

Received April 28, 2020, accepted May 11, 2020, date of publication May 22, 2020, date of current version June 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2996685

Discriminative Feature Selection-Based Motor Imagery Classification Using EEG Signal

MD. KHADEMUL ISLAM MOLLA¹, (Member, IEEE), ABDULLAH AL SHIAM²,
MD. RABIUL ISLAM³, AND TOSHIHISA TANAKA^{3,4,5}, (Senior Member, IEEE)

¹Department of Computer Science and Engineering, University of Rajshahi, Rajshahi 6205, Bangladesh

²Department of Computer Science and Engineering, Varendra University, Rajshahi 6204, Bangladesh

³Institute of Global Innovation and Research, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan

⁴Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan

⁵RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan

Corresponding authors: Md. Khademul Islam Molla (khademul.cse@ru.ac.bd) and Toshihisa Tanaka (tanakat@cc.tuat.ac.jp)

This work was supported in part by the Japan Society for Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (KAKENHI) Grant Numbers JP17H01760 and JP20H00235.

ABSTRACT Achieving a reliable classification of motor imagery (MI) tasks is a major challenge in brain-computer interface (BCI) implementation. The set of relevant and discriminative features plays an important role in the classification scheme. This paper presents a supervised approach to select discriminative features for the enhancement of MI classification using multichannel electroencephalography (EEG) signal. The dimension of multiband feature space is reduced using the feature selection method. Each trial of the multichannel EEG signal representing MI tasks is decomposed into a finite set of narrowband signals. The common spatial pattern-based features are extracted from each subband. The features obtained from the multiple subbands are combined to derive a high-dimensional feature vector. The neighborhood component analysis-based feature selection method is implemented to select the features that are relevant in performing an accurate classification. It is a nearest-neighbor-based approach to learn the feature weights with regularization by maximizing the average leave-one-out classification accuracy over the labeled training data. The selected features are used to train the support vector machine for classification. The features relatively irrelevant to the classification task are discarded, yielding a reduction of feature dimension. The evaluation of the proposed method is performed using BCI Competition III dataset 4a and IV dataset 2b. Both are publicly available datasets and are used as types of benchmark data to evaluate the MI classification algorithm to implement BCI. The obtained simulation results confirm the superiority of the proposed method compared to the recently developed algorithms.

INDEX TERMS Brain-computer interface (BCI), electroencephalography (EEG), machine learning, motor imagery (MI), subband decomposition, supervised feature selection.

I. INTRODUCTION

A Brain-Computer interface (BCI) decodes the movement imagination, also called the motor imagery (MI) of the brain, to issue a command without any peripheral nerve or muscle activity [1]. It has potential applications in neuroscience and neuro-engineering. MI has been used to encourage neuroplasticity in a patient's brain after a stroke [2]. Thus, recent applications of BCI with appropriate feedback offer neurorehabilitation to assist stroke patients in restoring their impaired motor functions [3], [4]. The usage of prosthetics, robots, and

other electronic devices used in neurorehabilitation tasks is fully controlled by motor imagination [1], [5]. The kinesthetic, auditory, or visual feedback to the subject is used to stimulate the response of the brain after a stroke. The use of non-invasive electroencephalography (EEG) is a comfortable and relatively easy method for BCI implementation. The BCI user's brain activity is typically measured using EEG [6]. The EEG-based design of BCI application is an extremely challenging task [7]. There are two types of MI-based BCI: asynchronous and synchronous. The subject controls the task and its timing without any external cues in asynchronous BCI [8]. It is more appropriate to implement real-time BCI applications but such BCI system requires processing of the

The associate editor coordinating the review of this manuscript and approving it for publication was Juan Wang¹.

brain signals continuously. In synchronous BCI, the cue is provided for a fixed time duration in which the subject needs to perform a mental task [9]. In this paper, the experimental setup is confined to implementing synchronous BCI.

The principal goal of BCI is to recognize the correct intention of brain activity during MI tasks, leading to a translation of the intention into an equivalent control command. To achieve this goal, suitable features are extracted and the appropriate machine learning scheme is employed to perform the classification of MI tasks recorded by multichannel EEG. Thus, discriminative feature extraction from EEG signal is one of the key stages in designing an MI-based BCI application. The common spatial pattern (CSP) is one of the methods used to extract potential features from multichannel EEG [10]. It implements optimal spatial filter from the training samples of recorded EEG signals. CSP derives a weighting matrix for the electrodes based on their significance in the classification task. Later, the performance of the MI-based BCI (MI-BCI) is improved with CSP by accomplishing different physiological events, such as event-related desynchronization and lateralized readiness potential [11]. Instead of using only spatial pattern, the common patterns of the spatial, time, and frequency domains are jointly considered to extract the potential features from EEG [12]. The features of the spatio-temporal discrepancy signal are derived from the EEG to classify two classes of MI tasks [13]. The deep learning approach with variational autoencoder is employed in [14] for EEG classification. The images achieved by time-frequency representation of EEG signals are used as the input to the deep neural network. The features obtained by regularized Riemannian transformation are used to classify MI tasks with reduced calibration time of EEG [15]. The narrowband signals containing the significant information about movement imagination enhances the MI classification performance. None of the methods [12]–[15] use subband filtering to extract the signal components representing the MI tasks.

The performance of MI-BCI significantly depends on the selection of frequency bands of the EEG signal from which the features are extracted [7]. The *mu* and *beta* rhythmic components evoked in response to the imagination of different movement tasks are promising sources of the features used in MI-BCI [16]. These rhythms are observed in the area of sensorimotor. The movement imagination of the hand and the foot produces the change in the *mu* rhythm at different brain regions [11]. It is also observed in [17] that the effective fluctuations in brain activity occur in low-frequency components of EEG. Moreover, the changes in *mu* and *beta* rhythms occur due to the subject's voluntary movements. To capture the changes within the narrow frequency band, CSP is implemented in each subband of EEG signals for discriminative feature extraction [18]. The selection of relevant features plays an important role in MI classification for BCI implementation. The feature selection approach is not implemented in previous studies [11], [16]–[18].

In MI-BCI studies, a number of methods including filter bank CSP (FBCSP) [19], [20], subband CSP [21], sparse filter-bank CSP [22], and discriminative filter bank CSP [23] have been proposed to extract the features from the narrowband EEG signals for MI classification. The sparse representation of CSP feature is implemented in [24] for two classes of MI discrimination. These works promote the implementation of subband CSP to obtain the discriminative features, thereby yielding a reliable classification of MI tasks. Therefore, the subband approach with CSP method is implemented in this work to extract the effective features from the narrowband EEG. In addition to the narrowband signals, the wideband signal can contain some apparent features to enhance MI classification. Without considering this issue, the narrowband signals are used for feature extraction in the subband CSP-based methods.

In the subband-based method, feature extraction is performed in individual subbands and combined to yield the feature vector. Thus, the derived feature space has a relatively higher dimension [25]. A number of features included in the feature vector might not be accurate, which can garble the machine learning algorithm and lead to the degradation of its performance [26]. An appropriate subgroup of the obtained features should be selected to perform a reliable classification of MI tasks. The others need to be removed to reduce the degradation of performance and reduce the overfitting and training time of the machine learning algorithm.

There are two types of feature selection techniques: unsupervised and supervised. Unsupervised algorithm selects relevant features without using the label information of the dataset available for training [25], whereas the supervised method requires the proper labels of training data for discriminative feature selection. Usually, it quantifies the distance between the features and the labelled training set by measuring mutual information, correlation, etc. [26]. The supervised technique is effective when the label information is available for training data. The label information is available with the data used in the experiments of this research and, thus, the supervised feature selection algorithm is implemented.

In this paper, subband CSP features are used and neighborhood component analysis (NCA)-based [27] supervised feature selection method is introduced to separate the highly discriminative features to enhance the performance of MI task classification. The multichannel EEG signal is recorded to represent the mental imagination in terms of MI task. The multichannel EEG is passed through a subband decomposition scheme to generate a finite set of narrowband signals. It localizes the signal components that are effective for MI classification within the subbands. The spatial features are extracted from each of the subbands by applying CSP. Then, the features obtained from the individual subbands are combined to yield a high-dimensional feature vector. Discriminative feature selection is performed by using NCA. Thus, the selected feature vector is used for the classification of MI tasks with the support vector machine (SVM). The effects of the number of features on classification accuracy

and the performance of the different classifiers are evaluated. The results of the experimental evaluation are compared with those obtained using the recently developed methods.

Regarding the organization of this paper, Section II describes the datasets used in the experiments, Section III details the methodology, Section IV illustrates the experimental results and discussion and, finally, Section V presents the conclusions.

II. DATA DESCRIPTION

Two publicly available BCI competition datasets are used to evaluate the performance of the proposed method. The datasets are described below.

A. BCI COMPETITION III DATASET 4a

The data were obtained from five healthy subjects, denoted as ‘aa’, ‘al’, ‘av’, ‘aw’, and ‘ay’ [11]. The age of each subject was between 24 to 25 years. Proper instructions were provided to the subjects, such that the experimental conditions were fulfilled. They sat on a comfortable chair and avoided eye movement. The visual stimulus was presented for 3.5 s. During that time duration, each participant was asked to perform three MI tasks, i.e., right hand, left hand, and right foot movement. Two MI tasks of the right hand and foot were taken into consideration for classification. A total of 280 trials of EEG with 118 channels were recorded for each subject while they performed the MI tasks according to the instruction. In particular, 168, 224, 84, 56, and 28 trials out of 280 were designed as training data for the subjects ‘aa’, ‘al’, ‘av’, ‘aw’, and ‘ay’, respectively. The other trials were kept for testing. The training data available with proper labels for the individual subjects are used in this study to evaluate the performances. The recorded signals are filtered using a band-pass filter between 0.05–200 Hz, sampled at 1000 Hz, and quantized by 16-bit resolution. The EEG signal is down-sampled at 100 Hz for further processing. The details of the experimental setup are provided in the study [11]. In this study, the 2 s length (0.5–2.5 s) EEG trial is extracted to obtain the meaningful feature that will be used in the classification. It is considered that the first 0.5 s (0–0.5 s) and the last 0.5 s (3.5–4.0 s) are the durations for pre- and post-imagination, respectively.

Channel Selection: A total of 118 EEG channels are used to record the mentioned dataset, including some irrelevant signals, because all of the channels are not required to discriminate the two MI tasks. The selection of relevant and minimum number of channels will be effective in terms of computational cost. The relevant zone of motor activity is the motor cortex region, including the primary, supplementary, and premotor cortex area [28]. The electrodes placed at these areas should be selected. Several studies are performed to use a selected number of channels to design the MI-BCI [29]–[31]. The 18 channels from the area of the sensorimotor cortex are used in [29], while 30 channels are selected in [30], [31] to classify two MI tasks. Considering the previous studies [30], [31], 30 channels in the sensory-motor cortex

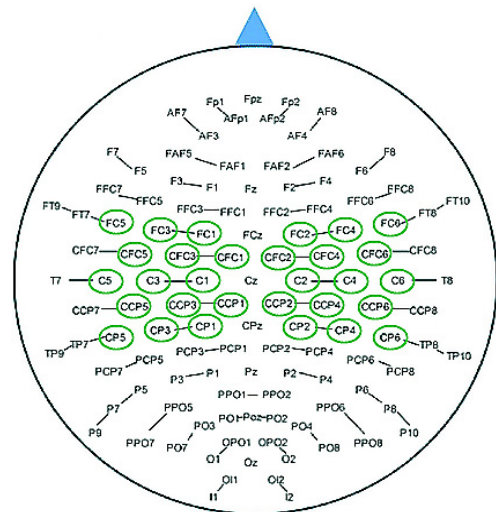


FIGURE 1. Thirty electrodes encircled in green are selected out of 118 and used in this study for experimental evaluation.

area are selected for MI task classification, as indicated in Fig. 1. Throughout this paper, the multichannel EEG refers to the signals recorded from the 30 selected electrodes.

B. BCI COMPETITION IV DATASET 2b

The EEG signals were recorded with three channels (C3, Cz, and C4) sampled at 250 Hz and band-pass filtered in the frequency range of 0.5–100 Hz. Nine subjects, namely, B01, B02, . . . , B09 participated in an experiment on MI by performing two different tasks (left hand and right hand). Five sessions were recorded for each subject. The first two sessions and the third session were designed for training without visual feedback and with visual feedback, respectively. The last sessions (4th and 5th) were recorded for testing. The EEG data of the three training sessions were used in this study to evaluate its performance. The 2 s trial length (0.5–2.5 s after starting the stimuli) is extracted to conduct the experiments. More details about the BCI Competition IV (2b) dataset could be found in [32].

III. METHODOLOGY

A multiband approach for dominant feature selection method is implemented here to enhance the classification accuracy of MI tasks for BCI application. A block diagram of the proposed method is shown in Fig. 2.

Subband CSP features are extracted from the EEG signals. A selected subset of features is used to classify the motor imageries with the SVM [33] classifier. Below are the steps to implement the method:

- (i) The multichannel EEG signal is decomposed into subbands
- (ii) CSP is used for feature extraction from each subband
- (iii) The features obtained from the individual subband are combined to derive a feature vector
- (iv) Discriminative features are selected using neighborhood component analysis based feature selection (NCFS)

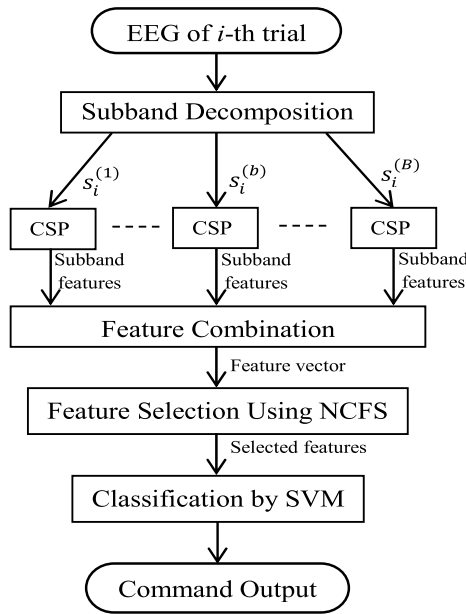


FIGURE 2. Block diagram of the proposed feature selection-based MI classification system, where $s_i^{(1)}, \dots, s_i^{(b)}, \dots, s_i^{(B)}$ represent the different subbands.

- (v) The SVM classifier is trained with selected features of the labeled training dataset and classification is performed using the test dataset
- (vi) Finally, the command is generated on the basis of MI classification for BCI implementation

A. SUBBAND DECOMPOSITION

The multichannel raw EEG is often contaminated by electro-physiological noise. Sometimes, such noise power is stronger than that of the EEG signal. Moreover, some narrowband components of EEG signal have a stronger response to the specific MI task. Therefore, the proper selection of subbands would intuitively provide a more accurate classification of MI tasks than using the full bandwidth of the EEG. Related studies claim that most of the brain activities related to MI tasks exist within the frequency band of 7–30 Hz [34]. Based on the experiments, four subbands are used within the frequency range of 8–35 Hz in this study. The subband decomposition is accomplished by applying Butterworth zero-phase band pass filter. The full band (8–35 Hz), mu band (8–13 Hz), low beta (13–22 Hz), and high beta with low gamma (22–35 Hz) are used here as the usable narrowband signals. The CSP-based features are extracted from each subband and combined to construct a feature vector with high dimensionality.

B. FEATURE EXTRACTION

The extraction of potential features is one of the most crucial stages in the field of BCI. Recent studies have generally investigated how to modify existing methods or develop novel techniques for feature extraction because of the features’ direct influence on the performance of the BCI system [1], [6]. One of the most successful and well-known methods in

BCI application to extract features from multichannel EEG is CSP [15], [16]. It decomposes a multichannel EEG into a number of additive components. Basically, it is a linear transformation to project a high-dimensional EEG signal into a low-dimensional spatial subspace with a projection matrix. Any row of the projection matrix consists of the weights of the EEG channels. Such transformation maximizes the variance of two-class signal matrices. It performs the simultaneous diagonalization of the covariance matrices derived from both of the classes [35]. The spatial filter is designed such that the variance of filtered data from one class is maximized while that of the other class is minimized. The resultant features minimize the intra-class variance while maximizing the inter-class variance. It increases the separation between the two classes in terms of variance [34]. Such attribute of CSP makes it an effective spatial filter to classify MI tasks using multichannel EEG classification. The first CSP-based spatial filter was implemented in [36] to effectively classify movement-related EEG for BCI implementation.

Let $E_{i,1}$ and $E_{i,2} \in \mathfrak{R}^{K \times L}$ denote the EEG training trials selected from the two different classes with dimensions $K \times L$, where K represents the number of channels and L is the number of discrete samples. CSP method derives the features based on the simultaneous diagonalization of the covariance matrices of both classes. It finds a spatial filter $w \in \mathfrak{R}^K$ to transform the EEG data with a projection matrix, such that the ratio of variance between the two classes becomes maximized [35].

$$w = \arg \max_w \frac{w^T \Lambda_1 w}{w^T \Lambda_2 w} \quad s.t. \|w\|_2 = 1 \quad (1)$$

where $\Lambda_c = \sum_{i=1}^{Y_c} E_{1,c} E_{i,c}^T / Y_c$ and Y_c is the number of trials belonging to class c ($c = 1, 2$). The optimal solution of Eq. (1) can be obtained by solving a generalized eigenvalue problem.

A matrix $w = [w_1, w_2, \dots, w_{2M}] \in \mathfrak{R}^{K \times 2M}$ including the spatial filters is formed by the eigenvectors corresponding to the M largest and smallest eigenvalues. For a given EEG sample E , the feature vector is constructed as $x = [x_1, x_2, \dots, x_{2M}]$ with entries [35]

$$x_m = \log \left(\text{var}(w_m^T E) \right), \quad m = 1, 2, \dots, 2M \quad (2)$$

where $\text{var}(\cdot)$ represents the variance. Log transformation is done in order to normalize the elements of x_m . A selected number of features are extracted from each subband using CSP. All the features obtained from the subbands are combined (simple concatenation) to generate the feature vector for the respective trial.

C. FEATURE SELECTION

Not all of the features in a high-dimensional feature vector are effective for classification. In fact, some features often degrade the performance of the machine learning algorithm. The aim of feature extraction is to provide appropriate discriminative information to enhance the object classification

performance. Hence, in machine learning approach, feature selection is an important part for choosing the best set of features from all that are available. Another objective of feature selection method is to suppress the irrelevant features with minimization of information loss.

The discriminative features are selected in this study from the raw features space using NCA [27]. It is a supervised learning method for classifying multivariate data into distinct classes according to a given distance metric over the data [37]. It is non-parametric, that is, it does not require any parameter or assumption about the statistical distribution of the samples. It ranks the features with regularization to learn the feature weights for minimization of an objective function that measures the average leave-one-out (LOO) classification loss on the labeled training data.

The set of training samples is defined as $S = \{(x_1, y_1), \dots, (x_n, y_n), \dots, (x_N, y_N)\}$, where x_n is the d -dimensional feature vector, $y_n \in \{1, 2, \dots, C\}$ represents its corresponding class label, C is the number of classes, and N is the number of training samples. The goal of the feature selection algorithm is to find a weighting vector w that leads itself to select features by optimizing the nearest neighbor classification. In terms of the weighting vector w , the weighted distance between the two samples x_i and x_j is defined by:

$$d_w(x_i, x_j) = \sum_{l=0}^d w_l^2 |x_{il} - x_{jl}| \tag{3}$$

where w_l is a weight associated with the l th feature. The LOO technique is considered to maximize the classification accuracy on training set S . The probability distribution is an effective assumption to select any reference point from S for classification. Here, the probability of x_i selecting x_j as a reference point is given by:

$$p_{ij} = \begin{cases} \frac{\tau(d_w(x_i, x_j))}{\sum_{k \neq i} \tau(d_w(x_i, x_k))}; & i \neq j \\ 0; & i = j \end{cases} \tag{4}$$

where $\tau(z) = e^{(-z/\alpha)}$ is a kernel function and its width α is an input parameter. It influences the probability of each point being selected as the reference point. There are two limiting cases ($\alpha \rightarrow 0$ and $\alpha \rightarrow +\infty$) of α . If $\alpha \rightarrow 0$, the term $(-z/\alpha)$ becomes undefined and unable to select reference points in a probabilistic way. The nearest neighbor of query point can be selected as the reference to resolve this exceptional case. For $\alpha \rightarrow +\infty$, $p_{ij} \rightarrow 1/N_c$ (except for $i = j$, N_c is the number of candidate points) and all the candidate points have the same chance to be selected as reference points apart from the query point. Then, the probability for correct classification of the query point x_i is given by:

$$p_i = \sum_j y_{ij} p_{ij} \tag{5}$$

with $y_{ij} = 1$ only for $y_i = y_j$, and 0 otherwise. Therefore, the objective function can be defined

as [37]:

$$\varphi(w) = \zeta(w) - \eta \sum_{l=1}^d w_l^2 \tag{6}$$

where $\zeta(w) = \sum_i \sum_j y_{ij} p_{ij}$ is the approximate LOO classification accuracy. If $\alpha \rightarrow 0$, $\zeta(w)$ becomes a true classification accuracy. In order to perform feature selection and manage overfitting, a regularized parameter $\eta (>0)$ is introduced. The value of η can be tuned by cross-validation. Being differentiable, the derivative of $\varphi(w)$ can be computed as:

$$\begin{aligned} \frac{\partial \varphi(w)}{\partial w_l} &= \sum_i \sum_j y_{ij} p_{ij} \left[\frac{1}{\alpha} p_{ij} (\sum_{k \neq i} p_{ik} |x_{il} - x_{kl}| \right. \\ &\quad \left. - |x_{il} - x_{jl}|) w_l \right] - 2\eta w_l \\ &= \frac{2}{\alpha} \sum_i (p_i \sum_{k \neq i} p_{ik} |x_{il} - x_{kl}| \\ &\quad - \sum_j y_{ij} p_{ij} |x_{il} - x_{jl}|) w_l - 2\eta w_l \\ &= 2 \left[\frac{1}{\alpha} \sum_i (p_i \sum_{j \neq i} p_{ij} |x_{il} - x_{jl}| \right. \\ &\quad \left. - \sum_j y_{ij} p_{ij} |x_{il} - x_{jl}|) - \eta \right] w_l \end{aligned} \tag{7}$$

The above-mentioned derivative leads to the corresponding gradient-based update equation. Thus, the obtained weight vector is used to select the features. Each feature is ranked using its corresponding weight and the desired number of top-ranked features are selected to be used in the classification. The steps for the NCFS method are illustrated in Algorithm 1 [37].

Algorithm 1 NCA-Based Feature Selection (NCFS)

- 1: NCFS ($S, \gamma, \alpha, \eta, \delta$)
- S : training set, γ : initial step length, α : kernel width
- η : regularization parameter, ε : small positive constant
- 2: Initialization: $w^{(0)} = (1, 1, \dots, 1)$, $\varepsilon^{(0)} = -\infty$, $t = 0$
- 3: Repeat
- 4: for $i = 1$ to N // Execute only step 5
- 5: Compute p_{ij} and p_i with $w^{(t)}$ using Eq. (4) and (5)
- 6: for $l = 1$ to d // Execute only step 7
- 7: $\Delta l = 2 \left[\frac{1}{\alpha} \sum_i (p_i \sum_{j \neq i} p_{ij} |x_{il} - x_{jl}| \right. \right.$
 $\left. \left. - \sum_j y_{ij} p_{ij} |x_{il} - x_{jl}|) - \eta \right] w_l^{(t)}$
- 8: $t = t + 1$
- 9: $w^{(t)} = w^{(t-1)} + \gamma \Delta$
- 10: $\varepsilon^{(t)} = \varphi(w^{(t-1)})$
- 11: if $\varepsilon^{(t)} > \varepsilon^{(t-1)}$ then
 $\gamma = 1.01\gamma$
- else
 $\gamma = 0.4\gamma$
- 12: Until $|\varepsilon^{(t)} - \varepsilon^{(t-1)}| < \delta$
- 13: $w = w^{(t)}$
- 14: Return w

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The well-known publicly available datasets BCI Competition III (4a) and IV (2b) are used to conduct the experiments to evaluate the performance of the proposed

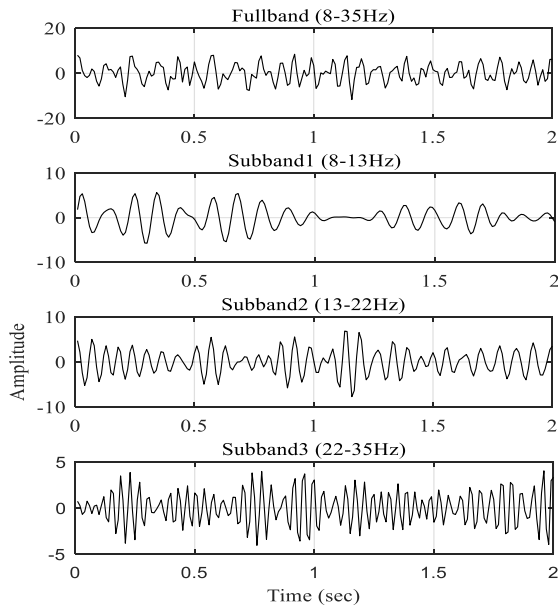


FIGURE 3. Four different subbands (8–35 Hz, 8–13 Hz, 13–22 Hz, and 22–35 Hz) obtained from channel C4 of subject 'aa' selected from the BCI Competition III dataset 4a.

method. Each trial of the datasets is decomposed into four subband signals and CSP-based features are extracted from each subband. The features of all subbands are combined to derive a high-dimensional feature vector. The NCFS-based supervised method is used to select the discriminative features. Thus, the obtained features are used for training SVM with linear kernel, followed by the evaluation of the classification performance with test data. For each subject, each trial of 2 s duration is extracted from the EEG data. The details of the data extraction method are described in Section II.

The fullband (8–35 Hz) and the three other subbands of channel C4 for subject 'aa' selected from dataset BCI Competition III (4a) are shown in Fig. 3. The CSP is applied on each frequency band to extract the spatial features. Four pairs of spatial filters resulting in eight features are selected from each subband. The CSP features obtained from each of the four bands are combined to comprise 32 ($= 4 \times 8$) dimensional feature vectors for each trial. Then, NCFS-based supervised feature selection algorithm is applied on the high-dimensional feature space. It uses the label information of training data and assigns a weight for each feature. The features are ranked based on the weights determined by NCFS method. A number of high-rank features are selected according to the obtained ranking. The UDFS, which is an unsupervised approach, is introduced in a previous work [35]. The class labels are available in the dataset BCI Competition III (4a). Thus, the supervised method for feature selection fits this experimental setup well. The values of the parameters used in the NCFS algorithm are set as: $\gamma = 2$, $\alpha = 1$, $\eta = 1/N$, $\delta = 1.0e-06$. The SVM is trained using the obtained feature of reduced dimension. The features of the same indices (as defined in the training set) are selected from the test dataset to evaluate the MI classification performance of the proposed method.

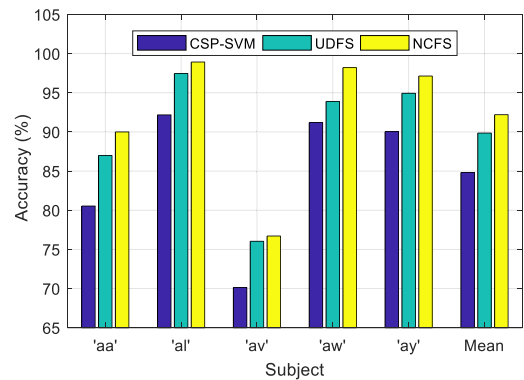


FIGURE 4. Performance (classification accuracy) comparison of the proposed method (NCFS) and the other two methods for individual subjects and their arithmetic mean using BCI Competition III dataset 4a.

The classification accuracy of each subject is measured by implementing the k -fold (here $k = 5$) cross-validation approach. For an individual subject, the dataset is divided randomly into k equal groups. The $(k-1)$ groups are assigned for training and one is designated for testing. The process is repeated k times. The classification accuracy is obtained by averaging the results of the k repetitions. The performance of the classification is evaluated by $A_{cc} = 100 \times (T_C/T_N)$, where T_N and T_C are the numbers of trials in the test dataset and the number of trials correctly recognized out of T_N , respectively.

Different experiments are conducted with BCI Competition III dataset 4a to illustrate the effectiveness of the proposed feature selection approach. The performances in terms of classification accuracy of MI-BCI using simple CSP-SVM (without feature selection), UDFS-based [35] feature selection, and the proposed NCFS-based feature selection methods for the five subjects are presented in Fig. 4. The CSP-SVM method is implemented with a full feature space. None of the features are discarded from the feature vector. The performance of CSP-SVM is always lower than that of the other two methods. It is observed that the dominant feature selection approach improves performance, whereas NCFS exhibits superior performance for all the subjects as well as on average across the subjects. The average MI classification accuracy of NCFS is 2.33% and 7.38% higher than that of UDFS and CSP-SVM, respectively. Using cross-validation procedure, individual repetition may produce a slightly different result. The average accuracy across all of the repetitions is taken as the final result for every subject.

After measuring the performance in terms of classification accuracy, the statistical test Friedman's one-way analysis of variance (ANOVA) is performed to study the significance level. Friedman's ANOVA is a non-parametric test [38] performed to detect the differences in the methods, including the proposed NCFS. Considering the result of Friedman's ANOVA, the methods have a significant main effect on accuracy ($p < 0.006$). To test the statistical significance of the methods, the Tukey–Kramer-based post-hoc test is performed [1]. From the results of the post-hoc test, the NCFS-based method achieves a more significant

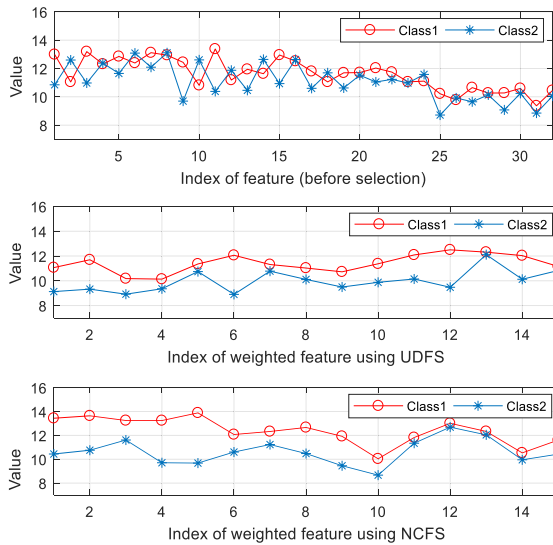


FIGURE 5. Significance of feature selection with BCI Competition III dataset 4a. Top panel: features of two classes before selection; middle panel: 15 top-ranked weighted features obtained by UDFS; bottom panel: 15 top-ranked weighted features obtained by NCFS.

improvement of classification accuracy for MI-BCI over all subjects compared with the other methods (NCFS vs. CSP-SVM: $p < 0.002$; NCFS vs. UDFS: $p < 0.03$).

An important reason for the performance improvement of the proposed method is the effective selection of features dominating in the correct classification. The two classes' raw features and the selected features using UDFS and NCFS of the same trial are shown in Fig. 5. The raw features include all of the 32 features, whereas the top-ranked 15 features are presented to illustrate the effectiveness of feature selection. It is observed that the top-ranked features are more separable from one class to another and, thus, are more discriminative than the raw features. The features selected using NCFS have higher disjointness between the classes than that of UDFS (as observed in Fig. 5).

The number of features selected for the MI classification is one of the vital factors that affect accuracy. The performances of the individual subjects as a function of the number of selected features using NCFS are illustrated in Fig. 6. In addition, the mean values across the subjects are presented together. The accuracies are varied over the dimension of selected features and the maximum accuracies of individual subjects are achieved with different numbers of selected features. The number of features corresponding to the maximum average accuracies across the subjects is taken as the feature dimension to conduct the rest of the experiments. The comparison of average accuracy between the UDFS [35] and the proposed NCFS as a function of the number of selected features is shown in Fig. 7. It is observed that the maximum classification accuracies of UDFS and NCFS are achieved by using 8 and 10 selected features, respectively. The average accuracy of the NCFS-based method is higher than that of UDFS method for the low-dimensional features space.

The MI classification accuracies of the proposed feature selection with different classifiers, namely, SVM [33], linear

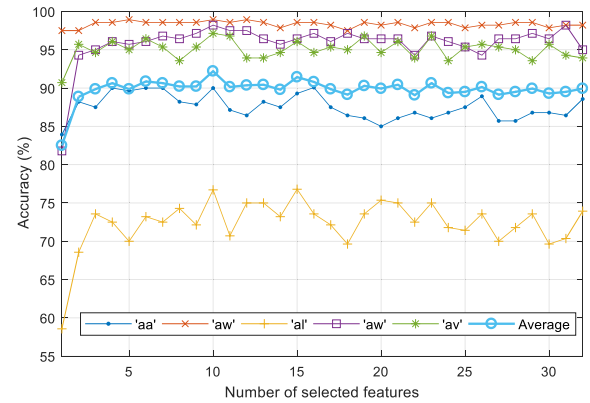


FIGURE 6. Classification performance of individual subjects ('aa', 'al', 'av', 'aw', 'ay') and their average as a function of the number of selected features (using NCFS).

TABLE 1. Comparison of different classifiers for UDFS and proposed NCFS methods.

Feature Selection	Classifier	Subjects					Mean±STD
		aa	al	Av	aw	ay	
UDFS [35]	SVM	86.98	97.45	76.04	93.90	94.94	89.86±8.65
	LDA	86.07	97.14	75.64	94.64	95.00	89.70±8.93
	KNN	84.29	97.50	75.00	94.29	94.64	89.14±9.36
NCFS	SVM	90.00	98.93	76.71	98.21	97.14	92.20±9.36
	LDA	90.07	97.50	76.07	97.07	96.07	91.36±9.05
	KNN	90.00	98.21	74.86	96.79	96.43	91.26±9.70

discriminant analysis (LDA) [39], and k-nearest neighbor (KNN) [40], are studied. The results for the BCI Competition III dataset 4a are presented in Table 1. In all of the cases, SVM performs better than LDA and KNN. Although LDA outperforms with subject 'aa' for NCFS, the average accuracy of SVM is significantly higher than that of the other classifiers. The proposed NCFS method with any classifier performs better than UDFS. The Tukey–Kramer-based post-hoc test is performed to verify the significance of SVM with NCFS compared to LDA and KNN. From the results of the Tukey–Kramer-based post-hoc test, the SVM classifier with the proposed NCFS achieves significant improvement of classification accuracy for MI-BCI over the subjects than other classifiers (NCFS-SVM vs. NCFS-LDA: $p < 0.05$; NCFS-SVM vs. NCFS-KNN: $p < 0.04$).

The BCI Competition III dataset 4a is also used to evaluate the MI classification performance of several recently reported methods [41]–[44]. The comparative performances in terms of classification accuracy of the proposed method with the recently developed algorithms are illustrated in Table 2. The average classification accuracy over all subjects of the proposed approach is 92.20%. The performance of this method is compared with the methods implemented using regularized Riemannian features (RRF) [15] and sparse group representation model (SGRM) of the CSP features [24]. The average classification accuracies of RRF and SGRM with dataset III (4a) are 87.21% and 77.70%, respectively. It is noted that the Riemannian manifold-based feature is used in

TABLE 2. Classification accuracy (%) on BCI competition III dataset 4a. the performance of the proposed method is compared with that of the recently developed seven algorithms. for each of the five subjects, the best result is marked in boldface.

Methods	Subjects					Mean±STD
	aa	al	Av	aw	ay	
RRF [15]	81.25	100.00	76.53	87.05	91.26	87.22±9.08
SGRM [24]	73.90	94.50	59.50	80.70	79.90	77.70±12.67
SSCSP [42]	72.32	96.42	54.10	70.54	73.41	73.36±13.50
SRCSP [43]	69.64	96.43	59.18	70.09	86.51	76.37±13.31
TKCSP [44]	68.10	93.88	68.47	88.40	74.93	78.76±10.54
AM-SVM [41]	86.61	100.00	66.84	90.63	80.95	85.00±11.00
UDFS [35]	86.98	97.45	76.04	93.93	94.94	89.86±8.65
Proposed NCFS	90.00	98.93	76.71	98.21	97.14	92.20±9.36

regularized Riemannian features (RRF) [15] method rather than CSP. The attractor metagene-based feature selection is used in [41] with proper parameter optimization of SVM (AM-SVM) to implement the MI classification for BCI application with an average accuracy of 85.00%. The NCFS-based method achieves a noticeable improvement of accuracy using the effective method for discriminative feature selection. It is observed that SSCSP method [42] uses sparse CSP to obtain an accuracy of 73.36%. The spatial regularization of CSP is implemented in SRCSP [43] with a classification accuracy of 76.37% using BCI Competition III (4a) dataset. The transfer kernel common spatial pattern (TKCSP) is introduced by Dai *et al.* [44]. The proposed method outperforms TKCSP by 13.44% in terms of classification accuracy. Moreover, the average performance across all subjects of our previous work involving UDFS [35] is 2.33% less than the accuracy of the NCFS-based proposed method.

Friedman’s ANOVA is performed to study the significance level. The test is performed to detect the differences in the performances of the various methods, including NCFS. According to the result of Friedman’s ANOVA, the methods have a significant main effect on classification accuracy ($p < 0.05$). To test the statistical significance of the methods mentioned in Table 2, the Tukey–Kramer-based post-hoc test is performed. Based on the results of this statistical test, the NCFS-based proposed method achieves a more significant improvement of performance for MI-BCI over the subjects than other methods (NCFS vs. RRF: $p < 0.04$; NCFS vs. SGRM: $p < 0.01$; NCFS vs. SSCSP: $p < 0.02$; NCFS vs. SRCSP: $p < 0.03$; NCFS vs. TKCSP: $p < 0.03$; NCFS vs. AM-SVM: $p < 0.03$; NCFS vs. UDFS: $p < 0.05$).

The MI classification accuracies of all the nine subjects of BCI Competition IV dataset 2b obtained by the proposed method NCFS are illustrated in Table 3. The values of the parameters used in the NCFS algorithm are kept similar to the implementation of BCI Competition III dataset 4a. To obtain the maximum average MI classification accuracy using BCI Competition IV dataset 2b, the eight top-ranked features are used. The average accuracy over the subjects achieved by NCFS is 81.52%. The results are compared with the recently

TABLE 3. Classification accuracy (%) on BCI competition IV dataset 2b. the performance of the proposed method (NCFS) is compared with that of the two recently developed algorithms (DLAV [14], SGRM [24], and UDFS [35]). for each of the nine subjects, the best result is marked in boldface.

Methods	Subjects									Mean±STD
	B01	B02	B03	B04	B05	B06	B07	B08	B09	
DLAV	76.10	67.30	71.80	95.40	82.30	82.10	77.50	75.30	75.90	78.19±07.96
SGRM	76.30	56.00	49.20	98.20	91.10	74.80	88.30	85.40	84.90	78.24±16.25
UDFS	76.09	58.64	53.45	99.38	83.83	76.96	83.15	90.66	83.48	78.40±14.53
NCFS	79.25	63.48	56.65	99.28	88.67	79.96	88.76	92.66	84.95	81.52±13.72

developed three methods, namely, deep learning with variational autoencoder (DLVA) [14], SGRM [24], and UDFS [35]. The average accuracies over the nine subjects derived by DLVA [14], SGRM [24], and UDFS [35] are 78.19%, 78.24%, and 78.40%, respectively. The average MI classification accuracy for the BCI Competition IV dataset 2b of the proposed method outperforms the three mentioned algorithms at least by 3.12%. The Tukey–Kramer-based post-hoc test is performed to test the statistical significance of the methods. Based on the results of this test, the NCFS-based proposed method achieves more significant improvement of performance for MI classification accuracy over the subjects than most of the mentioned methods (NCFS vs. UDFS: $p < 0.01$; NCFS vs. SGRM: $p < 0.03$; NCFS vs. DLVA: $p = 0.324$). Although the performance improvement of NCFS is not statistically significant compared to DLVA [14], the average classification accuracy of the proposed NCFS-based method outperforms DLVA by 3.33%.

The BCI Competition IV dataset 2b is also used in [13], [18]. In both of the methods, each EEG trial is divided into segments that are 2 s in length with a 1.9 s overlap. The effective length of the segment is only 0.1 s. Then, the maximum Kappa value is selected over the time course and employed as the evaluation criterion. It is a somewhat different way to evaluate the MI classification accuracy compared to the traditional approach, yet higher classification accuracies are obtained. In the proposed method and in the methods mentioned in Table 4, each trial is considered as a single sample for classification. Then, the classification accuracy (%) is measured by the number of correctly recognized samples over the total number of test samples.

Feature selection has a vital role in MI classification. The features that are relevant in performing the classification are selected while removing the irrelevant or less important features that do not contribute much to the target variable in order to achieve better accuracy for the classification. Irrelevant or partially relevant features can negatively impact model performance. Thus, the feature selection method has certain advantages in improving the classification performance. The mean accuracy (over all subjects) without feature selection is much lower than that of the methods with feature selection approach. The reason is that the method without feature selection uses additional features that are not relevant and also decrease the performance of the classifier.

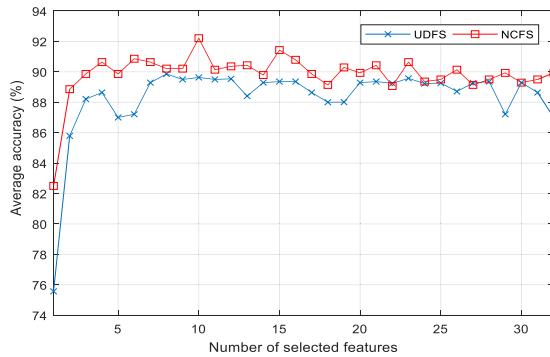


FIGURE 7. Average classification performance across five subjects (BCI Competition III dataset 4a) as a function of the number of selected features using UDFS and NCFS.

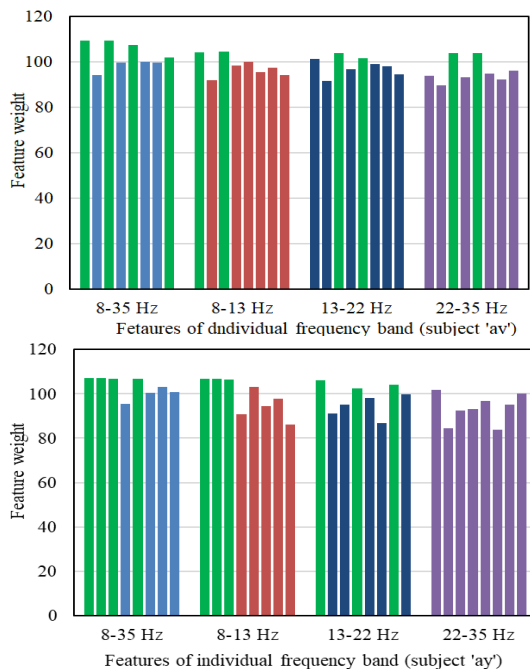


FIGURE 8. Features selected (green color) from the different subbands according to their weights assigned by the NCFS method.

The features extracted from the different narrowband signals also has a significant role in the improvement of classification accuracy. The features selected from the different subbands for subjects ‘av’ and ‘ay’ of dataset BCI Competition III (4a) are illustrated (single trial) in Fig. 8. The 10 desired features are selected from all subbands for subject ‘av’, whereas none of the 10 features are selected from subband 22–35 Hz for ‘ay’. It is noticed that different narrowband signals (subbands) contribute to building the subset of discriminative feature using the NCFS-based method. The evaluation results validate that the use of selected features improves the classifier performance. It is also observed that the proposed NCFS-based approach outperforms the recently reported algorithms.

The proposed NCFS-based method has some limitations. A fixed time window starting at the same temporal location of recorded EEG trial is used in this study. The latency in

responding to the stimuli and the duration of MI are subject dependent. Therefore, the use of such time window is not compliant with the concept of BCI. The CSP-based features used in this work are suitable only for the binary class and not extendable for multiclass MI classification problems. Keeping these limitations in mind, this work can be extended in the future to overcome the mentioned drawbacks.

V. CONCLUSIONS

A supervised feature selection method is implemented in this paper for MI classification using EEG signals. The experimental evaluation is performed by publicly available BCI Competition III dataset 4a and BCI Competition IV dataset 2b. With the first dataset, 30 out of 118 channels are used to represent a two-class (right hand and right foot movement) MI task for EEG classification in the BCI paradigm. The multichannel EEG is decomposed into four subbands that include mu, low beta, high beta, and fullband within the frequency range of 8–35 Hz. Four pairs of CSP features are extracted from each subband and then combined to derive a high-dimensional feature vector. Not all of the features are always relevant for classification. The proper elimination of irrelevant and redundant features makes the feature vector more discriminative and, thus, improves the classification performance. The proposed NCFS method effectively selects the discriminative features.

An unsupervised feature selection method is implemented in previous work [35]. Given the label information available in both BCI Competition III dataset 4a and BCI Competition IV dataset 2b, the supervised approach is more suitable. The proposed supervised feature selection method outperforms the unsupervised approach UDFS [35], as illustrated in Table 2 and 3. It is a non-parametric method, that is, it does not require any information about the statistical distribution of the samples. Along with reducing the amount of data used in machine learning, it alleviates the effect of the problem of dimensionality to improve the algorithms’ generalization performance. In this study, the optimal number of features for all subjects in a dataset is effectively implemented. It is extendable for multiclass problems of MI classification in the BCI paradigm.

Instead of using a filter bank to separate the rhythmic components, a number of bandpass filters are designed to extract the narrowband signals containing the components suitable for movement-related MI classification. In addition to the subband signals, the CSP-based features are also extracted from the full band (8–35 Hz) EEG signals. The inclusion of the fullband signal has a vital role in the discrimination of MI tasks. The scenario becomes clear when the proposed feature selection approach is implemented. The discriminative features are selected using NCFS from different subbands, as well as fullband EEG signals. Different experimental evaluations are conducted for the two-class MI-based EEG classification problem. The obtained results are compared with different recently developed algorithms. The experimental results establish that NCFS-based supervised feature

selection with SVM classifier outperforms the recently developed algorithms. Thus, the proposed combination of fullband and subband signals, as well as the implementation of feature selection approach, enhances the MI classification accuracy, as presented in Table 3 and 4. The foregoing therefore establishes the superiority of the proposed method.

REFERENCES

- [1] M. K. I. Molla, N. Morikawa, M. R. Islam, and T. Tanaka, "Data-adaptive spatiotemporal ERP cleaning for single-trial BCI implementation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 7, pp. 1334–1344, Jul. 2018.
- [2] T. Mulder, "Motor imagery and action observation: Cognitive tools for rehabilitation," *J. Neural Transmiss.*, vol. 114, no. 10, pp. 1265–1278, Oct. 2007.
- [3] L. van Dokkum, T. Ward, and I. Laffont, "Brain computer interfaces for neurorehabilitation—its current status as a rehabilitation strategy post-stroke," *Ann. Phys. Rehabil. Med.*, vol. 58, pp. 3–8, Feb. 2015.
- [4] S. R. Soekadar, N. Birbaumer, M. W. Slutzky, and L. G. Cohen, "Brain-machine interfaces in neurorehabilitation of stroke," *Neurobiol. Disease*, vol. 83, pp. 172–179, Nov. 2015.
- [5] P. Nuyujukian, J. M. Fan, J. C. Kao, S. I. Ryu, and K. V. Shenoy, "A high-performance keyboard neural prosthesis enabled by task optimization," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 1, pp. 21–29, Jan. 2015.
- [6] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, 2018, Art. no. 031005.
- [7] M. R. Islam, T. Tanaka, and M. K. I. Molla, "Multiband tangent space mapping and feature selection for classification of EEG during motor imagery," *J. Neural Eng.*, vol. 15, no. 4, Aug. 2018, Art. no. 046021.
- [8] S. U. Kumar and H. H. Inbarani, "PSO-based feature selection and neighborhood rough set-based classification for BCI multiclass motor imagery task," *Neural Comput. Appl.*, vol. 134, pp. 3239–3258, Mar. 2017.
- [9] K. Keng Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 2390–2397.
- [10] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.
- [11] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.
- [12] H. Higashi and T. Tanaka, "Common spatio-time-frequency patterns for motor imagery-based brain machine interfaces," *Comput. Intell. Neurosci.*, vol. 2013, Nov. 2013, Art. no. 537218.
- [13] J. Luo, Z. Feng, and N. Lu, "Spatio-temporal discrepancy feature for classification of motor imageries," *Biomed. Signal Process. Control*, vol. 47, pp. 137–144, Jan. 2019.
- [14] M. Dai, D. Zheng, R. Na, S. Wang, and S. Zhang, "EEG classification of motor imagery using a novel deep learning framework," *Sensors*, vol. 19, no. 3, p. 551, Jan. 2019.
- [15] A. Singh, S. Lal, and H. W. Guesgen, "Small sample motor imagery classification using regularized Riemannian features," *IEEE Access*, vol. 7, pp. 46858–46869, 2019.
- [16] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topography*, vol. 12, pp. 177–186, Mar. 2000.
- [17] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, Nov. 1999.
- [18] J. Luo, J. Wang, R. Xu, and K. Xu, "Class discrepancy-guided sub-band filter-based common spatial pattern for motor imagery classification," *J. Neurosci. Methods*, vol. 323, pp. 98–107, Jul. 2019.
- [19] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, p. 39, Mar. 2012.
- [20] H. Higashi and T. Tanaka, "Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1100–1110, Apr. 2013.
- [21] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (SBCSP) for brain-computer interface," in *Proc. 3rd Int. IEEE/EMBS Conf. Neural Eng.*, May 2007, pp. 204–207.
- [22] 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, Nov. 2015.
- [23] K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, and K. K. Ang, "A new discriminative common spatial pattern method for motor imagery brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, pp. 2730–2733, Nov. 2009.
- [24] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, and A. Cichocki, "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 631–641, Mar. 2019.
- [25] J. G. Dy and C. E. Brodley, "Feature selection for unsupervised learning," *J. Mach. Learn. Res.*, vol. 5, pp. 845–889, Jan. 2004.
- [26] L. Song, A. Smola, A. Gretton, K. M. Borgwardt, and J. Bedo, "Supervised feature selection via dependence estimation," in *Proc. 24th Int. Conf. Mach. Learn. (ICML)*, 2007, pp. 823–830.
- [27] J. Goldberger, S. Roweis, G. E. Hinton, and R. R. Salakhutdinov, "Neighborhood components analysis," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 17, 2005, pp. 513–520.
- [28] L. He, D. Hu, M. Wan, Y. Wen, K. M. von Deneen, and M. Zhou, "Common Bayesian network for classification of EEG-based multiclass motor imagery BCI," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 46, no. 6, pp. 843–854, Jun. 2016.
- [29] S. Wang and C. J. James, "Extracting rhythmic brain activity for brain-computer interfacing through constrained independent component analysis," *Comput. Intell. Neurosci.*, vol. 2007, Aug. 2007, Art. no. 41468.
- [30] S. R. Sreeja, J. Rabha, D. Samanta, P. Mitra, and M. Sarma, "Classification of motor imagery based EEG signals using sparsity approach," in *Proc. Int. Conf. Intell. Hum. Comput. Interact.*, 2017, pp. 47–59.
- [31] S. R. Sreeja, J. Rabha, K. Y. Nagarjuna, D. Samanta, P. Mitra, and M. Sarma, "Motor imagery EEG signal processing and classification using machine learning approach," in *Proc. Int. Conf. New Trends Comput. Sci. (ICTCS)*, Oct. 2017, pp. 61–66.
- [32] R. Leeb, C. Brunner, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, *BCI Competition 2008-Graz Data Set B*. Graz, Austria: Graz Univ. Technology, 2008.
- [33] F. Dmitriy and M. Ilya, "Support Vector Machines for Classification," in *Discrete Methods Epidemiology (DIMACS Series in Discrete Mathematics and Theoretical Computer Science)*, vol. 70. Camden, NJ, USA: Rutgers Univ., 2006, pp. 13–20.
- [34] J. Müller-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. 787–798, May 1999.
- [35] A. A. Shiam, M. R. Islam, T. Tanaka, and M. K. I. Molla, "Electroencephalography based motor imagery classification using unsupervised feature selection," in *Proc. Int. Conf. Cyberworlds (CW)*, Oct. 2019, pp. 239–246.
- [36] Y. Zhang, Y. Wang, G. Zhou, J. Jin, B. Wang, X. Wang, and A. Cichocki, "Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces," *Expert Syst. Appl.*, vol. 96, pp. 302–310, Apr. 2018.
- [37] W. Yang, K. Wang, and W. Zuo, "Neighborhood component feature selection for high-dimensional data," *J. Comput.*, vol. 7, no. 1, pp. 161–168, Jan. 2012.
- [38] K. Suefusa and T. Tanaka, "Asynchronous brain-computer interfacing based on mixed-coded visual stimuli," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 2119–2129, Sep. 2018.
- [39] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Mullers, "Fisher discriminant analysis with kernels," in *Proc. Neural Netw. Signal Process. 9th Proc. IEEE Signal Process. Soc. Workshop*, Aug. 1999, pp. 41–48.
- [40] G. Shakhnarovich, T. Darrell, and P. Indyk, *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice*. Cambridge, MA, USA: MIT Press, 2005.
- [41] S. Selim, M. M. Tantawi, A. S. Howida, and A. Badr, "A CSPAM-BA-SVM approach for motor imagery BCI system," *IEEE Access*, vol. 6, pp. 49192–49208, 2018.
- [42] M. Arvaneh, C. Guan, K. K. Ang, and H. C. Quek, "Spatially sparsified common spatial pattern to improve BCI performance," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 2412–2415.
- [43] F. Lotte and C. Guan, "Spatially regularized common spatial patterns for EEG classification," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3712–3715.
- [44] M. Dai, D. Zheng, S. Liu, and P. Zhang, "Transfer kernel common spatial patterns for motor imagery brain-computer interface classification," *Comput. Math. Methods Med.*, vol. 2018, Mar. 2018, Art. no. 9871603.



MD. KHADEMUL ISLAM MOLLA (Member, IEEE) received the B.Sc. and M.Sc. degrees in electronics and computer science from the Shahjalal University of Science and Technology, Bangladesh, in 1995 and 1997, respectively, and the Ph.D. degree from The University of Tokyo, Tokyo, Japan, in 2006.

In 2006, he joined as an Associate Professor with the Department of Computer Science and Engineering, University of Rajshahi, Bangladesh, where he is currently serving as a Professor. He was a JSPS Postdoctoral Fellow with the Department of Information and Communication Engineering, The University of Tokyo, from September 2006 to September 2008. He was a Research Fellow with the University of Alberta, Edmonton, AB, Canada, from November 2010 to October 2011. From January 2012 to June 2012, he was an Associate Professor with King Khalid University, Saudi Arabia. He was an NICT Fellow with The University of Tokyo, from January 2013 to July 2013. He was a JSPS Invited Researcher and a Visiting Professor with The University of Tokyo and the Tokyo University of Agriculture and Technology, Tokyo, the Kyushu Institute of Technology, Fukuoka, Japan, in several times. He is a member of APSIPA. He was the Organizing Chair of Asia-Pacific Signal and Information Processing Association (APSIPA) BioSiPS 2014. His research interests include application of signal processing, brain-computer interface, biomedical engineering, and time series analysis.



ABDULLAH AL SHIAM received the B.Sc. and M.Sc. degrees in computer science and engineering from Rajshahi University, Bangladesh, in 2017 and 2019, respectively. He was a Research Fellow with the Information and Communication Technology (ICT), Bangladesh, from 2018 to 2019. He is currently working as a Lecturer with the Department of Computer Science and Engineering, Varendra University, Rajshahi, Bangladesh. His research interests include brain signal

(EEG/MEG) processing, brain-computer interface (BCI), and biomedical engineering.



MD. RABIUL ISLAM received the B.Sc. and M.Sc. degrees from the Department of Computer Science and Engineering, University of Rajshahi, Bangladesh, in 2008 and 2009, respectively, and the Ph.D. degree from the Department of Electronic and Information, Tokyo University of Agriculture and Technology, Koganei, Tokyo, Japan. He was a Research Fellow with the National Science, Information and Communication Technology (NSICT), Bangladesh, from 2008 to 2009.

From October 2014 to September 2015, he was a Research Student with Tokyo University and Agriculture and Technology. He is currently an Assistant Professor (research) with the Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology. He is also working as an Assistant Professor (on leave) with the Department of Computer Science and Engineering, Pabna Science and Technology University, Pabna, Bangladesh. His research interests include brain signal (EEG/MEG) processing, audio and biomedical signal processing, brain-computer interface, and epilepsy.



TOSHIHISA TANAKA (Senior Member, IEEE) received the B.E., M.E., and Ph.D. degrees from the Tokyo Institute of Technology in 1997, 2000, and 2002, respectively. From 2000 to 2002, he was a JSPS Research Fellow. From October 2002 to March 2004, he was a Research Scientist with the RIKEN Brain Science Institute. In April 2004, he joined the Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, where he is currently a

Professor. In 2005, he was a Royal Society Visiting Fellow with the Communications and Signal Processing Group, Imperial College London, U.K. From June 2011 to October 2011, he was a Visiting Faculty Member with the Department of Electrical Engineering, University of Hawaii at Manoa. His research interests include a broad area of signal processing and machine learning, including brain and biomedical signal processing, brain-machine interfaces, and adaptive systems. He is a member of IEICE, APSIPA, the Society for Neuroscience, and the Japan Epilepsy Society. He is also a Co-Editor of *Signal Processing Techniques for Knowledge Extraction and Information Fusion* (with Mandic, Springer), 2008, and a Leading Co-Editor of *Signal Processing and Machine Learning for Brain-Machine Interfaces* (with Arvaneh, IET, U.K.), 2018. He has served as an Associate Editor and a Guest Editor of special issues in journals, including *Neurocomputing* and *IEICE Transactions on Fundamentals and Computational Intelligence and Neuroscience* (Hindawi). He currently serves as the Editor-in-Chief of *Signals* (MDPI). He serves as an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, *Neural Networks* (Elsevier), *Applied Sciences* (MDPI), and *Advances in Data Science and Adaptive Analysis* (World Scientific). Furthermore, he served as a Member-at-Large, board of governors (BoG) of Asia-Pacific Signal and Information Processing Association (APSIPA). He is a Distinguished Lecturer of APSIPA.

...