combination of various cortical sources located in different areas because of volume conduction effect. Hence, raw scalp EEG potentials have poor spatial resolution [25]. A spatial filtering is an essential step to remove the volume conduction effect and reach the actual underlying signal sources [26].

Common Spatial Pattern (CSP) is an algorithm commonly used in BCI systems to preprocess the EEG signals [21,27,28]. The algorithm finds optimal spatial filters that are functional in discriminating two classes of EEG signals in MI-based BCI. It maximizes the ratio of average variances that belong to two different classes. A computed CSP spatial filter projects the multi-dimensional EEG time domain signal into a one-dimensional time domain signal. The most challenging task of CSP based BCI is to select the spatial filters for the extraction of features. In all of the existing works the spatial filters are selected manually. For example, in [29,30], 3 pairs of filters are used as recommended in [25]; only single filter pair is used in [31,32]; in [33] experimentation is prepared by using 4 pairs of filters, and in [34] the classification performance with 1, 2, and 3 pairs of filters is compared. However, the manual selection of CSP filters does not confirm that the approach will yield best accuracy.

In this paper, we propose a novel approach of selecting the best CSP filter pair and corresponding most discriminative features which will eventually improve classification performance. In this approach, grid search method is used to select the filter pair and features by searching the best linear discriminant analysis (LDA) scores.

The paper is organized as follows – Section 2 and Section 3 discusses literature review and feature extraction techniques, respectively. Section 4 contains the description of the proposed method. The basics of linear discriminant analysis (LDA) technique is explained in Section 5 and the experimental results are illustrated in Section 6. Finally, Section 7 includes discussion and some concluding remarks.

## 2. LITERATURE REVIEW

The CSP algorithm handles two classes at the same time and simultaneously diagonalizes the covariance matrices of both classes [35]. Also, it is proved that the algorithm

is efficient in BCI competitions [36,37]. CSP optimizes the ratio of average variances of two classes, and consequently requires only one average covariance matrix for each class. It creates difficulty during handling EEG like non-stationary signals as the covariance matrix of an EEG signal may change over time due to artifact [38]. The fitness function constraint is another drawback of the CSP. CSP does not allow different types of fitness function, which may be more useful in different situation [39]. Overfitting occurs during the optimization of the ratio of average variances of the two classes due to its outlier sensitivity [40].

The shortcomings of the CSP lead researchers to breed some modified versions of the CSP. Regularized CSP (RCSP) is one of them introduced, aiming at computing optimal spatial pattern by familiarizing a regularization term to the CSP formula [26,29,38]. The RCSP method attempts to use a priori knowledge in the spatial filter optimization by imposing variant constraints in the CSP's formulation [39]. Some examples of RCSP method are (i) Composite CSP (CCSP) algorithm: it works on the basis of useful information transfer from subject to subject by regularizing the covariance matrices and using other subject's data [34], (ii) Regularized CSP with Generic Learning (GLRCSP) approach: the aim of this approach is to regularize the covariance matrix estimation, using data from other subjects [29,41]. It uses regularization parameters to shrink the covariance matrix towards both the identity and a generic covariance matrix, (iii) Regularized CSP with Selected Subjects (SSRCSP): unlike CCSP or GLRCSP, SSRCSP uses data only from selected subjects. Due to potentially large inter-subjects variability, the data can be irrelevant to all of them [29], (iv) Spatially Regularized CSP (SRCSP): the main target of this approach is to obtain spatially smooth filters for which neighboring electrodes have similar weights [29]. Here, the a priori knowledge that neighboring neurons tend to have as a similar function is used, (v) stationary CSP (sCSP): the purpose of the sCSP is to extract features that are stationary as well as discriminative between two classes [38]. sCSP assumes that non-stationaries in EEG data come from processes that are not task-related. (vi) CSP with Tikhonov Regularization (TRCSP): the TRCSP algorithm [29] functions based on the regularization of the CSP. The regularization involves penalizing results with large weights. Since the CSP formulation considers the variances of two classes, CSP is very sensitive to outlier that sometimes limits the effectiveness of the approach. A robust L1-norm based CSP algorithm is called CSP-L1 that efficiently suppresses the leading influence of the outliers [42].

### 3. FEATURE EXTRACTION METHODS

#### 3.1 The CSP Algorithm

CSP is a feature extraction technique used in signal processing for separating a multivariate signal into additive sub-components. The technique is used to design spatial filters; that is the variance of the filtered data from one class is maximized while the variance of the filtered data from the other class is minimized. Thus, the resulting feature vectors increase the discriminability between the two classes by means of minimizing the intra class variance and maximizing the inter class variance [43]. This property builds CSP as one of the most effective spatial filters for EEG signal processing. The method of CSP was first introduced to EEG analysis for detection of abnormal EEG [44] and effectively applied to movement-related EEG for the classification purpose [25,27]. The target of the CSP is to project the multichannel EEG data into low dimensional spatial subspace with a projection matrix, using linear transformation [45].

For details explanation of the CSP algorithm, let us assume the original EEG data matrix  $E_k^i$  from trial  $i$  for class  $k$ . The dimension of each  $E_k^i$  is  $D \times T$ , in which  $D$  is the number of channels and  $T$  is the number of samples per channel. For the explanation, the EEG data of a single trial ( $i = 1$ ) is represented as  $E_{k \in (h,f)}$  in which  $h$  denotes hand and  $f$  denotes foot movement. The normalized spatial covariance of the EEG for hand movement,  $C_h$  and for the foot movement,  $C_f$  can be calculated as:

$$C_h = \frac{E_h E_h'}{\text{tr}(E_h E_h')}, \quad C_f = \frac{E_f E_f'}{\text{tr}(E_f E_f')} \quad (1)$$

where  $E_h$  and  $E_f$  represent the original EEG matrices for hand and foot movement respectively,  $(\cdot)'$  is the transpose operator and  $\text{tr}(\cdot)$  represents the sum of the diagonal elements of any given matrix. The composite spatial covariance,  $C$  is the sum of the averaged normalized spatial covariance  $\bar{C}_h$  and  $\bar{C}_f$ . The  $\bar{C}_h$  and  $\bar{C}_f$  are estimated by averaging over all the trials of each class. The composite spatial covariance,  $C$  is calculated as

$$C = \bar{C}_h + \bar{C}_f = \Sigma \lambda \Sigma' \quad (2)$$

where  $\Sigma$  is the matrix of Eigenvectors and  $\lambda$  is the diagonal matrix of Eigenvalues. The averaged normalized spatial covariance  $\bar{C}_h$  and  $\bar{C}_f$  are transformed as

$$J_h = X \bar{C}_h X' \quad \text{and} \quad J_f = X \bar{C}_f X' \quad (3)$$

where  $X = \Sigma' / \sqrt{\lambda}$  is the whitening transformation matrix.  $J_h$  and  $J_f$  share common eigenvectors and the sum of corresponding eigenvalues for the two matrices

will always be one. If  $J_h = Y \Lambda_h Y'$  and  $J_f = Y \Lambda_f Y'$ , then  $\Lambda_h + \Lambda_f = I$ , where  $I$  is the identity matrix. Since the sum of two corresponding eigenvalues is always one, a high eigenvalue for  $J_h$  means a high variance for EEG in hand movement and a low variance for the EEG in foot movement (low eigenvalue for  $J_f$ ) and vice versa. The classification operation is done based on this property. The projection of whitened EEG into the eigenvectors  $Y$  corresponding to the largest  $\Lambda_h$  and  $\Lambda_f$  will give feature vectors that significantly enhance the discrimination ability.

The goal of the CSP is to find  $B$  spatial filters to create a projection matrix  $W$  of dimension  $D \times B$  (each column is a spatial filter). The projection matrix  $W$  is represented as

$$W = Y' X \quad (4)$$

The projection matrix  $W$  transforms the original EEG linearly into uncorrelated components, according to:

$$Z = W^T E \quad (5)$$

In the original EEG,  $E$  can be reconstructed by  $E = W^{-1} Z$  where  $W^{-1}$  is the inverse matrix of  $W$ . The columns of  $W^{-1}$  are spatial patterns that describe the variance of the EEG. The first and last columns contain the most discriminatory spatial patterns that explain the high variance of one class and the low variance of the other.

#### 3.2 The Composite CSP (CCSP)

The CSP algorithm does not consider the inter-subject information. It only exploits covariance matrices on a subject-by-subject basis. The Composite CSP (CCSP) algorithm, a modification of the CSP algorithm is made by Kang *et al.* [34] in which they consider subject-to-subject transfer and exploit a linear combination of covariance matrices. In the CCSP, covariance matrices,  $M_c$  can be measured as a weighted sum of the covariance matrices of the other subjects,  $M_c = \sum_{s \in B} U_c^s V_c^s V_{t,c}^{-1}$  where

$B$  is a set of subjects,  $U_c^s$  is the spatial covariance matrix for class  $c$  and subject  $s$ ,  $V_c^s$  is the number of EEG trials used to calculate  $U_c^s$ , and the total number of EEG trials for class  $c$  is  $V_{t,c}$ .

#### 3.3 Spatially Regularized CSP (SRCSP)

The CSP algorithm overlooks the spatial position of EEG electrodes though it is used to learn spatial filters. The

main target of the Spatially Regularized CSP (SRCSP) is to consider this spatial information [29]. The aim of this algorithm is to get spatially smooth filters so that adjacent electrodes can ensure comparable weights. The spatial smoothness of the filters  $\phi$ , can be achieved by using Laplacian penalty term and the regularization matrix  $R$  [29,46].

$$R = D_N - N, \quad N(p, q) = \exp\left(-\frac{\|\omega_p - \omega_q\|^2}{2d^2}\right) \quad (6)$$

where  $\omega_p$  is the vector containing the 3D coordinates of the  $p^{\text{th}}$  electrode,  $D_N$  is the diagonal matrix and  $d$  is the distance between two electrodes. Once the weights of adjacent electrodes are very close to each other, the penalty term  $P(\phi) = \phi^T R \phi$  will be of small value since  $\phi^T (D_N - N) \phi = \sum_{p,q} N(p, q) (\phi_p - \phi_q)^2$  [47].

### 3.4 Regularized CSP with Generic Learning (GLRCSP)

The Regularized CSP with Generic Learning (GLRCSP) [41] algorithm uses data from other subjects and regularizes the estimation of the covariance matrix. The identity matrix and generic covariance matrix are shrunk by the algorithm. Like CCSP, in this algorithm the covariance matrix  $M_c$  is calculated using the covariance matrices of other subjects.

$$M_c = H_M \sum_{s \in B} U_c^s, \quad H_M = \frac{1}{(1 - \beta)G_c + \beta \sum_{s \in B} G_{U_c^s}} \quad (7)$$

where  $G_c$  is the number of trials used to calculate the covariance matrix  $c$  and  $\beta$  is the regularization term.

### 3.5 CSP with Tikhonov Regularization (TRCSP)

The CSP with Tikhonov Regularization (TRCSP) [29] works grounded on the regularization of the CSP with quadratic penalties. The Tikhonov Regularization is

made up of penalizing explanation with large weights. When the penalty term  $P(\phi) = \|\phi\|^2 = \phi^T \phi = \phi^T I \phi$ , the TRCSP is obtained by using regularization matrix  $R = I$ .

## 4. PROPOSED METHOD

A common problem of the CSP is that the relation between the CSP filter optimization and the classification performance is not direct enough. The selected spatial filters do not certainly map the signals into spaces where the best classification accuracy is found by the corresponding extracted features [39]. The proposed CSP based Grid Search (GS-CSP) method searches the best combination of spatial filter pair and features (SFPP) that provide maximum classification accuracy. In the pre-processing stage, a number of channels that roughly cover the motor cortex of the brain are selected. The training and test sets of the raw EEG are separated and individually filtered by a band pass filter. Figure 1 shows the block diagram of the proposed GS-CSP approach. For the selection of suitable SFPP, the training set EEG signal is used. The GS-CSP comprises three stages and detail of each is described below.

### 4.1 Spatial Filtering and Features Extraction

The CSP algorithm is greatly successful in calculating spatial filters. In this stage, the filtered training set is divided into two subsets: training subset and test subset. To calculate spatial filters the CSP algorithm is used on the training subset. The filtering is accomplished by linearly transforming the EEG measurements by using Equation (5). Using the calculated spatial filters, CSP features are then extracted from each of the trial of both training and test subset of EEG. It is done by projecting the EEG data into the CSP filters. The CSP features are generated as the log variance of the projected signals.



**Figure 1:** Block diagram of the proposed GS-CSP based spatial filter pair and features selection

## 4.2 Features Ranking

In this stage the extracted training subset features are ranked on the basis of values of mutual information with training subset EEG class. The test subset features are ranked in view of the ranking of training subset features. Mutual information is a measure of the amount of information by which one random variable tells about another random variable. The Mutual Information ( $MI$ ) between two random variables  $\mathbf{A}$  and  $\mathbf{B}$  is [48]

$$MI(\mathbf{A}; \mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}). \quad (8)$$

where  $\mathbf{A} = \{A_1, A_2, \dots, A_d\}$  is  $ad$ -dimensional random variable. The entropy of  $\mathbf{A}$  is

$$H(\mathbf{A}) = - \sum_{a \in \mathbf{A}} p(a) \log_2 p(a); \quad (9)$$

The conditional entropy of random variables  $\mathbf{A}$  and  $\mathbf{B}$  is

$$H(\mathbf{B}|\mathbf{A}) = - \sum_{a \in \mathbf{A}} \sum_{b \in \mathbf{B}} p(a, b) \log_2 p(b|a); \quad (10)$$

where  $p(\cdot)$  are probability mass functions.

In the EEG classification case, the CSP features are continuous variables and class levels are discrete values. Hence, the  $MI$  between the CSP features  $\mathbf{X}$  and class  $\mathbf{X}_C$  is calculated as [33]

$$MI(\mathbf{X}; \mathbf{X}_C) = H(\mathbf{X}_C) - H(\mathbf{X}_C|\mathbf{X}), \quad (11)$$

where  $\varphi \in \mathbf{X}_C = \{1, \dots, N_\varphi\}$ ; and the conditional entropy is

$$H(\mathbf{X}_C|\mathbf{X}) = - \int_{\mathbf{X}} \sum_{\varphi=1}^{N_\varphi} p(\varphi|x) \log_2 p(\varphi|x) dx, \quad (12)$$

where  $N_\varphi$  is the number of classes.

## 4.3 Filter Pair and Features Selection

In the third stage, the GS-CSP based SFPF selection is implemented with LDA. During LDA training the discriminant hyperplane coefficients  $\kappa_i$  are computed for the training subset ranked features  $i$  ( $i = 1, \dots, 2m$  where  $m$  is the CSP filter pair). The LDA accuracy score of the test subset ranked feature  $\theta_j$  ( $j = 1, \dots, 2m$ ) is now measured by the computed coefficients. For every value of  $\kappa_i$  and their corresponding test subset features  $\theta_j$ , LDA accuracy scores are measured as  $S_{m,j} = \kappa_i \Theta \theta_j$  where the symbol  $\Theta$  represents LDA test operator. In this approach, a grid of accuracy scores is formed from where the appropriate

SFPF by which the maximum accuracy score is yielded are selected.

The SFPF selection method is summarized as:

- (1) Learning spatial filters by the training subset EEG using CSP algorithm
- (2) Extracting CSP features by the spatial filters from both training subset and test subset data. For  $m$  pairs of spatial filters,  $2m$  numbers of CSP features are extracted
- (3) Calculating mutual information of the extracted training subset features with their corresponding EEG class
- (4) Ranking the extracted training subset features on the basis of values of the mutual information. Also the extracted test subset features are ranked according to the order of training subset features.
- (5) Computing discriminant hyperplane coefficients (bias and slope of the discriminant hyperplane)  $\kappa_i$  for the ranked training subset features.
- (6) Measuring the LDA accuracy score  $S_{m,j}$  of ranked test subset feature  $\theta_j$  based on the computed coefficients  $\kappa_i$
- (7) Finally, selecting the best SFPF that produce maximum LDA accuracy scores.

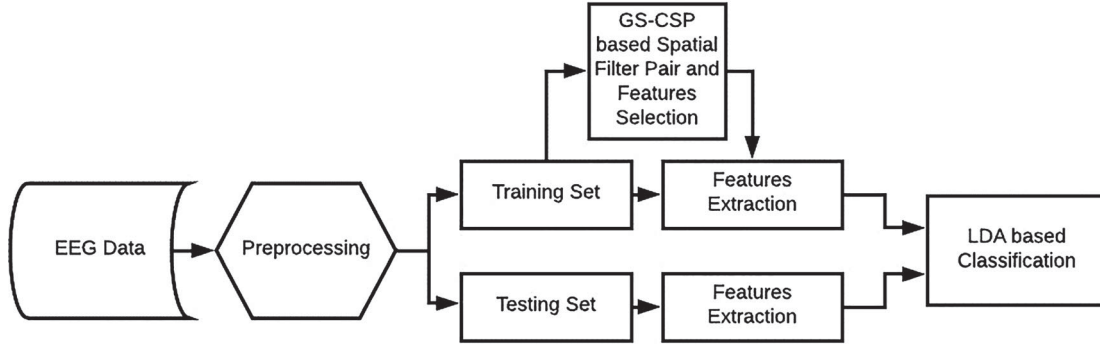
## 5. LDA BASED CLASSIFICATION

LDA, also known as Fisher's linear discriminant analysis is a technique used to find a linear combination of features that separates two or more classes of data. It is typically used as a dimensionality reduction step before classification [49]. It reduces dimensionality but at the same time preserves as much of the class discriminatory information as possible. The goal of the LDA is to use a separating hyperplane that maximally separates the data representing different classes. The hyperplane is found by selecting the projection, in which the same classes are projected very close to each other, and the distance between the values of two classes remains as maximum as possible [50].

Let us assume that we have  $K$  classes, each containing  $N$  observations  $x_i$ . The within-class scatter,  $\tilde{S}_w$  for all  $K$  classes can be calculated as:

$$\tilde{S}_w = \sum_{k=1}^K f_k S_w^k \quad (13)$$

where the within-class covariance matrix  $S_w^k$  and the fraction of data  $f_k$  are calculated according to the following



**Figure 2:** Block diagram of motor imagery based BCI with the proposed approach

formulas:

$$S_w^k = \sum_{i=1}^{N_k} (x_i^k - \mu^k)(x_i^k - \mu^k)^T \quad (14)$$

$$f_k = \frac{N_k}{\sum_{j=1}^K N_j} \quad (15)$$

where  $N_k$  is the number of observations of  $k^{\text{th}}$  class and  $\mu^k$  indicates the means of the all observations  $x_i$  for  $k^{\text{th}}$  class. The between-class scatter  $\tilde{S}_b$  for all  $K$  classes is calculated as:

$$\tilde{S}_b = \sum_{k=1}^K f_k S_b^k \quad (16)$$

where the between-class covariance matrix,  $S_b^k$  can be estimated as

$$S_b^k = \sum_{k=1}^K (\mu^k - \mu)(\mu^k - \mu)^T \quad (17)$$

where  $\mu$  indicates the means of all observations  $x_i$  for all classes. The main objective of LDA is to find a projection matrix that maximizes the ratio of the determinant of  $\tilde{S}_b$  to the determinant of  $\tilde{S}_w$ . The projections providing the best class separation are eigenvectors with the highest eigenvalues of matrix  $M$ :

$$M = \frac{\tilde{S}_b}{\tilde{S}_w} \quad (18)$$

Since the matrix  $M$  is asymmetric, the calculation of eigenvectors can be difficult. This difficulty can be minimized by using generalized eigenvalue problem [44]. Now, the aim of the LDA is to seek  $(K-1)$  projections  $[y_1, y_2, y_3, \dots, y_{K-1}]$  by means of  $(K-1)$  projection

vectors. The transformed data set  $y$  is obtained as a linear combination of all input features  $x$  with weights  $W$ .

$$y = x^T W \quad (19)$$

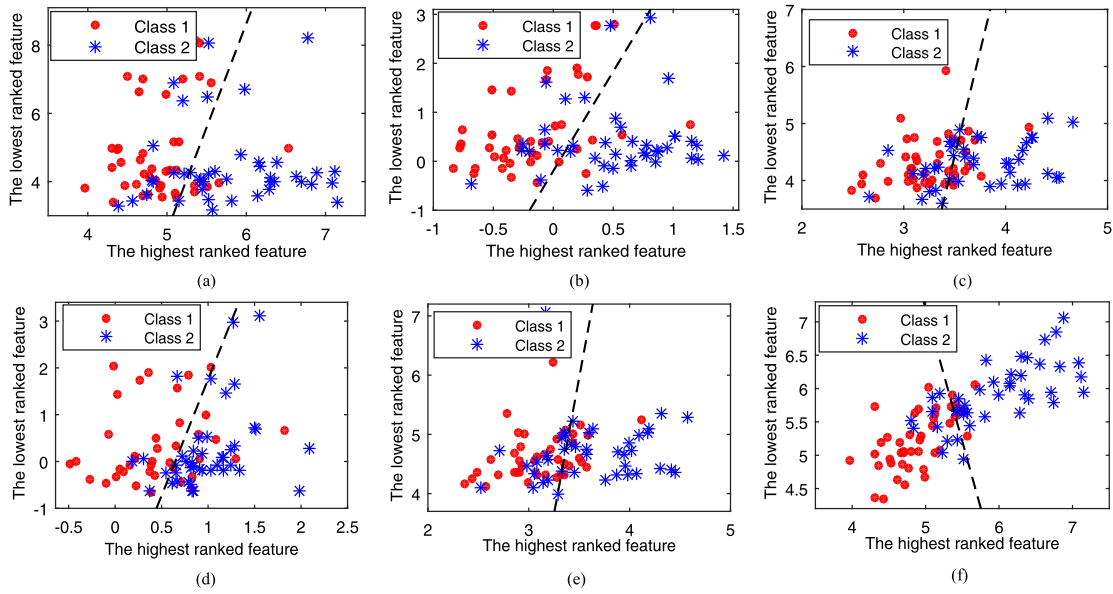
where  $W = [w_1, w_2, w_3, \dots, w_H]$  is a matrix form with the  $H$  eigenvectors of matrix  $M$  associated with the highest eigenvalues. The LDA reduces the original feature space dimension to  $H$ . The LDA performs well when the discriminatory information of data depends on the means of the data. But it does not work for the variance depended discriminatory informative data. Also, the performance of the LDA is not good for nonlinear classification. To implement MI based BCI, the work flow of this paper is presented with the block diagram as shown in Figure 2. The GS-CSP approach is employed on the training set EEG data to select the best spatial filter pairs and a number of dominating features. The CSP features of training set are finally extracted, using the selected filter pairs. Besides, the selected features are used in the LDA classifier to classify the testing set.

## 6. EXPERIMENTAL RESULTS

In this section, the performance of the proposed method is evaluated by conducting classification experiments on MI movement data. We applied the proposed method to two undermentioned datasets, where the datasets are widely used as publicly available dataset of Dataset IIIa and IVa from BCI competition III [21].

### 6.1 Dataset IVa, BCI Competition III (BCIC III-IVa)

The dataset contains data from four initial sessions without feedback and is recorded from five healthy subjects (labeled *aa*, *al*, *av*, *aw*, *ay*) who performed right hand and right foot movement imagination. A total of 118 electrodes are used for recording EEGs with a sample rate of 100 Hz. A training set and test set with different size for each subject is available. The data for each subject



**Figure 3:** Features extracted with different types of spatial filtering method using spatial filter pair  $m = 3$  and the GS-CSP using  $FP = 3$ ,  $F_s = 2$  from the training set of subject *av* of dataset *BCIC III-IVa*: (a) CSP, (b) CCSP, (c) SRCSP, (d) GLRCSP, (e) TRCSP and (f) GS-CSP. The class borders are indicated by dashed lines

comprises 280 trials, among which 168, 224, 84, 56, and 28 composed the training set for subject *aa*, *al*, *av*, *aw*, and *ay*, respectively, while the remaining of the trials composed their test set.

**6.2 Dataset IIIa, BCI Competition III (BCIC III-IIIa)**

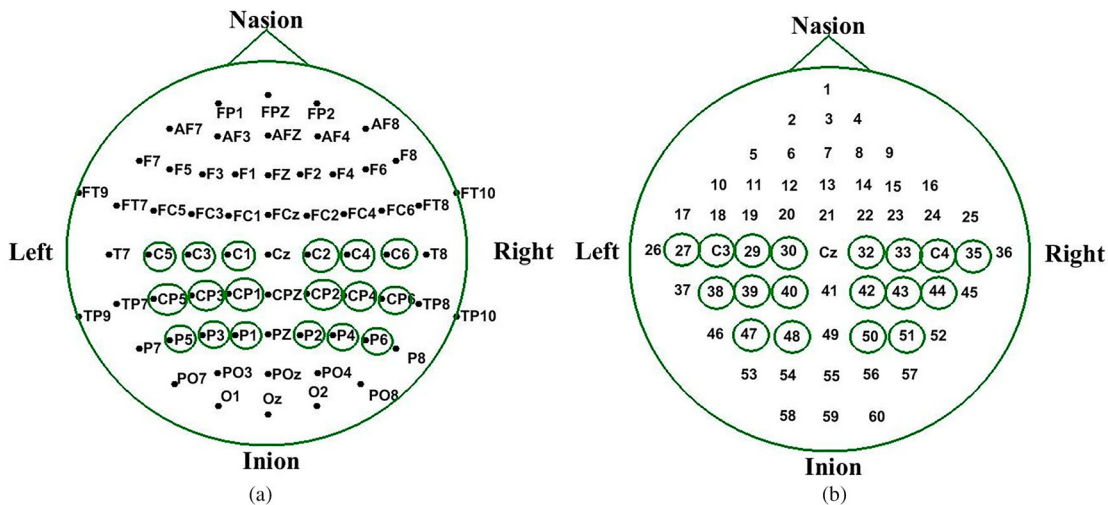
The dataset comprises EEG signals from three subjects who performed four different MI tasks: left hand, right hand, foot, and tongue. In this experiment, we consider MI binary classification: left-hand versus right-hand, left-hand versus foot, left-hand versus tongue, right-hand versus foot, right-hand versus tongue, and foot versus tongue.

This paper focuses on binary classification, not multiclass MI classification, since the target is to select a dominant feature. EEG signals are recorded using 60 electrodes with a sample rate of 250 Hz. A training set and a test set are available for each subject. Both sets contain 45 trials per class for subject *k3b*, and 30 trials per class for subject *k6b* and *l1b*.

**6.3 Pre-processing**

In this experiment, a class is assigned to each trial, i.e. the discrete classification of each class is considered.

The time of 0–0.5 s and 3.5–4.0 s are imagination preparation stage and post imagination stage, respectively.



**Figure 4:** Selected electrodes (circled) for (a) BCI competition III dataset IVa and (b) BCI competition III dataset IIIa

During this imagination features do not contain significant information [51]. For each dataset and trial, the data from time segment located from 0.5 to 2.5 s after the visual cue instructing the subject to perform MI is considered. The brain rhythmic components alpha contains frequency band 8–12 Hz and beta contains frequency band 12–30 Hz. A fourth-order Butterworth band pass

IIR filter with zero phases is used to filter the alpha and beta rhythmic components (8–30 Hz) from each trial.

#### 6.4 Channel Selection

The MI response of brain is more active in the motor cortex [52]. In this experiment, out of the 118 EEG

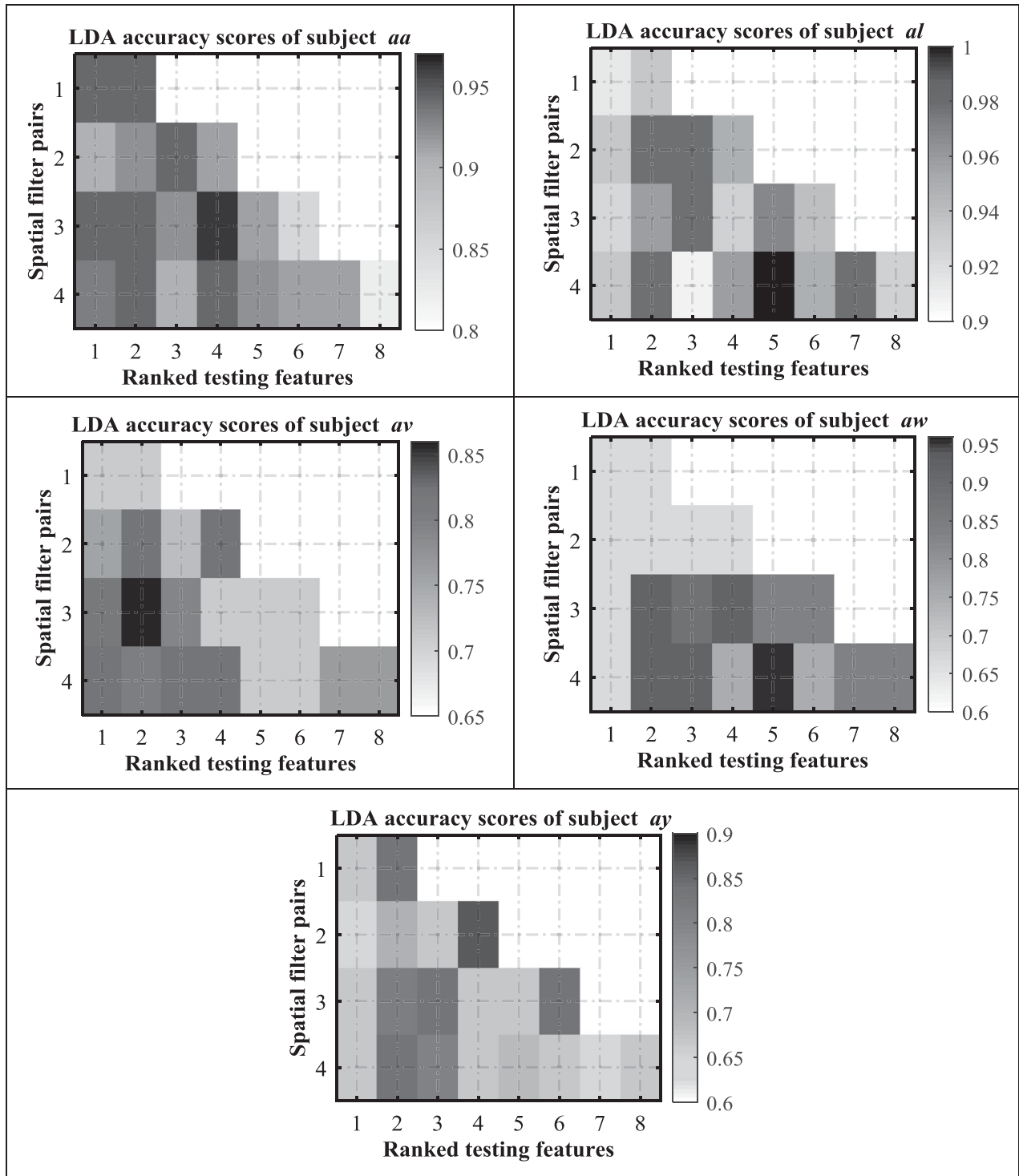
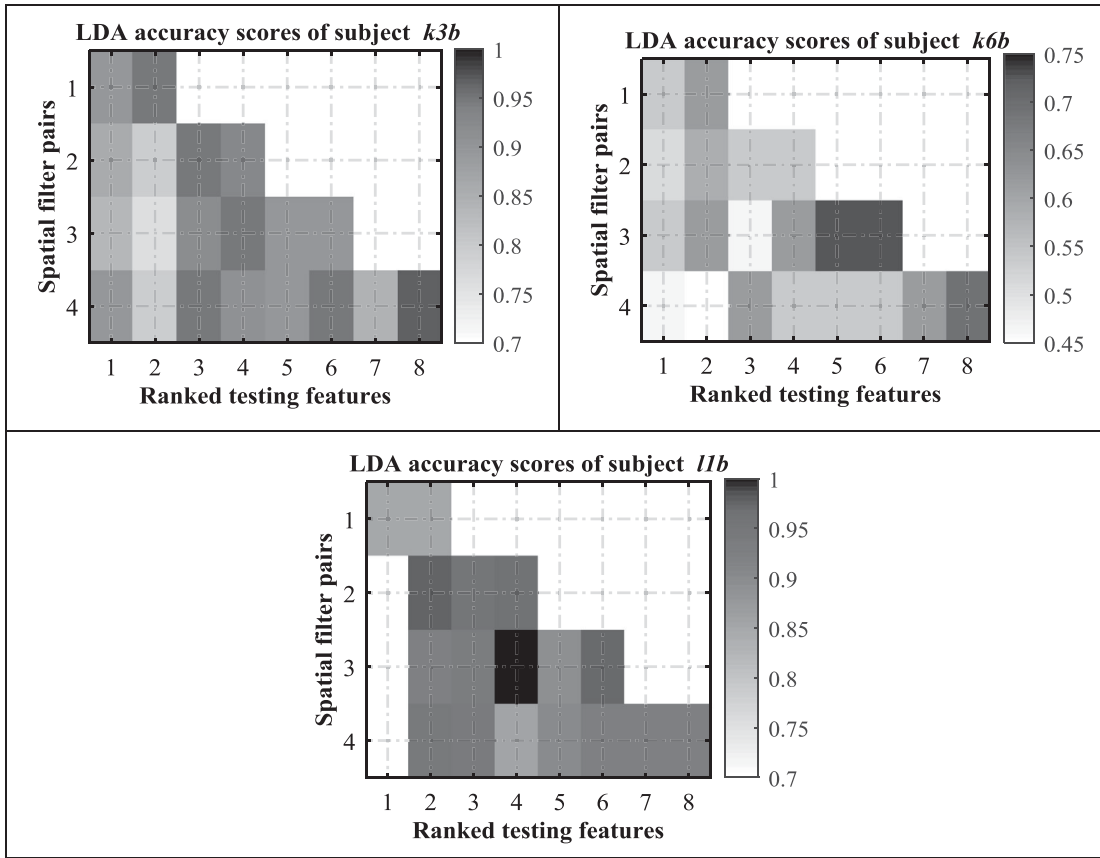


Figure 5: GS-CSP based LDA accuracy scores for all the five subjects of dataset BCIC III-IVa



**Figure 6:** GS-CSP based LDA accuracy scores for all the three subjects of dataset BCIC III-IIIa

channels of BCIC III-IVa and 60 EEG channels of BCIC III-IIIa, from both datasets 18 electrodes around the sensory motor cortex are manually selected for classification as mentioned in [53]. Selected electrodes for the two datasets are shown in Figure 4.

### 6.5 Spatial filter pair and features selection

We employed the proposed state-of-the-art GS-CSP to select the most significant spatial filter pair and discriminative features in different subjects of aforementioned datasets. The training subset comprises 80% trials of the EEG training set and the test subset includes remaining trials. To extract features from both training subset and test subset data, the CSP algorithm with CSP filter pairs  $m$  ( $m = 1, 2, 3$  and  $4$  as it varies from 1 to 4 in most of the existing work) is used in this experiment. For every combination of spatial filter pair and number of features (FP, Fs), a grid of accuracy scores is produced. For the different subject of BCIC III-IVa and BCIC III-IIIa, the GS-CSP based LDA accuracy scores are shown in Figures 5 and 6, respectively. These figures represent the calculated LDA accuracy scores for every value of spatial filter pair with their corresponding discriminant hyperplane coefficients and ranked test subset features.

Based on the maximum score we select the best combination which is subsequently used in classification. In this experiment, the combinations (3, 4), (4, 5), (3, 2), (4, 5), and (2, 4) are selected for the subject *aa*, *al*, *av*, *aw* and *ay*, respectively, as shown in Figure 5. In Figure 6, for the subject *k3b*, *k6b*, and *l1b* the selected combinations are (4, 8), (3, 5), and (3, 4), respectively, in the occasion of left hand versus right hand binary classification. The combination with the smallest number of filter pair is selected if more than one combination produces maximum score since the combination required less training time. In this experiment, the appropriate combinations that are selected for various MI tasks and subjects are tabulated in Table 1. The selected SFPF are then used

**Table 1: Selected combination of spatial filter pairs and number of features (FP, Fs) for various motor imagery task and subject of dataset BCIC III-IIIa**

Motor imagery tasks	Subjects		
	<i>k3b</i>	<i>k6b</i>	<i>l1b</i>
Left hand–Right hand	(4,8)	(3,5)	(3,4)
Left hand–Foot	(3,6)	(2,4)	(1,2)
Left hand–Tongue	(4,7)	(4,8)	(3,4)
Right hand–Foot	(1,2)	(3,4)	(4,6)
Right hand–Tongue	(4,5)	(4,8)	(2,3)
Foot–Tongue	(3,6)	(2,4)	(3,4)

**Table 2: Classification accuracies (mean and standard deviation (SD) in %) of the BCI competition III dataset IVa using the proposed GS-CSP and reported CSP variants. the best result for each subject is displayed in bold characters**

Methods	Spatial filter pairs	BCIC III-IVa					Mean $\pm$ SD
		Subject					
		<i>aa</i>	<i>al</i>	<i>av</i>	<i>aw</i>	<i>ay</i>	
CSP	1	81.89	93.68	70.75	84.71	66.71	79.55 $\pm$ 10.88
	2	82.75	94.29	69.89	87.18	<b>93.46</b>	85.51 $\pm$ 9.93
	3	85.50	96.18	70.04	87.75	90.46	85.99 $\pm$ 9.77
	4	85.95	95.86	70.07	87.43	90.50	85.96 $\pm$ 9.66
CCSP	3	81.29	95.71	72.39	87.75	89.17	85.26 $\pm$ 8.88
SRCSP	3	87.25	94.32	66.71	87.32	89.60	85.04 $\pm$ 10.64
GLRCSP	3	80.43	95.46	69.92	72.82	87.85	81.29 $\pm$ 10.55
TRCSP	3	87.00	95.25	65.53	87.75	92.78	85.66 $\pm$ 11.77
<b>GS-CSP</b>	1, 2, 3, 4	<b>87.96</b>	<b>97.78</b>	<b>72.54</b>	<b>87.95</b>	<b>93.46</b>	<b>87.94 <math>\pm</math> 9.55</b>

**Table 3: Classification accuracies (mean and standard deviation (SD) in %) of the BCI competition III dataset IIIa using the proposed GS-CSP and reported CSP variants**

Methods	Spatial filter pairs	BCIC III-IIIa (Left hand-Right hand)				BCIC III-IIIa (Left hand-Foot)			
		Subject				Subject			
		<i>k3b</i>	<i>k6b</i>	<i>l1b</i>	Mean $\pm$ SD	<i>k3b</i>	<i>k6b</i>	<i>l1b</i>	Mean $\pm$ SD
CSP	1	93.58	73.88	79.21	82.22 $\pm$ 10.19	94.38	80.39	<b>97.37</b>	90.71 $\pm$ 9.06
	2	92.97	76.21	94.29	87.82 $\pm$ 10.08	95.84	<b>84.78</b>	94.88	91.83 $\pm$ <b>6.13</b>
	3	93.67	76.24	97.02	88.98 $\pm$ 11.16	<b>96.65</b>	80.16	94.72	90.51 $\pm$ 9.02
	4	<b>95.30</b>	80.66	96.03	90.66 $\pm$ 8.67	94.36	80.96	94.42	89.91 $\pm$ 7.75
CCSP	3	94.80	80.24	92.97	89.34 $\pm$ <b>7.93</b>	<b>96.65</b>	78.65	94.72	90.01 $\pm$ 9.88
SRCSP	3	94.11	78.92	97.34	90.12 $\pm$ 9.84	<b>96.65</b>	82.42	93.09	90.72 $\pm$ 7.41
GLRCSP	3	94.56	65.16	94.13	84.62 $\pm$ 16.85	94.65	76.98	95.83	89.15 $\pm$ 10.56
TRCSP	3	93.86	80.34	95.41	89.87 $\pm$ 8.29	96.65	82.43	92.90	90.66 $\pm$ 7.37
<b>GS-CSP</b>	1, 2, 3, 4	<b>95.30</b>	<b>81.69</b>	<b>98.55</b>	<b>91.85 <math>\pm</math> 8.94</b>	<b>96.65</b>	<b>84.78</b>	<b>97.37</b>	<b>92.93 <math>\pm</math> 7.07</b>
Methods	Spatial filter pairs	BCIC III-IIIa (Left hand-Tongue)				BCIC III-IIIa (Right hand-Foot)			
		Subject				Subject			
		<i>k3b</i>	<i>k6b</i>	<i>l1b</i>	Mean $\pm$ SD	<i>k3b</i>	<i>k6b</i>	<i>l1b</i>	Mean $\pm$ SD
CSP	1	97.79	87.17	95.13	93.36 $\pm$ 5.53	<b>96.01</b>	75.89	93.91	88.60 $\pm$ 11.06
	2	98.17	87.43	95.51	93.70 $\pm$ 5.59	95.01	80.29	94.06	89.78 $\pm$ 8.24
	3	96.46	88.50	95.25	93.40 $\pm$ 4.28	94.01	87.86	93.78	91.88 $\pm$ 3.48
	4	96.83	<b>90.35</b>	94.64	93.94 $\pm$ <b>3.29</b>	94.01	87.29	96.59	92.63 $\pm$ 4.80
CCSP	3	96.12	82.43	95.25	91.26 $\pm$ 7.66	94.01	87.86	93.72	91.86 $\pm$ 3.47
SRCSP	3	98.46	88.53	95.25	94.08 $\pm$ 5.07	94.01	78.38	94.10	88.83 $\pm$ 9.05
GLRCSP	3	98.02	88.37	95.25	93.88 $\pm$ 4.97	96.62	75.10	95.88	89.20 $\pm$ 12.22
TRCSP	3	98.46	89.22	95.25	94.31 $\pm$ 4.69	93.12	87.86	94.10	91.69 $\pm$ <b>3.36</b>
<b>GS-CSP</b>	1, 2, 3, 4	<b>98.90</b>	<b>90.35</b>	<b>97.88</b>	<b>95.71 <math>\pm</math> 4.66</b>	<b>96.01</b>	<b>89.34</b>	<b>96.67</b>	<b>94.01 <math>\pm</math> 4.05</b>
Methods	Spatial filter pairs	BCIC III-IIIa (Right hand-Tongue)				BCIC III-IIIa (Foot-Tongue)			
		Subject				Subject			
		<i>k3b</i>	<i>k6b</i>	<i>l1b</i>	Mean $\pm$ SD	<i>k3b</i>	<i>k6b</i>	<i>l1b</i>	Mean $\pm$ SD
CSP	1	96.31	83.43	94.53	91.42 $\pm$ 6.98	86.16	90.31	73.57	83.35 $\pm$ <b>8.72</b>
	2	96.20	86.62	94.97	92.60 $\pm$ 5.21	86.55	<b>96.94</b>	75.17	86.22 $\pm$ 10.89
	3	96.20	85.31	94.81	92.10 $\pm$ 5.93	<b>93.79</b>	94.32	75.79	87.63 $\pm$ 10.29
	4	96.73	<b>89.39</b>	94.70	93.60 $\pm$ 3.79	91.47	94.23	73.25	86.32 $\pm$ 11.40
CCSP	3	98.20	89.30	91.72	93.07 $\pm$ 4.60	89.79	93.32	73.82	85.64 $\pm$ 10.39
SRCSP	3	97.20	88.66	94.06	93.31 $\pm$ 4.32	90.64	96.32	74.18	87.05 $\pm$ 11.50
GLRCSP	3	96.20	84.04	91.88	90.71 $\pm$ 6.16	89.79	93.32	75.77	86.29 $\pm$ 9.28
TRCSP	3	96.20	89.20	93.99	93.13 $\pm$ <b>3.58</b>	90.75	96.32	75.41	87.49 $\pm$ 10.83
<b>GS-CSP</b>	1, 2, 3, 4	<b>98.90</b>	<b>89.39</b>	<b>96.93</b>	<b>95.07 <math>\pm</math> 5.01</b>	<b>93.79</b>	<b>96.94</b>	<b>78.94</b>	<b>89.89 <math>\pm</math> 9.61</b>

Note: The best result for each subject is displayed in bold characters

for the classification of test set EEG of respective subjects. Figure 3 shows an example about separation of classes by using CSP, CCSP, SRCSP, GLRCSP, TRCSP, and GS-CSP. To scatter plot the features, spatial filter pair 3 ( $m = 3$ ) and the highest and the lowest ranked features

are used. In the case of GS-CSP the selected (3, 2) combination,  $FP = 3$  and  $Fs = 2$ , is considered. As shown in Figure 3, with the proposed approach the extracted features are properly organized compared to other methods.

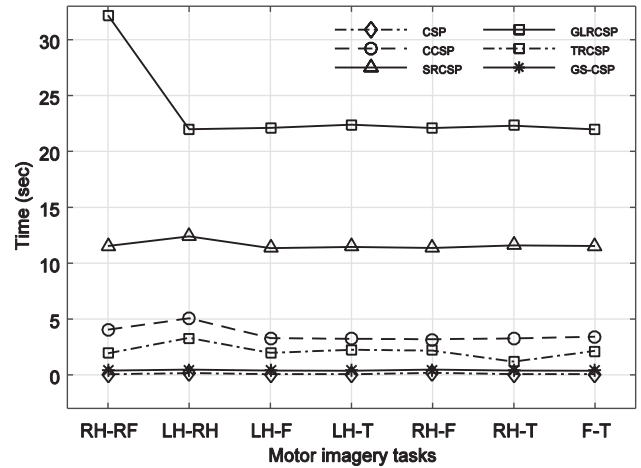
## 6.6 Classification

For each subject, the classification performance is evaluated, using the selected spatial filter pair and a number of ranked CSP features. The selected spatial filter pair is used during features extraction. Subsequently, only the selected number of mutual information-based ranked features are considered to test classification accuracy. The binary classification (right-hand versus right-foot) performances of datasets *BCIC III-IVa* is presented in Table 2. Table 3 demonstrates the result of the six types of MI binary classification (left-hand versus right-hand, left-hand versus foot, left-hand versus tongue, right-hand versus foot, right-hand versus tongue, and foot versus tongue) of dataset *BCIC III-IIIa*. Although there is some variability in classification performance over subjects, the Tables show that the proposed method clearly outperforms the other reported methods. Besides, it outperforms them as it gives less execution time. To estimate the performance a validation technique is used in this experiment. In all our experiments 10-fold cross validation is used. In this cross validation technique, the dataset is randomly divided into 10 equal subsets where one of the subsets is used for the test while rests 9 are used to the training. The cross validation is repeated 10 times, and then the results of 10 times are averaged to yield a single classification rate. The execution time is estimated by summing up the expended time during spatial filtering (training) and classification (testing). To calculate the execution time the 10-fold cross validation with the technical software MATLAB R2014b is used. During the performance estimation the system environment was Intel(R) Core(TM) i5-4310U @ 2.00 GHz processor, 4.00 GB RAM and 64-bit Windows 7 Ultimate operating system. Figure 7 shows the average execution time of different subject for various MI tasks.

## 7. DISCUSSION AND CONCLUSIONS

Before classification, it needs to reduce dimensionality of original EEG data so that the redundant information from the data can be removed. The techniques, PCA and ICA can be used to reduce dimensionality in BCI applications. In the research in [18,22,24], very promising classification results are achieved by selecting features using the PCA. The features can be reduced to low dimensional data using both PCA and ICA [23]. In this study, features are extracted by CSP and the dominant CSP features from the extracted CSP features are selected.

A novel method to classify EEG of imagined movement is presented in this paper. The EEG is filtered into the most dominant frequency bands alpha and beta (8–30 Hz).



**Figure 7:** Average execution time of different subject for various motor imagery tasks; RH-RF: Right Hand versus Right Foot, LH-RH: Left Hand versus Right Hand, LH-F: Left Hand versus Foot, LH-T: Left Hand versus Tongue, RH-F: Right Hand versus Foot, RH-T: Right Hand versus Tongue, and F-F: Foot versus Tongue

The training and test sets of the EEG are separated and GS-CSP is applied to only the training set in order to select spatial filter pair and discriminative number of CSP features. In doing so, the training set is further divided into two sets; training subset and test subset. The discriminative CSP features are then extracted from both parts. Mutual information concerning the training subset CSP features and their corresponding given class are computed. Later on, features are put in order based on mutual information. Subsequently, the CSP features of the test subset are ranked in accordance with the order of the ranked training subset CSP features. The discriminant hyperplane coefficients are computed for all of the ranked training CSP features. These coefficients are used to calculate the LDA accuracy scores for the respective testing CSP features. The number of spatial filter pair doubles the CSP features and influences classification accuracy. In most of the existing BCI study a certain pair of spatial filter is selected manually, however, that does not assure the best accuracy result. Spatial filter pair selection differs from study to study mostly from 1 to 4. The highest result can be achieved by appropriate selection of spatial filter pair and number of CSP features. In this study, all of the spatial filter pairs of 1–4 and their corresponding CSP features are considered. The LDA accuracy scores are calculated for all of the combinations of spatial filter pairs and CSP features, (FP,Fs) where  $FP = 1, \dots, 4$  and  $Fs = 1, \dots, 2m$ . Based on the grid search the top accuracy score is obtained and hence corresponding combination is selected for classification. The accuracy scores are subject to the combination e.g. the combination of (3, 4), (4, 5), (3, 2), (4, 5), and

(2, 4) that yields top scores for the subjects *aa*, *al*, *av*, *aw* and *ay*, respectively, as shown in Figure 5. In the case of left hand versus right hand MI task, the best score for the subject *k3b*, and *l1b* are produced for the combination of (4, 8) and (3, 4), respectively, as shown in Figure 6. Two combinations e.g. (3, 5) and (3, 6) generate the highest score for the subject *k6b* but only (3, 5) is selected for classification because of less features. The combination with less filter pair should be selected if more than one combination with different filter pair produces the highest score since the pair gives a reduced amount of training time. Finally, the selected combination is used to compute the classification accuracy of the test set EEG.

It is concluded that on average even for individual subject the maximum classification accuracy is achieved by the proposed method. Moreover, in the case of execution time the proposed method outperforms some other methods as shown in Figure 7. The execution time of the methods CCSP, SRCSP, GLRCSP, and TRCSP is more than one second because of more training time. As described in the subsections 3.2–3.5, the algorithm uses data from other subjects and regularizes the estimation of the covariance matrix. Since the methods consider the information of other subject during the training period, the algorithms consume much more training time than standard CSP. The proposed feature selection approach GS-CSP searches the best feature combination using the standard CSP. The approach needs little more time compared to CSP for the searching but much lower than the other methods as shown in Figure 7.

Therefore, experimental studies on two public EEG datasets (BCI competition III dataset IVa and BCI competition III dataset IIIa) showed that the proposed GS-CSP method yielded higher overall classification accuracy and lower execution time in contrast to some other stated methods.

## ACKNOWLEDGEMENTS

The authors Md. Sujun Ali and Mst. Jannatul Ferdous are funded by the ICT Division of the Ministry of Post, Telecommunication and Information Technology of the Government of Bangladesh through a fellowship program. Therefore, the authors would like to acknowledge the State Ministry of Bangladesh.

## ORCID

Jannatul Ferdous  <http://orcid.org/0000-0002-6233-3703>

## REFERENCES

1. S. C. Kleih, T. Kaufmann, C. Zickler, S. Halder, F. Leotta, F. Cincotti, F. Aloise, *et al.*, “Out of the frying pan into the fire – the P300-based BCI faces real-world challenges,” *Prog. Brain Res.*, Elsevier, Vol. 194, pp. 27–46, 2011.
2. D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, “Electroencephalographic (EEG) control of three-dimensional movement,” *J. Neural Eng.*, Vol. 7, no. 3, pp. 036007, 2010.
3. F. Cincotti, D. Mattia, F. Aloise, S. Bufalari, G. Schalk, G. Oriolo, A. Cherubini, M. G. Marciani, and F. Babiloni, “Non-invasive brain–computer interface system: towards its application as assistive technology,” *Brain Res. Bull.*, Vol. 75, no. 6, pp. 796–803, 2008.
4. S. Janjarasjitt, “Epileptic seizure classifications of single-channel scalp EEG data using wavelet-based features and SVM,” *Med. Biol. Eng. Comput.*, Vol. 55, no. 10, pp. 1743–61, 2017.
5. W. Mumtaz, S. S. A. Ali, M. A. M. Yasin, and A. S. Malik, “A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD),” *Med. Biol. Eng. Comput.*, Vol. 56, no. 2, pp. 233–46, 2018.
6. K. Samiee, P. Kovács, and M. Gabbouj, “Epileptic seizure detection in long-term EEG records using sparse rational decomposition and local Gabor binary patterns feature extraction,” *Knowl. Based. Syst.*, Vol. 118, pp. 228–40, 2017.
7. F. Akram, M. K. Metwally, H.-S. Han, H.-J. Jeon, and T.-S. Kim, “A novel P300-based BCI system for words typing,” in 2013 International Winter Workshop on Brain-Computer Interface (BCI), IEEE, 2013, pp. 24–5.
8. D. La Rocca, P. Campisi, and J. Solé-Casals, “EEG based user recognition using BUMP modelling,” in 2013 International Conference of the BIOSIG special interest Group (BIOSIG), IEEE, 2013, pp. 1–12.
9. B. Rebsamen, E. Burdet, C. Guan, H. Zhang, C. L. Teo, Q. Zeng, C. Laugier, and M. H. Ang Jr., “Controlling a wheelchair indoors using thought,” *IEEE Intell. Syst.*, Vol. 22, no. 2, pp. 18–24, 2007.
10. N. Birbaumer, “Brain-computer-interface research: Coming of age,” *Clin. Neurophysiol.*, Vol. 117, no. 3, pp. 479–83, 2006.
11. S. Kumar, K. Mamun, and A. Sharma, “CSP-TSM: Optimizing the performance of Riemannian tangent space mapping using common spatial pattern for MI-BCI,” *Comput. Biol. Med.*, Vol. 91, pp. 231–42, 2017.
12. S. Kumar, A. Sharma, K. Mamun, and T. Tsunoda, “A deep learning approach for motor imagery EEG signal classification,” in 2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), IEEE, 2016, pp. 34–9.

13. S. Kumar, A. Sharma, and T. Tsunoda, "An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information," *BMC Bioinformatics*, Vol. 18, no. 16, pp. 545, 2017.
14. M. Miao, A. Wang, and F. Liu, "A spatial-frequency-temporal optimized feature sparse representation-based classification method for motor imagery eeg pattern recognition," *Med. Biol. Eng. Comput.*, Vol. 55, no. 9, pp. 1589–1603, 2017.
15. A. Yuksel, T. Olmez, and L. Bianchi, "A neural network-based optimal spatial filter design method for motor imagery classification," *PloS One*, Vol. 10, no. 5, pp. e0125039, 2015.
16. J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, Vol. 113, pp. 767–91, 2002.
17. E. Niredermeyer, and F. Lopes da Silva, "Electroencephalography: BasicPrinciples", in *Clinical Applications and Related Fields*, 5th ed., Baltimore, MD, USA: Lippincott Williams Wilkins, 2005, pp. 1–1256.
18. K. Gorur, M. R. Bozkurt, M. S. Bascil, and F. Temurtas, "Glossokinetic potential based tongue-machine interface for 1-D extraction," *Australas. Phys. Eng. Sci. Med.*, Vol. 41, no. 2, pp. 379–91, 2018.
19. G. Pfurtscheller, and F. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin. Neurophysiol.*, Vol. 110, no. 11, pp. 1842–57, 1999.
20. A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian based kernel for BCI applications," *Neurocomputing*, Vol. 112, pp. 172–78, 2013.
21. H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabilitation Eng.*, Vol. 8, no. 4, pp. 441–46, 2000.
22. M. S. Bascil, A. Y. Tesneli, and F. Temurtas, "Multi-channel EEG signal feature extraction and pattern recognition on horizontal mental imagination task of 1-D cursor movement for brain computer interface," *Australas. Phys. Eng. Sci. Med.*, Vol. 38, no. 2, pp. 229–39, 2015.
23. M. S. Bascil, A. Y. Tesneli, and F. Temurtas, "Spectral feature extraction of EEG signals and pattern recognition during mental tasks of 2-D cursor movements for BCI using SVM and ANN," *Australas. Phys. Eng. Sci. Med.*, Vol. 39, no. 3, pp. 665–76, 2016.
24. M. S. Bascil, "A new approach on HCI extracting conscious jaw movements based on EEG signals using machine learnings," *J. Med. Syst.*, Vol. 42, no. 9, pp. 169, 2018.
25. B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process Mag.*, Vol. 25, no. 1, pp. 41–56, 2008.
26. B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, K. R. Muller, and V. V. Nikulin, "Invariant common spatial patterns: Alleviating non stationarities in brain-computer interfacing," *Adv. Neural. Inf. Process. Syst.*, Vol. 20, pp. 113–20, 2008.
27. G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K. R. Muller, "Optimizing spatio-temporal filters for improving brain-computer interfacing," *Adv. Neural. Inf. Process. Syst.*, J. Platt, Ed., Vol. 18, pp. 315–22, (2006).
28. J. Muller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, Vol. 110, no. 5, pp. 787–98, 1999.
29. F. Lotte, and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, Vol. 58, no. 2, pp. 355–62, 2011.
30. F. Yger, F. Lotte, and M. Sugiyama, "Averaging covariance matrices for EEG signal classification based on the CSP: an empirical study," in 23rd European Signal Processing Conference (EUSIPCO), 2015, pp. 2771–75.
31. Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface," *J. Neurosci. Methods*, Vol. 255, pp. 85–91, 2015.
32. H. Suk, and S.-W. Lee, "A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 35, no. 2, pp. 286–99, 2013.
33. K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in IEEE International Joint Conference on Neural Networks, 2008, pp. 2390–97.
34. H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Process Lett.*, Vol. 16, no. 8, pp. 683–6, 2009.
35. O. Falzon, K. P. Camilleri, and J. Muscat, "The analytic common spatial patterns method for EEG-based BCI data," *J. Neural Eng.*, Vol. 9, no. 4, pp. 045009, 2012.
36. B. Blankertz, K. R. Muller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, et al., "The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials," *IEEE Trans. Biomed. Eng.*, Vol. 51, no. 6, pp. 1044–51, 2004.

37. B. Blankertz, K. R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, et al., "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, Vol. 14, no. 2, pp. 153–59, 2006.
38. W. Samek, C. Vidaurre, K. R. Muller, and M. Kawanabe, "Stationary common spatial patterns for brain–computer interfacing," *J. Neural Eng.*, Vol. 9, no. 2, pp. 026013, 2012.
39. D. Fattahi, B. Nasihatkon, and R. Boostani, "A general framework to estimate spatial and spatio-spectral filters for EEG signal classification," *Neurocomputing*, Vol. 119, pp. 165–74, 2013.
40. B. Reuderink, and M. Poel. "Robustness of the common spatial patterns algorithm in the BCI-pipeline," 2008.
41. H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial patterns with generic learning for EEG signal classification," in 31st Annual International Conference of the IEEE EMBS, 2009, pp. 6599–602.
42. H. Wang, Q. Tang, and W. Zheng, "L1-norm-based common spatial patterns," *IEEE Trans. Biomed. Eng.*, Vol. 59, no. 3, pp. 653–62, 2012.
43. R. P. N. Rao. *Brain-Computer Interfacing: An Introduction*. New York: Cambridge University Press, 2013.
44. Z. J. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG," *Electroencephalogr. Clin. Neurophysiol.*, Vol. 79, no. 6, pp. 440–7, 1991.
45. Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selection in motor imagery based brain computer interface," in IEEE Engineering in Medicine and Biology, 27th Annual Conference, 2005, pp. 5392–5.
46. Y. Xi, U. Hasson, P. J. Ramadge, and Z. J. Xiang, "Boosting with spatial regularization," *Adv. Neural. Inf. Process. Syst.*, Vol. 1, no. 3, pp. 2107–15, 2009.
47. D. Cai, X. He, Y. Hu, J. Han, and T. Huang, "Learning a spatially smooth subspace for face recognition," in 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007.
48. T. M. Cover, and J. A. Thomas. *Elements of information theory*. 2nd edition. New York: Wiley, 2006, pp. 1–7.
49. M. Kantardzic. *Data mining: concepts, models, methods, and algorithms*. New Jersey: IEEE Press & John Wiley, November 2002.
50. M. R. Hasan, M. I. Ibrahimy, S. M. A. Motakabber, and S. Shahid, "Classification of multichannel EEG signal by linear discriminant analysis," in 23rd International Conference on Systems Engineering, Advances in Intelligent Systems and Computing 1089, 2015, pp. 279–82.
51. L. Song, and J. Epps, "Classifying EEG for brain-computer interface: Learning optimal filters for dynamical system features," *Comput. Intell. Neurosci.*, Vol. 2007, pp. 1–11, 2007.
52. C. Neuper, G. R. Muller-Putz, R. Scherer, and G. Pfurtscheller, "Motor imagery and EEG-based control of spelling devices and neuroprostheses," *Prog. Brain Res.*, Vol. 159, pp. 393–409, 2006.
53. S. Wang, and C. J. James, "Extracting rhythmic brain activity for brain-computer interfacing through constrained independent component analysis," *Comput. Intell. Neurosci.*, Vol. 2007, pp. 1–9, 2007.

## Authors



**Md. Sujan Ali** received BSc and MSc degrees in Electrical and Electronic Engineering from Islamic University, Kushtia, Bangladesh in 2002 and 2003 respectively. During 2006–2008, he was a lecturer in the Department of Information and Communication Technology, Metropolitan University, Bangladesh. After that, he joined as a Lecturer in the Department of Computer Science and Engineering, Jatiya Kabi Kazi Nazrul Islam University, Mymensingh, Bangladesh where he promoted as an Assistant Professor in 2011 and Associate Professor in 2016. He received his PhD degree from the faculty of Engineering, University of Rajshahi, Bangladesh in 2019. His research interest includes brain signal processing, brain–computer interface (BCI), biomedical signal and image processing.

**Email:** sujan\_cse@jkkniu.edu.bd



**Mst. Jannatul Ferdous** achieved her BSc and MSc degree in Electrical and Electronic Engineering from Islamic University, Kushtia, Bangladesh in 2002 and 2003 respectively. She was a Lecturer in the Department of Information and Communication Technology, Metropolitan University, Bangladesh. In 2010, she joined as a Lecturer and promoted as an Assistant Professor in 2011 and Associate Professor in 2016 in the Department of Computer Science and Engineering, Jatiya Kabi Kazi Nazrul Islam University, Mymensingh, Bangladesh. She received her PhD degree from the faculty of Engineering, University of Rajshahi, Bangladesh in 2019. Her research interest includes digital signal processing, EEG signal enhancement, biomedical signal and image processing.

**Corresponding author. Email:** mjferdous\_cse@jkkniu.edu.bd



**Md. Ekramul Hamid** received his BSc and MSc degree from the Department of Applied Physics and Electronics, Rajshahi University, Bangladesh. After that he obtained the Masters of Computer Science from Pune University, India. He received his PhD degree from Shizuoka University, Japan. During 1997–2000, he was a lecturer in the Department of Computer Science and Engineering, Rajshahi University. Dr Hamid was working as Assistant Professor at the King Khalid University, Abha, KSA from 2009 to 2011. He is currently working as a Professor in the Department of Computer Science and Engineering, University of Rajshahi. His research interests include Digital Signal Processing, Analysis and Synthesis of Speech Signal, Speech Enhancement, and Image Processing.

**Email:** ekram\_hamid@yahoo.com



**Md. Khademul Islam Molla** received his BSc and MSc degrees in electronics and computer science from Shahjalal University of Science and Technology, Sylhet, Bangladesh in 1995 and 1997, respectively. He joined at the same department as a Lecturer in 1997. He obtained his PhD degree from the Department of Frontier Informatics under the Graduate School of Frontier Sciences, the University of Tokyo, Tokyo, Japan in 2006. He was working as Lecturer and Assistant Professor in the Department of Computer Science and Engineering of the University of Rajshahi, Bangladesh up to August 2006. After completing his PhD, he joined in the same department as Associate Professor in August 2006, and he has been a Professor since May 2012. From September 2006 to September 2008, he was working as JSPS Postdoctoral Research Fellow in the Department of Information and Communication Engineering, The University of Tokyo, Tokyo, Japan. He was a research fellow at the University of Alberta, Canada, from Nov. 2010 to Oct. 2011. He visited several universities in Japan as guest researcher. Presently, he is a visiting scientist at the University of Tokyo, Japan. His research interests include audio signal processing, blind source separation, brain–computer interface (BCI), biomedical signal, and image processing. He is a member of the Institute of Electrical and Electronics Engineers (IEEE). In 2007, he received the Best Paper Award from the Research Institute of Signal Processing, Japan (RISP).

**Email:** khademul.cse@ru.ac.bd