Neurocomputing 368 (2019) 133–145

Contents lists available at ScienceDirect

Neurocomputing

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

Classification of multiclass motor imagery EEG signal using sparsity approach

S. R. Sreeja*, Debasis Samanta

Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur 721302, India

ARTICLE INFO

Article history:
Received 14 December 2018
Revised 31 July 2019
Accepted 14 August 2019
Available online 27 August 2019

Communicated by Shaoting Zhang

Keywords:
Electroencephalogram
Motor imagery
Brain computer interface
Sparse representation
Sparsity-based classification

ABSTRACT

Motor imagery (MI) based brain–computer interface systems involving multiple tasks are highly required in many real-time applications such as hands and touch-free text entry, prosthetic arms, virtual reality systems, movement of a wheel chair, cursor movement, etc. The classification of MI data is the core computing in all these systems. However, the existing classification techniques are either computationally expensive or not so accurate or both. To address this limitation, in this work, a sparse representation based classification technique has been proposed to classify multi-tasks MI electroencephalogram data. The proposed method computes only wavelet energy directly from the segmented MI data and constructs a dictionary. The sparse representation from the dictionary is then used to classify given a test data. The proposed approach is faster as it works with only a single feature and without the need for any preprocessing. Further, with a reduced length of an imaging period, the proposed method provides accurate classification in a lesser computation time. The performance of the proposed approach has been evaluated and also compared with other classifiers reported in the literature. The results substantiate that the proposed sparsity approach performs significantly better than the existing classifiers.

1. Introduction

Brain–computer interface systems (BCIs) aim to provide a better and quality life for the people with disabilities [1]. A number of physiological sensors are known to develop a BCI system, out of which electroencephalogram (EEG) is popular among the practitioners as it satisfies both convenience criteria (i.e., non-intrusiveness, non-obtrusiveness, and simplicity) and effectiveness criteria (i.e., sensitivity, efficiency, and compatibility) [2]. A BCI system with an EEG captures the brain signal and the captured signals are then translated into control commands in several applications, such as navigation in virtual reality (VR) environments [3], movement of a cursor in computer screen [4], movement of a wheelchair [5], control of robot [6], prosthetic arm movements [7], etc. The EEG-based BCIs are mostly built using visually evoked potentials (VEPs), slow cortical potentials (SCPs), event-related potentials (ERPs), and sensorimotor rhythms (SMRs) [2]. Of these, a SMR-based BCI provides high degrees of freedom and has the potential to offer physical interaction as an alternative to bodily motor pathways for disabled users [2]. In addition, SMRs can be easily detectable in both healthy and disabled individuals. As EEG signal is not stationary and of low signal-to-noise ratio (SNR), hence in its strictest sense, classifying different motor imagery (MI) intention is a challenging task.

Of late, a number of initiatives have been taken to make robust MI-based BCIs. [8–20]. Looking at the existing literature, many works to classify MI tasks extract features from Common Spatial Patterns (CSP) followed by a classifier [8–15]. As MI-based EEG signals exhibit spatio-temporal patterns, CSP is a highly successful algorithm for capturing the appropriate MI features. Several variants of CSP have been proposed to capture spectral patterns [12,14] and to tackle multiclass problems [8,13,15,21]. However, in CSP-based methods, it is quite difficult to figure out the approximate number of CSP filters, it is computationally expensive, sometimes overfits with the data and works better only for binary classification problem [21].

The majority of the existing works on multiclass MI EEG signal data classification are based on support vector machine (SVM) [16,22], logistic regression (LR) [21], multivariate empirical mode decomposition [17], adaptive stacked regularized linear discriminant analysis (adaptive SRLDA) [18], linear prediction singular value decomposition (LP-SVD) based logistic model [19], probabilistic models such as naive Bayes Parzen window classifier [12], Bayesian network using Gaussian mixture model (GMM) [20,23], etc. Most of these classifiers are primarily constructed for binary classification problems. For classifying multiple MI tasks,

* Corresponding author.
E-mail addresses: sreejasr@iitkgp.ac.in (S. R. Sreeja), dsamanta@iitkgp.ac.in (D. Samanta).

https://doi.org/10.1016/j.neucom.2019.08.037
0925-2312/

these classifiers break into a series of binary classifiers using one-versus-one or one-versus-all strategy [16]. Nevertheless, it is time-consuming to train several binary classifiers when the number of MI tasks is large. Other than feature extraction-based methods, recently, deep learning models have also been implemented to MI EEG data. In [24], CSP followed by deep neural networks (DNN) have been employed for classifying two-class MI data. In [25], convolutional neural networks (CNN) and long short term memory (LSTM) networks have been combined to extract spatial and temporal features of EEG data and classify using deep forest models. In [26], frequential deep belief network (FDBN) has been employed using frequency domain features from MI EEG data. It may be noted that a deep learning model needs a large training data set and fine-tuning the neural network parameters, in fact, a non-trivial task.

Recently, sparse representation has aroused a significant resurgence of interest due to its tremendous performance in many applications, such as face recognition [27], speech recognition [28], image classification [29], signal denoising [30], medical image analysis [31,32], computer vision [33,34], and cardiac motion estimation [35]. In sparse representation, signals are represented as a sparse linear combination of few columns taken from a dictionary. Several predefined dictionaries, which use off-the-shelf bases have been reported in the literature such as Fourier, wavelets, Dirac, discrete cosine transform (DCT), etc. [36]. Nevertheless, the latest researches on sparse representation and dictionaries have shown that the dictionaries learned from data itself outperforms the predefined ones [30,35]. Alternatively, a dictionary can be designed or learned using a set of representative samples, for instance, a training data. Generally, these dictionaries are preferred to be over-complete, that is, the number of columns in the dictionary is more than the dimension of a column. As of now, limited researches have been done on dictionary learning and sparse representation based classification (SRC) to classify MI-based EEG signals [11,37–42].

A work on SRC for MI-based BCI has been reported by Shin et al. [37]. They applied CSP for signal pre-processing and conventional SRC for classification. However, the computational complexity of SRC needs to be minimized. Aiming to reduce the computational complexity, a simple adaptive SRC scheme has been developed by Shin et al. [11]. The results showed a comparatively better classification performance on two-class MI signals. Later, Zhang et al. [10] found a sparse method to automatically select significant filter bands to improve classification. Sreeja et al. [40] classified two-class MI-based EEG signals making use of band-pass filter and CSP, extracted wavelet energy as the feature to construct the dictionary and the conventional SRC for classification. They achieved a classification accuracy, which is better than the conventional SVM and LR methods. Jiao et al. [41] applied sparse group representation model (SGRM) for MI EEG data classification. However, the computational time still need to be reduced for real-time BCI applications. In fact, the performances of the existing SRC methods for MI-based BCIs [10,11,37,40,41,43] are solely governed by either band-pass or CSP filtering.

In this work, a new sparsity-based framework for multiclass MI EEG signals has been introduced to classify different MI tasks accurately in a lesser time. The proposed sparsity-based framework works without using any pre-processing technique such as band-pass and CSP filtering. The imaging period (it is the stimulation time of MI signals, which varies from one user to another) is a hindering factor in MI data classification, and it may be a reason that the machine learning based classification algorithms perform inefficiently. This work aims to make a better approach with a reduced length of an imaging period and without compromising the accuracy. Another objective is to develop an approach which would work better for all BCI users, independent of their efficiency in BCI training. In summary, the proposed sparsity-based MI data classification holds the following merits.

1. It works well without the use of band-pass and CSP spatial filters.
2. It performs well on any user irrespective of their proficiency in BCI training.
3. It works well with the channels which are present over the motor cortex region only.
4. It works fine with a reduced length of an imaging period, which is, in fact, a necessity for real-time BCI applications.
5. It performs well with the dictionary constructed using wavelet energy and this essentially speed up the system.
6. It is comparatively with a lower computational cost than the machine learning based classifiers.

The rest of the paper is organised as follows. In Section II, the proposed methodology is discussed in detail. The experimental results are reported in Section III. In Section IV, a discussion of the performance of the proposed approach is given. Finally, the paper concludes in Section V.

2. Proposed methodology

This section describes the MI data used for this research and then the steps followed in the proposed methodologies, namely data segmentation, feature extraction, designing a dictionary and sparsity-based classification approach.

2.1. Dataset description

The dataset IIIa from BCI competition III[1] has been used in this work. It is a dataset of four classes (motor imagery movements of left-hand, right-hand, foot and tongue) from three subjects (K3, K6, and L1) who had a different amount of proficiency in BCI training. Subject K3 was the most experienced, whereas L1 had a little experience and K6 was a beginner in BCI training. The recording was made with a 64-channel EEG amplifier from the Neuroscan, using the left mastoid for reference and the right mastoid as ground. The experiment was designed to instruct the subjects to perform imagery movements provoked by a visual cue. Each trial began with a blank black screen and at $t = 2$ s, a short beep tone was presented and a cross '+' displayed on the screen to raise the subject's attention (also see Fig. 1(b)). Then at $t = 3$ s, an arrow appearing for 1.25 s pointed either to the left, right, upward or downward. The subjects were asked to perform an imagery, say, left-hand, right-hand, foot or tongue movements according to the cues appeared. The respective MI should be executed until the cross disappears at $t = 7$ s. The number of runs for the subject K3 was *nine* whereas for K6 and L1 was *six*. Each of the four MI tasks was trained 10 times in each run. The EEG was sampled with 250 Hz and it was filtered between 1 and 50 Hz with notch-filter on. A notch filter was enabled to suppress line noise [16]. The BCI competition III dataset IIIa consists of both labeled and unlabeled data, where the class-labels are denoted as {1, 2, 3, 4} for the labeled trial set of left-hand, right-hand, foot and tongue MI tasks, respectively, and NaN for the unlabeled set. The data from the 60 channels had been recorded and stored in GDF format. The characterization of the dataset is given in Table 1. The detailed description of the dataset can be found in [16].

2.2. Methodology

The framework of the proposed approach is shown in Fig. 2. The raw EEG data is segmented and then a feature vector is

[1] http://www.bbci.de/competition/iii

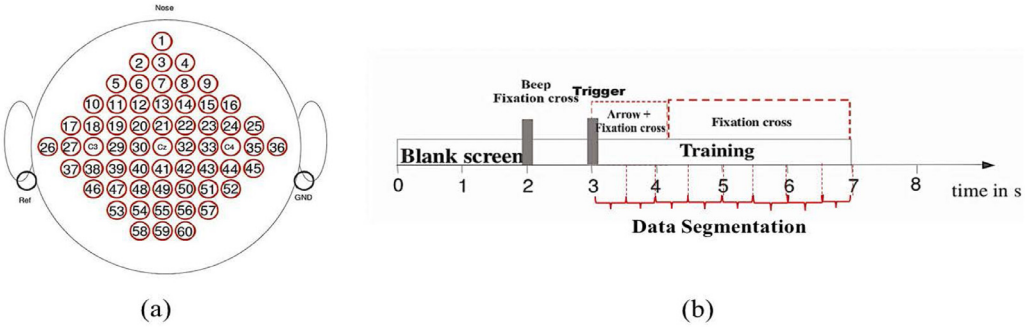

Fig. 1. (a) EEG electrode positions; (b) Training time of a trial.

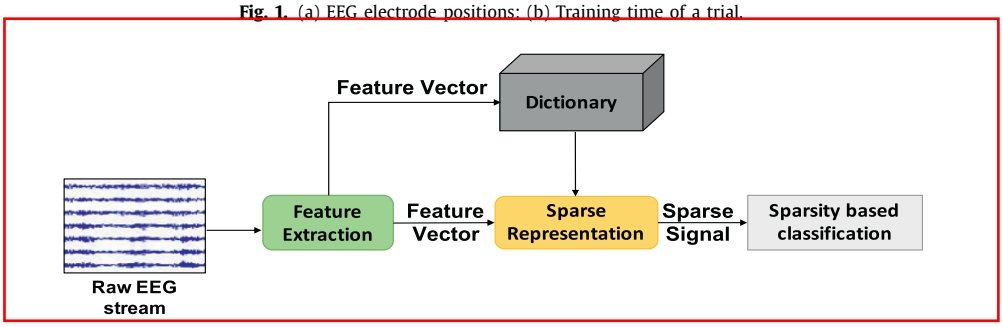

Fig. 2. Flowchart of the proposed approach.

Table 1
Characterization of the dataset.

Subjects	No. of runs	No. of trials	No. of labeled trials
K3	9	$9 \times 10 \times 4$	180
K6	6	$6 \times 10 \times 4$	120
L1	6	$6 \times 10 \times 4$	120

computed. It may be noted that the feature vector for a trial is typically a size of *number of segments* $\times$ *number of channels*. After segmentation, the given dataset is randomly split into training and testing data using k-fold cross-validation. The dictionary is built using the extracted feature vectors from all trials meant for training data. Finally, for the given test data, a sparse coefficient vector is estimated and sparsity-based classification is carried out. A detail explanation of each of the above-mentioned processes is given in the following.

2.2.1. Feature extraction

A raw MI-based EEG signal is divided into smaller segments. The segmentation is done where four second continuous imaging time ($T = 3$ s to $T = 7$ s) of each trial is divided into eight segments with half a second data in each segment. Because the imaging period of one MI could not last more than half a second [20]. Each segmentation is called as an *epoch* and the size of an epoch is $\mathbb{R}^{N_{t1} \times N_{ch}}$, where $N_{t1} = 125$, number of samples in half a second and $N_{ch} = 60$, number of channels.

Sreeja et al. [40] found that wavelet energy produces a good differentiation between the MI classes and yields a better accuracy than with other features. With this revelation, in this work, the wavelet energy has been extracted as the feature from each epoch $\mathbf{X} \in \mathbb{R}^{N_{t1} \times N_{ch}}$ of each class C_i, $i = \{1, 2, 3, 4\}$. The wavelet energy $\mathbf{x_w}$ is calculated using *coif1* wavelet with a level-1 decomposition [44]. As the main objective of this work is to reduce the overall computation time, the number of wavelet decomposition level is chosen as one. It has been observed in our experiments that the performance of *coif1* wavelet function is comparable to other wavelet functions, such as Daubechies, Symlets, etc. The energy of the approximate ($\mathbf{x_{wA}} \in \mathbb{R}^{1 \times N_{ch}}$) and detail ($\mathbf{x_{wD}} \in \mathbb{R}^{1 \times N_{ch}}$) coefficients are calculated separately and appended into a single vector, that is, $\mathbf{x_w} = \mathbf{x_{wA}} \frown \mathbf{x_{wD}}$, where $\frown$ denotes appending operation, $\mathbf{x_w} \in \mathbb{R}^{1 \times m}$ and $m = 2 \times N_{ch}$. The transpose of the feature vector $(\mathbf{x_w})^T \in \mathbb{R}^{m \times 1}$ of each epoch forms the training sample.

2.2.2. Dictionary construction

The next goal is to construct a dictionary, say, $\mathbf{D}$, which is capable of retaining all possible information about the input signal. The feature vectors extracted from each epoch of a class, say, C_i are stacked together to create a sub-dictionary matrix, say, $\mathbf{D_i}$, and it is defined as

$$\mathbf{D_i} = [\mathbf{d_{i,1}}, \mathbf{d_{i,2}}, \mathbf{d_{i,3}}, \ldots, \mathbf{d_{i,n_i}}] \in \mathbb{R}^{m \times n_i} \quad (1)$$

where $\mathbf{d_{i,j}} \in \mathbb{R}^{m \times 1}$, m is the dimension of the single feature vector and $j = 1, 2, \ldots, n_i$, and n_i is the total number of training samples extracted for the class C_i, where $i = \{1, 2, 3, 4\}$. Concatenating the dictionary matrix of all the classes yields an over-complete ($n > m$) dictionary $\mathbf{D}$, that is, $\mathbf{D} := [\mathbf{D_1}, \mathbf{D_2}, \mathbf{D_3}, \mathbf{D_4}] \in \mathbb{R}^{m \times n}$ where $n = \sum n_i$. The dictionary can be built pondering the feature vectors of all the subjects into a single dictionary. Indeed, it increases the dimension of the dictionary and time-consuming. Since the objective of this work is to reduce the time taken and to make the approach work better for all the users, a distinct dictionary for each subject has been constructed. A schematic representation of the dictionary $\mathbf{D}$ for a subject is shown in Fig. 3.

2.2.3. Sparse representation

For a given test signal, its feature vector is computed using the similar procedure used to create the atoms (i.e., columns) in $\mathbf{D}$. Thus, the test data and the atoms in the dictionary are of the same dimension. According to the CS (compressive sensing) theory [45], the signals of interest can be sparsely represented using some convenient basis. Given an over-complete dictionary matrix $\mathbf{D} \in \mathbb{R}^{m \times n}$ and the test data $\mathbf{y_i} \in \mathbb{R}^m$ of a class C_i, the objective of sparse representation is to calculate the sparse coefficient vector $\alpha_\mathbf{i} \in \mathbb{R}^n$, such that $\mathbf{y_i} = \mathbf{D}\alpha_\mathbf{i}$. In other words, the test data $\mathbf{y_i}$ can be

Fig. 3. A typical look of a dictionary for a subject.

sparsely represented as a linear combination of atoms of the dictionary $\{\mathbf{d_{i,j}}\}_{j=1}^{n_i}$. More precisely, it can be defined as

$$\mathbf{y_i} = \mathbf{D}\alpha_\mathbf{i} \tag{2}$$

which can be further represented as follows.

$$\mathbf{y_i} = \alpha_{\mathbf{i},\mathbf{1}}\mathbf{d_{i,1}} + \alpha_{\mathbf{i},\mathbf{2}}\mathbf{d_{i,2}} + \cdots + \alpha_{\mathbf{i},\mathbf{n_i}}\mathbf{d_{i,n_i}} \tag{3}$$

where $\alpha_\mathbf{i} = [\mathbf{0}, \ldots, \mathbf{0}, \alpha_{\mathbf{i},\mathbf{1}}, \alpha_{\mathbf{i},\mathbf{2}}, \ldots, \alpha_{\mathbf{i},\mathbf{n_i}}, \mathbf{0}, \ldots, \mathbf{0}]^\mathbf{T}$ is a sparse coefficient vector. The non-zero entries in $\boldsymbol{\alpha_i}$ is associated with the class C_i. The sparsest solution for $\mathbf{y_i} = \mathbf{D}\alpha_\mathbf{i}$ can be obtained by minimizing l_0-norm [46–48] and it is given as

$$\min_{\alpha_\mathbf{i}} \|\alpha_\mathbf{i}\|_0 \text{ subject to } \mathbf{y_i} = \mathbf{D}\alpha_\mathbf{i} \tag{4}$$

where $\|.\|_0$ is l_0-norm, which counts the non-zero entries in α_i. However, finding the sparsest solution for Eq. (4) is both numerically unstable and NP-hard [49], as it requires an extensive list of all feasible combinations for the positions of nonzero elements in α_i. According to the CS theory [45], under the assumption that the solution (α_i) is sparse enough, it is feasible to achieve a good approximate solution using l_0-minimization problem [36].

Many algorithms are available in the literature to solve Eq. (4) and they are broadly classified under two categories: convex relaxation methods [50,51] and greedy approximation methods [46,47]. The convex relaxation methods relax l_0-minimization to l_1-minimization problem (e.g., linear programming (LP), iterative shrinkage and thresholding algorithm (ISTA) [50], least absolute shrinkage and selection operator (LASSO) [51], etc.). The greedy based techniques, on the other hand, does not solve the l_0-minimization problem but gives an approximate solution by finding the best optimal solution (e.g., orthogonal matching pursuit (OMP) [47], matching pursuit (MP) [46], etc.). In this work, OMP algorithm has been employed due to its simplicity and efficiency [47].

The OMP algorithm [47] (also see Algorithm 1) is used to locate the sparse coefficients that approximately solves Eq. (4). This OMP algorithm attempts to choose the atoms that maximally correlate with the residual **r**. Then it augments the index of the chosen atom p_t to the active set P_t at each iteration, until the stopping condition is achieved. The natural way to stop the OMP algorithm is when the sparsity threshold is reached, but the threshold value has to be determined empirically and rather its not a convenient way of control. Therefore, the proposition is to use the residual **r** to control the moment when the algorithm is to be stopped. Hence, the stopping criteria is usually $\|\mathbf{r}\|_2 \leq \|\boldsymbol{\epsilon}\|_2$ [52], where ϵ is equivalent to the variance of the data $\mathbf{y_i}$. In the next step in Algorithm 1, $\mathbf{d}_{P_t}^\dagger$ indicates Moore–Penrose pseudo-inverse of $\mathbf{d}_{P_t}$, that is, $\mathbf{d}_{P_t}^\dagger := (\mathbf{d}_{P_t}^T \mathbf{d}_{P_t})^{-1}\mathbf{d}_{P_t}^T$, which employs orthogonalization at each iteration.

Algorithm 1: OMP algorithm.

Task: Approximate the optimization problem
$\min_\alpha \|\alpha\|_0$ subject to $\mathbf{y} = \mathbf{D}\alpha$
Input: Test data **y**, Dictionary **D**
Output: α
1: Initialize the residual $\mathbf{r}_0 = \mathbf{y}$, index set $P_0 = \emptyset$ and the iteration counter $t = 1$
2: Find the best matching atom,
$p_t = \arg\max_{k=1,2,..,n_i} \left|\langle \mathbf{r}_{t-1}, \mathbf{d}_{i,k}\rangle\right|$, $\sim i \in \{1, 2, 3, 4\}$
3: Update the index set, $P_t = P_{t-1} \cup \{p_t\}$
4: Compute the sparse coefficient $\alpha_t = \mathbf{d}_{P_t}^\dagger \mathbf{y}$
5: Update the residual $\mathbf{r}_t = \mathbf{y} - \mathbf{D}\alpha_t$
6: If the stopping condition is achieved ($\|\mathbf{r}_t\|_2 \leq \|\epsilon\|_2$), stop the algorithm and return $\alpha = \alpha_t$. Else, set $t = t + 1$ and return to Step 2.

2.2.4. Sparsity-based classification

Ideally, the sparse coefficients obtained will have entries corresponding to the atoms of the dictionary of that particular class. For example, if a test set from the foot MI class (C_3) is taken, the region where the non-zero entries lie in the sparse vector correspond to the atoms of the sub-dictionary D_3. A typical snapshot of the sparse representation model is shown in Fig. 4.

Since EEG signals are not stationary and of low SNR, the non-zero entries in the sparse vector may lie on another sub-dictionaries corresponding to other MI classes. Hence, specifying simple classification rules (like computing energy, number of non-zero entries, variance, minimum deviation, etc.) on the sparse coefficient vector helps to assign the test data $\mathbf{y}_i$ to the class-label C_i. In this study, the following four classification rules have been proposed.

$$R_1(y_i) = \arg\max_i \|\alpha_i\|_2 \tag{5}$$

$$R_2(y_i) = \arg\max_i \ \max(\text{non-zero}(\alpha_i)) \tag{6}$$

$$R_3(y_i) = \arg\max_i \ \max(\text{variance}(\alpha_i)) \tag{7}$$

$$\begin{aligned} r(y_i) &= \|\mathbf{y_i} - \mathbf{D}\alpha_\mathbf{i}\|_2 \\ R_4(y_i) &= \arg\min_i \ r(y_i) \end{aligned} \tag{8}$$

where $\|.\|_2$ is l_2-norm, which computes the energy of the sparse coefficients in Eq. 5, $R_2(y)$ computes the maximum number of non-zero entries in the sparse coefficient vector and $R_3(y)$ calculates the maximum of the variance in the sparse vector. The classification rule $R_4(y)$ measures the residual (deviation) between the test data with the training atoms and the estimated sparse vector, such that, $R_4(y) = \arg\min_i \|\mathbf{y_i} - \mathbf{D}\alpha_\mathbf{i}\|_2$. Apparently, smaller residual reveals a larger contribution to represent the test data.

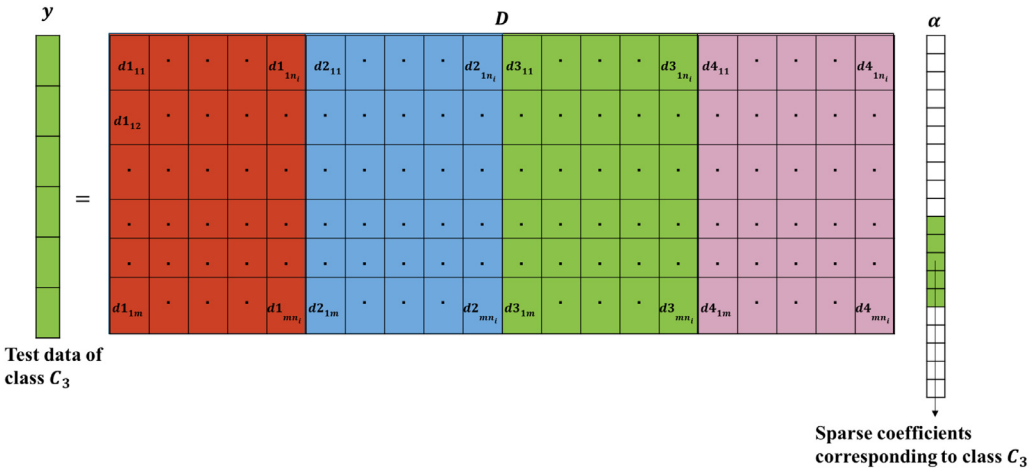

Fig. 4. Sparse representation model.

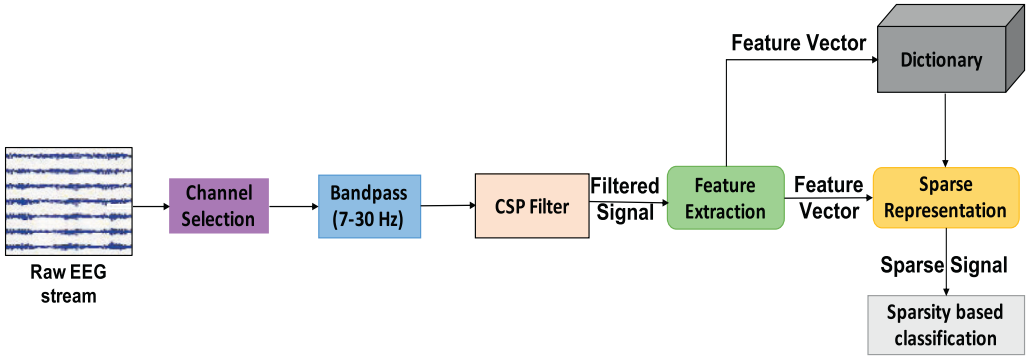

Fig. 5. A typical BCI system using sparsity approach.

2.3. System for MI data classification

A typical BCI system using sparsity approach is shown in Fig. 5. The EEG signals due to a brain activity are collected from the various channels and passed through the band-pass filter (single or multiple sub-bands) followed by the CSP spatial filter. Then feature extraction is done where features form the atoms of the dictionary. Once the dictionary is ready, sparsity-based classification is carried out to classify the different MI tasks.

3. Experiments and experimental results

In this section, the various experiments which have been carried out and the results obtained using the proposed method are presented. The following are the objectives of the experiments.

- To show the impact of filters in the sparsity-based classification technique.
- To show that the channels present over the motor areas are enough to be considered for sparsity approach.
- To find the best MI time duration for real-time BCI applications.
- To prove that wavelet energy is the best feature for constructing a dictionary.
- To prove that the proposed sparsity approach is better than the existing classifiers on classifying multi-class MI signals.

All the experiments have been executed on a machine with 3.2 GHz CPU, 4 GB RAM, and software (Python 2.7)[2]. The results have been recorded and evaluated using various evaluation measures such as k-fold cross-validation accuracy ($k = 10$), confusion matrix, precision-recall plot and ROC plot. The results observed vis-a-vis the objectives of the experiments are presented in the following.

3.1. Experiment with band-pass and CSP filter

In order to show the impact of filters on the proposed sparsity method, the different cases in the experiment, which have been considered are as follows.

Case 1: With various sub-bands and CSP filtering.
Case 2: With various sub-bands and without CSP filtering.
Case 3: Without band-pass and CSP filtering.

Case 1

An experiment using a band-pass filter with different frequency sub-bands followed by a CSP filter had been done. The experiment with commonly used sub-bands such as one [8–30 Hz], two [8–13 and 13–30 Hz], three [8–12, 12–16 and 16–30 Hz], six [8–12, 12–16, 16–20, 20–24, 24–28 and 28–32 Hz] and nine [4–8, 8–12, ..., and 36–40 Hz] using Chebyshev filter Type II with order 6 [12] was carried out. 10 CSP filters had been chosen empirically (a detail explanation of CSP filtering can be found in [40]). Then wavelet feature was calculated from the spatially filtered signal and the dictionary is constructed. It is noteworthy that increase in sub-bands increases the dimension of an atom in the dictionary. The 10-fold cross-validation accuracy obtained using various sub-bands and CSP for the three subjects and for four different classification rules are listed in Table 2.

[2] https://github.com/BCI-HCI-IITKGP/MI-classification-using-sparsity

Table 2
Performance of the sparsity approach with various sub-bands and CSP.

No. of band-pass and selected bands	Subjects	*k*-fold cross-validation accuracy (Mean ± std)					
		$R_1(y)$	$R_2(y)$	$R_3(y)$	$R_4(y)$	Training time (s)	Testing time (s)
1 [8-30] Hz	K3	**94.86 ± 0.30**	94.78 ± 0.33	**94.86 ± 0.30**	94.67 ± 0.36	122.84	0.559
	K6	90.37 ± 0.32	90.70 ± 0.25	90.37 ± 0.32	**90.88 ± 0.24**	120.48	0.556
	L1	90.44 ± 0.44	90.47 ± 0.43	90.44 ± 0.44	**90.48 ± 0.32**	120.48	0.557
2 [8-13, 13–30] Hz	K3	**94.88 ± 0.23**	94.82 ± 0.32	**94.88 ± 0.23**	94.76 ± 0.32	180.75	0.679
	K6	90.43 ± 0.46	90.76 ± 0.28	90.43 ± 0.46	**90.76 ± 0.20**	180.49	0.676
	L1	90.32 ± 0.20	90.28 ± 0.18	90.32 ± 0.20	**90.78 ± 0.24**	180.45	0.678
3 [8-12, 12–16, 16–30] Hz	K3	94.11 ± 0.28	94.18 ± 0.25	94.11 ± 0.28	**94.32 ± 0.22**	235.36	0.728
	K6	**89.65 ± 0.65**	89.50 ± 0.71	**89.65 ± 0.65**	89.63 ± 0.72	235.21	0.726
	L1	**90.95 ± 0.82**	90.56 ± 0.96	**90.95 ± 0.82**	90.52 ± 1.03	235.35	0.726
6 [8-12, 12–16, ..., 28–32] Hz	K3	94.82 ± 0.33	94.45 ± 0.45	94.82 ± 0.33	**94.90 ± 0.32**	342.46	0.812
	K6	90.88 ± 0.40	**91.02 ± 0.30**	90.88 ± 0.40	90.75 ± 0.30	341.41	0.810
	L1	**90.25 ± 0.32**	89.22 ± 0.28	**90.25 ± 0.32**	89.04 ± 0.20	341.40	0.813
9 [4-8, 8–12, ..., 36–40] Hz	K3	**94.88 ± 0.12**	94.85 ± 0.16	**94.88 ± 0.12**	94.88 ± 0.20	958.62	0.924
	K6	90.20 ± 0.23	90.24 ± 0.18	90.20 ± 0.23	**90.30 ± 0.15**	958.55	0.921
	L1	90.56 ± 0.34	**90.68 ± 0.28**	90.56 ± 0.34	90.52 ± 0.31	958.51	0.921

Table 3
Performance of the sparsity approach with various sub-bands and without CSP.

No. of band-pass and selected bands	Subjects	*k*-fold cross-validation accuracy (Mean ± std)					
		$R_1(y)$	$R_2(y)$	$R_3(y)$	$R_4(y)$	Training time (s)	Testing time (s)
1 [8-30] Hz	K3	94.04 ± 0.28	94.42 ± 0.30	94.04 ± 0.28	**94.50 ± 0.28**	90.75	0.511
	K6	90.80 ± 0.25	90.80 ± 0.25	90.80 ± 0.25	**90.84 ± 0.24**	90.70	0.511
	L1	**90.56 ± 0.32**	90.50 ± 0.30	**90.56 ± 0.32**	90.45 ± 0.33	90.45	0.512
2 [8-13, 13–30] Hz	K3	**93.92 ± 0.26**	93.69 ± 0.28	**93.92 ± 0.26**	93.85 ± 0.30	147.79	0.534
	K6	90.05 ± 0.16	90.26 ± 0.21	90.05 ± 0.16	**90.26 ± 0.18**	147.54	0.533
	L1	**90.32 ± 0.26**	90.00 ± 0.28	**90.32 ± 0.26**	90.24 ± 0.20	147.50	0.534
3 [8-12, 12–16, 16–30] Hz	K3	**94.42 ± 0.32**	93.65 ± 0.30	**94.42 ± 0.32**	94.34 ± 0.34	183.50	0.689
	K6	89.38 ± 0.42	89.45 ± 0.30	89.38 ± 0.42	**89.55 ± 0.32**	183.26	0.686
	L1	89.98 ± 0.56	89.60 ± 0.48	89.98 ± 0.56	**90.14 ± 0.56**	183.28	0.686
6 [8-12, 12–16, ..., 28–32] Hz	K3	93.34 ± 0.45	93.65 ± 0.42	93.34 ± 0.45	**93.90 ± 0.38**	238.74	0.789
	K6	**90.08 ± 0.35**	89.82 ± 0.27	**90.08 ± 0.35**	89.85 ± 0.42	237.88	0.782
	L1	**90.02 ± 0.18**	89.86 ± 0.28	**90.02 ± 0.18**	89.88 ± 0.19	237.80	0.784
9 [4-8, 8–12, ..., 36–40] Hz	K3	94.98 ± 0.86	**95.85 ± 1.03**	94.98 ± 0.86	94.88 ± 0.90	355.76	0.886
	K6	89.65 ± 0.78	89.54 ± 0.74	89.65 ± 0.78	**90.04 ± 0.80**	355.02	0.880
	L1	**90.83 ± 0.66**	90.04 ± 0.72	**90.83 ± 0.66**	90.74 ± 0.91	354.98	0.878

Table 4
Performance of the sparsity approach without band-pass and CSP.

Subjects	*k*−fold cross-validation accuracy (Mean ± std)					
	$R_1(y)$	$R_2(y)$	$R_3(y)$	$R_4(y)$	Training time (s)	Testing time (s)
K3	93.23 ± 0.28	91.85 ± 0.32	93.34 ± 0.26	**94.12 ± 0.21**	76.46	0.466
K6	89.42 ± 0.29	89.05 ± 0.23	89.00 ± 0.29	**90.64 ± 0.22**	75.98	0.462
L1	89.84 ± 0.18	89.90 ± 0.16	89.92 ± 0.18	**89.98 ± 0.18**	75.34	0.463

The average execution time for training and testing taken by this experiment was noted down and it is given in Table 2. The training time includes the execution time required for band-pass filtering, CSP filtering, feature extraction and construction of the dictionary for the entire training data. The testing time includes band-pass filtering, CSP filtering, feature extraction, finding sparse vector and classification of a single test data. Using sparsity approach the approximate average accuracy obtained for the subjects K3, K6 and L1 are 94%, 90%, and 90%, respectively. The reason for the subject K3 with higher accuracy may be due to additional 3 runs of BCI training data. Paired *t*-test has been carried out on the results which signifies that there is no significant difference in the accuracy between single band-pass and multiple bands with $p = .45$ (at significance level, $\alpha = 5\%$).

Observation

- With the increased number of band-pass filters, both the training and testing time increases.
- The accuracy of a classifier with a single band-pass filter is more or less equivalent to that of multiple bands.

Case 2

In this case, the experiment was done only using band-pass filtering. The wavelet feature was calculated from the band-pass filtered signal, and the dictionary was constructed. The result observed is shown in Table 3. The reported training time includes the execution time of band-pass filtering, feature extraction, and dictionary design. The testing time includes band-pass filtering, feature extraction, finding sparse vector and classification of a single test data. The paired *t*-test on the result confirms that there is no significant difference between the accuracy of Case 1 and Case 2, with $p = .33$ at significance level, $\alpha = 5\%$.

Observation

- The accuracy obtained for this experiment is more or less equivalent to the accuracy obtained in Case 1.
- The execution time taken by this experiment is lesser than Case 1.

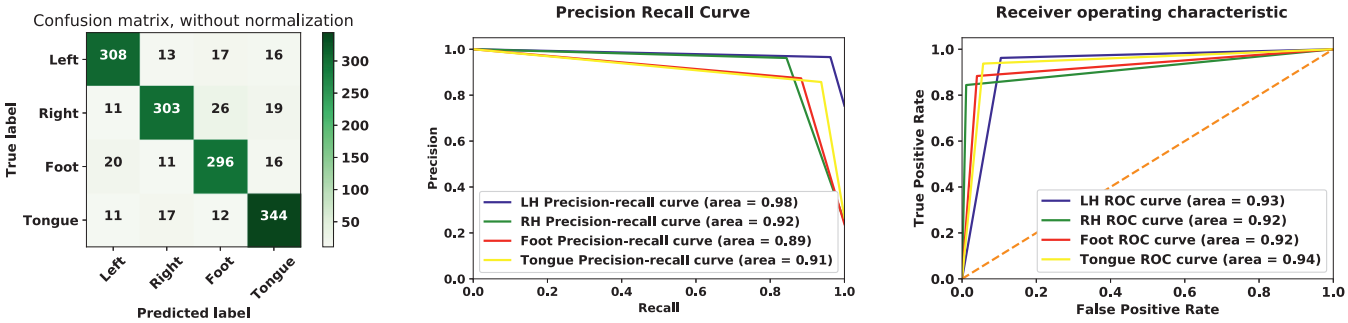

Fig. 6. Performance of the subject K3 with single band-pass and CSP.

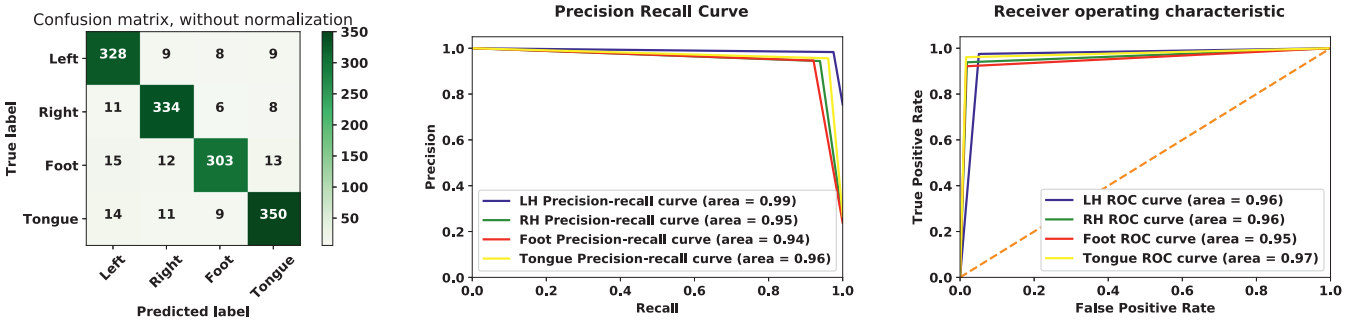

Fig. 7. Performance of the subject K3 with single band-pass and without CSP.

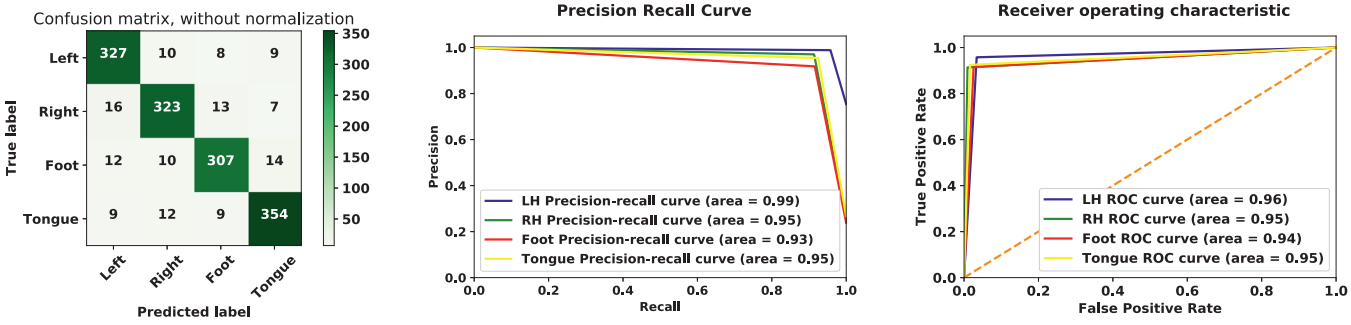

Fig. 8. Performance of the subject K3 without band-pass and CSP.

Case 3

In this case, the dictionary was constructed using wavelet features computed directly from the segmented EEG signal without passing through band-pass filter and CSP filter. The result observed is shown in Table 4. The paired t-test on the result confirms that there is no significant difference between the accuracy of Case 1 and Case 3, with $p = .21$ at significance level, $\alpha = 5\%$.

Observation

- The sparsity approach without using any filtering gives almost similar result to that of with sparsity approach using band-pass filter.
- The execution time required in this experiment is less when compared to Case 1 and Case 2.
- Case 3 proves to be an efficient sparsity approach for real-time BCI applications since it takes very less time for computation without loosing the accuracy.

Discussion

The experiment with filtering shows that increase in the number of sub-bands increases the computation time of the sparsity approach. Moreover, the results furnished in Table 4 prove that the sparsity approach performs well without using any pre-processing technique and with a less computation time. The results for Case 1, 2 and 3 in terms of confusion matrix, precision-recall plot and ROC plot using $R_4(y)$ are shown in Figs. 6, 7 and 8, respectively. In summary, the results confirm that the sparsity approach performs well in all the cases.

3.2. Experiment with a selected set of channels

Research on human brain proves that during MI tasks only the motor cortex gets activated [53,54]. The electrode positions C3 and C4 capture the neural activations of sensorimotor areas during movement imagination [55]. Hence, in this experiment, eight channels neighboring to C3 and C4 were considered. The selected electrodes were 17–19, 27, C3, 29, 37–39, 23–25, 33, C4, 35, 43–45 (see Fig. 1 (a)). 10-fold cross-validation accuracy for the subjects K3, K6 and L1 using 18 channels are listed in Table 5. The values in Table 5 clearly show a slight increase in accuracy and lesser execution time than the values in Table 4. Similarly, for the subject L1, the performance measures shown in Fig. 10 using 18 channels are better than the measures shown in Fig. 9 using all the 60 channels.

Table 5
Performance of the sparsity approach with selected channels.

Subjects	*k*-fold cross-validation accuracy (Mean ± std)					
	$R_1(y)$	$R_2(y)$	$R_3(y)$	$R_4(y)$	Training time (s)	Testing time (s)
K3	94.44 ± 0.24	92.93 ± 0.28	94.44 ± 0.24	**94.70 ± 0.18**	65.24	0.421
K6	89.79 ± 0.26	89.12 ± 0.25	89.79 ± 0.26	**90.83 ± 0.22**	64.86	0.420
L1	90.04 ± 0.16	90.20 ± 0.16	90.04 ± 0.16	**90.24 ± 0.18**	65.00	0.420

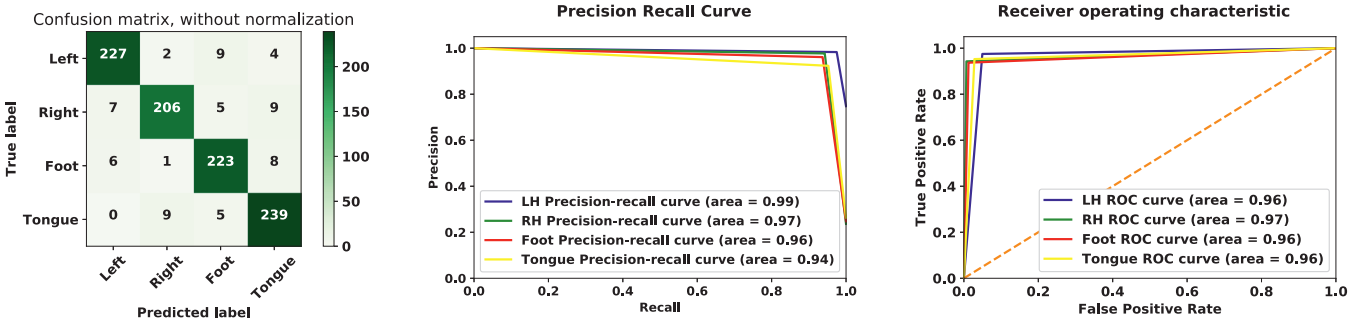

Fig. 9. Performance of the subject L1 using all the 60 channels.

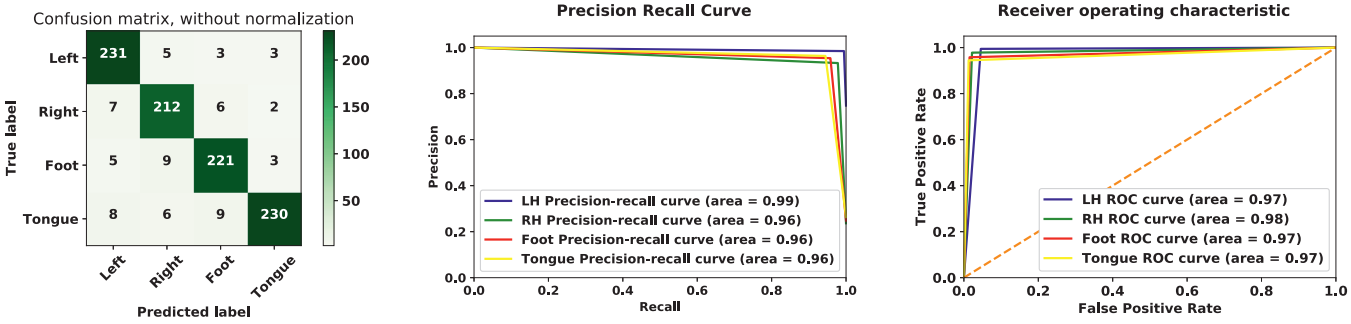

Fig. 10. Performance of the subject L1 with selected channels.

3.3. Experiment with different lengths of imaging period

In BCI competition III dataset IIIa, the interval of one MI is at least 3.5 s. Depending on subject's competency, one can perform MI for several times during a trial. In the initial experiment, the imaging period was considered as 0.5 s, since the imaging period could not last more than half a second [20]. Decreasing the length of the imaging period, without negotiating the performance will be greatly beneficial for real-time BCI applications. Hence, the experiment had been conducted with the different length of imaging periods, such as 0.1, 0.2, 0.3, 0.4 and 0.5 s. Note that, increase in the number of segments increases the dimension of the dictionary. Hence, the training time increases. However, for the experiment with the different length of the imaging period, the amount of non-zero coefficients differs. The sparse coefficients obtained for the subject K6 is shown in Fig. 11. The *x*-axis represents the dimension of the sparse vector and the *y*-axis represents the coefficient value.

From Tables 2 to 5 it is interesting to note that the results produced by the classification rules $R_1(y)$ and $R_3(y)$ are same in all the cases and most of the time $R_4(y)$ produces better accuracy. Hence, from this experiment on-wards only the classification rule $R_4(y)$ is considered. The results shown in Table 6 prove that when the imaging period is 0.1s, the testing time taken may be less but the accuracy of the sparsity approach is not good whereas when the imaging period is above 0.1 s, it achieves relatively a better performance. The reason is when the imaging time is very short, there won't be sufficient MI information for classification. Hence, one can consider 0.2 s or 0.3 s as the length of the imaging period for classifying different MI tasks, since it achieves better accuracy in a lesser computational time. From Table 6, it is also observed that subjects K3 and L1 achieve better performance with 0.2 s, whereas the subject K6 achieves the same with 0.3 s. This may be due to subject's proficiency with BCI training. Fig. 12 substantiates the performance of the subject L1 using 0.2 s as the length of an imaging period.

Table 6
Performance of the sparsity approach with different lengths of imaging period.

Imaging period (s)	Subjects			Training time (s)	Testing time (s)
	K3	K6	L1		
0.1	73.45 ± 0.46	70.92 ± 0.53	71.90 ± 0.57	148.04	**0.314**
0.2	**94.94 ± 0.25**	90.25 ± 0.23	**90.84 ± 0.17**	123.80	0.353
0.3	94.08 ± 0.16	**90.46 ± 0.19**	90.12 ± 0.26	92.30	0.406
0.4	94.26 ± 0.24	90.02 ± 0.38	90.72 ± 0.35	78.16	0.417
0.5	94.70 ± 0.18	90.83 ± 0.22	90.24 ± 0.18	65.03	0.421

3.4. Experiment with different features

Many features like time-domain, frequency-domain, wavelet-domain have been considered to classify MI EEG data using machine learning based classification algorithms. To substantiate the efficacy of the use of a large features set, in this study, nearly thirty different features [56] were extracted from the segmented EEG signal (taking 0.2 s as the length of the imaging period). A

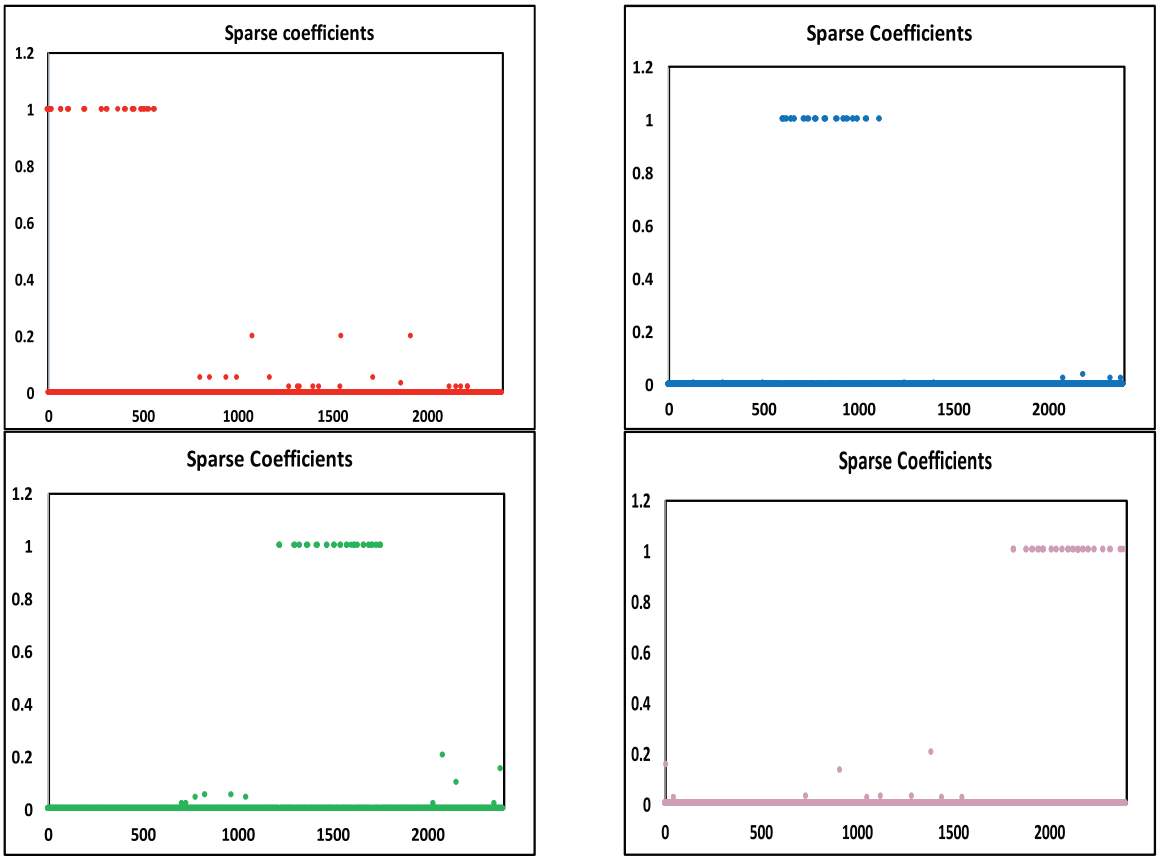

Fig. 11. Sparse coefficients for the subject K6, with 0.2 s as imaging period. The red, blue, green, and pink color sparse coefficients belong to left-hand, right-hand, foot and tongue MI's, respectively.

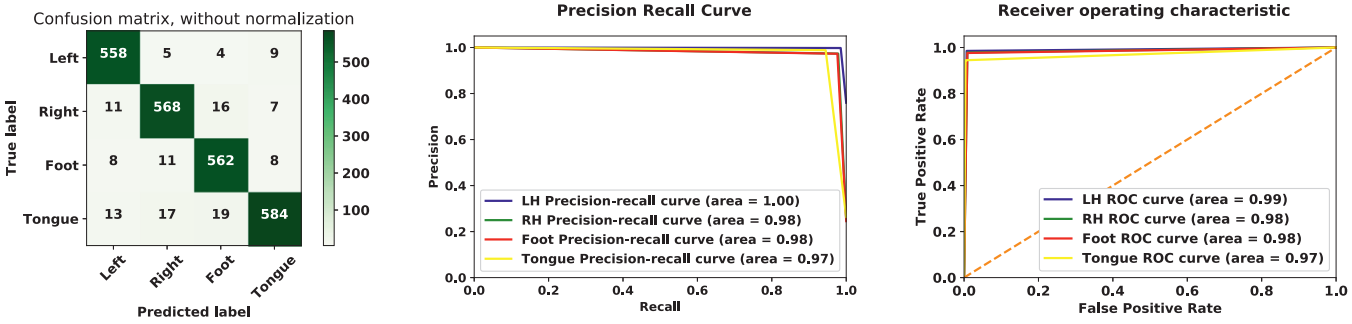

Fig. 12. Performance of L1 with 0.2s as imaging period.

single over-complete dictionary could be constructed considering all the features, but the dictionary size became too large, which took a lot of time for computation. Hence, it was essential to find out the best feature for sparsity-based classification. To do so, individual dictionaries were constructed using the extracted features and sparsity-based classification techniques were then carried out. Some of the features are listed in Table 7 which performed well with the sparsity approach. From Table 7, it can be observed that wavelet energy performs better than other features. It is also observed that, when different features are merged together, the sparsity approach performs well in terms of accuracy, but it is computationally too expensive.

3.5. Comparison with other classifiers

In order to prove the efficiency of the proposed method, a comparison between the proposed method with the four classifiers, namely SVM, LDA, k-NN and with a deep learning model (LSTM followed by multilayer perceptron (MLP)) is carried out. Since the above-mentioned machine learning classifiers are primarily designed for binary classification problem, to make it useful for multi-class classification problem, the one-versus-one scheme had been followed to break into a series of binary classifiers. Therefore, in total $\frac{C(C-1)}{2}$ binary classifiers were needed for C-class problem. These classifiers are implemented using Scikit Learn[3] and Tensorflow[4] in Python. The results obtained using different classifiers, following different steps, with 0.2 s as the length of the imaging period is shown in Table 8. The main objective of this comparison is to show that the machine learning methods applied to

[3] http://scikit-learn.org
[4] https://www.tensorflow.org

Table 7
Performance of the sparsity approach with different features.

Features	Subjects			Training time (s)	Testing time (s)
	K3	K6	L1		
Wavelet energy	**94.94 ± 0.25**	90.46 ± 0.19	90.12 ± 0.26	123.80	**0.353**
Wavelet entropy	94.32 ± 0.28	90.32 ± 0.18	90.01 ± 0.22	124.02	**0.353**
Bandpower (β)	92.13 ± 1.26	88.48 ± 1.16	89.76 ± 0.84	123.00	0.354
Bandpower (μ)	91.18 ± 0.84	88.74 ± 0.77	89.83 ± 0.95	123.76	0.354
Kurtosis	88.11 ± 1.18	84.83 ± 1.25	85.54 ± 1.33	122.80	**0.353**
Wavelet energy + wavelet entropy	94.90 ± 0.22	**90.88 ± 0.19**	**90.35 ± 0.22**	139.42	0.367
Wavelet energy + bandpower (β)	93.86 ± 0.25	89.72 ± 0.24	90.12 ± 0.31	142.78	0.368
Wavelet energy + bandpower (μ)	93.80 ± 0.29	90.02 ± 0.26	90.00 ± 0.25	142.79	0.368
Wavelet entropy + bandpower (β)	93.78 ± 0.43	89.82 ± 0.58	89.70 ± 0.54	142.82	0.368
Wavelet entropy + bandpower (μ)	93.16 ± 0.39	89.78 ± 0.62	89.86 ± 0.44	142.80	0.368
Bandpower (β) + bandpower (μ)	90.32 ± 0.73	87.42 ± 0.86	88.32 ± 0.65	139.80	0.366

Table 8
Comparison of the proposed sparsity approach with other classifiers.

Steps followed	Subjects	Metrics	SVM (rbf)	SVM (poly)	LDA	k-NN	LSTM+MLP	Sparsity approach
No band-pass and No CSP	K3	Accuracy	67.42 ± 2.92	61.83 ± 2.45	58.39 ± 2.60	60.44 ± 2.56	86.74 ± 1.18	**94.94 ± 0.25**
		Testing time	0.392	0.395	0.396	0.396	0.402	**0.353**
	K6	Accuracy	59.88 ± 2.20	57.32 ± 3.01	55.76 ± 2.72	58.62 ± 2.74	67.80 ± 1.36	**90.46 ± 0.19**
		Testing time	0.392	0.395	0.396	0.396	0.402	**0.353**
	L1	Accuracy	64.32 ± 1.53	60.60 ± 2.83	44.76 ± 2.98	45.65 ± 2.93	74.48 ± 1.78	**90.12 ± 0.26**
		Testing time	0.392	0.395	0.396	0.396	0.402	**0.353**
Single band-pass and CSP	K3	Accuracy	88.90 ± 1.32	86.45 ± 2.14	85.52 ± 2.42	88.86 ± 1.58	89.22 ± 1.46	**94.88 ± 0.28**
		Testing time	0.492	0.504	0.498	0.496	0.504	**0.456**
	K6	Accuracy	69.53 ± 1.15	65.62 ± 2.76	65.49 ± 2.78	68.82 ± 1.79	70.84 ± 1.86	**90.46 ± 0.19**
		Testing time	0.492	0.504	0.498	0.496	0.504	**0.448**
	L1	Accuracy	76.45 ± 1.47	75.32 ± 1.84	75.98 ± 3.04	76.15 ± 1.90	78.15 ± 1.63	**90.12 ± 0.26**
		Testing time	0.492	0.504	0.498	0.496	0.504	**0.452**

MI EEG signal processing are computationally expensive than the sparsity approach.

The parameters used for the above-mentioned classifiers are discussed here. For SVM, the experiment was carried out with choosing the kernel as *sigmoid, poly, rbf* and *linear*. With *rbf* and *poly* (degree = 3) as kernel and penalty parameter C as 1, a better accuracy was achieved. For LDA, the experiment with the parameters shrinkage as *auto* (automatic shrinkage using the Ledoit–Wolf lemma) and solver as *lsqr* (Least squares solution) was carried out. For k-NN, the number of nearest neighbors were chosen as 5, 10, 20, 50 and 100. With 50 nearest neighbors and with weights as the *distance* (Euclidean distance) better accuracy was achieved. Further, several deep learning models had been considered and the best model consists of a single layer of LSTM stacked with MLP of 3 hidden layers. The parameters were set as *num_units* in the range 800–900, learning rate in the range 0.00005–0.00007, batch size as 30–35 and epochs in the range 250–500. The window size is fixed as 125 and *tanh* as activation function. The best accuracy achieved is reported in Table 8.

Next, MI dataset was segmented with 0.2 s as length of the imaging period and processed with band-pass and CSP filter. For machine learning algorithms, thirty feature vectors (mentioned in Section IV-D) were extracted from each epoch. The best k optimal feature vectors were selected using sequential feed forward feature selection method (SFFS) [57]. These features were randomly split into training and testing data using k-fold cross-validation. The classifiers were trained with these training data and validated using the testing data. In case of deep learning, the processed EEG data was given as input to the model. The results obtained using different classifiers, following different steps are given in Table 8. The values listed in Table 8, show that the conventional classifiers perform well only with the use of band-pass filter and a spatial filter. Moreover, the time taken by the conventional classifiers are significantly higher than that of the proposed approach. The classifiers were also trained and tested only with wavelet energy as feature. However, the accuracies obtained by the conventional classifiers were not up to the mark. Only for the subject K3, the accuracy was nearly 80%. For other subjects, the accuracy obtained was approximately 60%. The following reasons may be attributed why conventional classifiers are not able to yield better result than sparsity-based approach.

1. The artifacts in EEG signal may be a deterring factor for the machine learning based classifiers.
2. Choosing appropriately hyper parameters of the models using machine learning and deep learning may be another concern.
3. Machine learning based classification techniques usually perform good with linearly separable and binary class data, whereas EEG data is nonlinear in nature and of multi-class.
4. Another deciding factor for the better performance is the selection of appropriate feature set representing the EEG data.

3.6. Comparison with state-of-the-art techniques

The proposed sparsity approach has been compared against the state of the art techniques. For multi-class MI classification, a number of works have been reported in the literature using machine learning algorithms, deep learning algorithms and sparsity based methods. In order to make the comparison worthy, the kappa (κ) coefficient values and accuracies had been calculated and compared with the existing results. Table 9 summarizes the results achieved by the proposed approach and the existing methods. It should be noted that, among the listed sparsity-based methods [38,39,42], only [38] has given the accuracy of multi-class MI EEG classification. Other methods [39,42] computed accuracies considering pair-wise combination of classes (left hand-right hand, left hand-foot, etc.). It is worth highlighting that in all reported studies, subject K3 achieved the highest accuracies while subject K6 produces the worst results. From the results in Table 9, it can be seen that the proposed method outperforms all other methods

Table 9
Comparison with state-of-the-art techniques.

Methodology	κ coefficient values				Classification Accuracy (%)			
	K3	K6	L1	Mean	K3	K6	L1	Mean
Fisher ratios of channel-freqency-time bins, mu and beta band, Multiclass CSP, SVM [22]	0.82	0.76	0.80	0.79	–	–	–	–
Surface laplacian, 8–30-Hz filter, CSP, SVM+kNN+LDA [22]	0.90	0.43	0.71	0.69	–	–	–	–
Infomax ICA, Amplitude spectra (Welch), linear PCA, and SVM [22]	0.95	0.41	0.52	0.63	–	–	–	–
Common Bayesian Network and SVM [20]	0.98	0.88	0.82	0.90	–	–	–	–
Multicalss information theoretic feature extraction (ITFE), and LR [21]	–	–	–	–	94.20	69.00	78.60	80.60
LP-SVD, logistic model tree (using Q- and Hotelling's T^2 statistics) [19]	–	–	–	–	90.00	76.25	77.91	81.38
Spectral features, modified mixed alternating least square based non-negative matrix factorization (MALS-NMF), SVM [38]	–	–	–	–	90.38	60.02	63.25	71.22
Band-pass filter, CSP, PSD and Variance, dictionary pair learning (DPL) [39] (Pair-wise combination of classes)	–	–	–	–	87.41	80.55	89.16	85.70
Band-pass filter, CSP, Temporally constrained sparse group spatial patterns (TSGSP), SVM [42] (Pair-wise combination of classes)	–	–	–	–	97.80	77.50	90.25	88.52
Band-pass filter, Multiclass CSP, Multiclass support matrix machine (MSMM) [15]	0.95	–	0.81	0.88	–	–	–	–
Proposed sparsity approach (without bandpass and CSP)	**0.96**	**0.90**	**0.91**	**0.92**	**94.94**	**90.46**	**90.12**	**91.84**

significantly. Particularly, for the subject K6, who was the beginner to BCI training, the proposed method achieves 90% accuracy, which indeed signifies a great improvement.

4. Discussion

The traditional practice to classify EEG data is using machine learning algorithms. Such algorithms are computationally expensive and the accuracy is not up to the level of acceptance. More significantly, the existing algorithms are feature-based and not able to take into account the subject variability, which is an important issue, in particular, in the context of MI EEG data. The proposed sparsity-based approach is able to overcome the aforementioned limitations. In the following, observed results according to the proposed approach are justified.

4.1. Why the accuracy is better with sparsity-based approach?

In the sparsity-based approach, an over-complete dictionary has been constructed using the feature extracted from the given MI signal without any filtering step. Hence, such a dictionary helps to retain the entire characteristics of the MI signal. Moreover, an over-complete dictionary is constructed in the sense that it is completely incoherent. That is, the dictionary $\mathbf{D := [D_1, D_2, D_3, D_4]}$ is composed of four sub-dictionary matrices corresponding to four MI classes and the atoms from different sub-dictionaries are uncorrelated (low coherence). Therefore, every test signal is mostly represented using signals within its own sub space. Hence, the signals can be sparsely represented such that, it cannot be represented simultaneously in other classes. This greatly helps in improving the performance of the classification.

4.2. Why is the proposed sparsity approach faster?

Generally, the algorithms available for MI classification follow a series of procedures, such as signal pre-processing (artifact removal and band-pass filtering), spatial filtering, feature extraction, feature reduction and classification. But, in the sparsity-based approach, all those tasks have been avoided, which significantly helps to improve the computation time. The features are directly extracted from the raw EEG signal, and an incoherent over-complete dictionary is constructed. Once the dictionary is obtained, the sparse coefficients are estimated and classification is carried out. Further, the main objective of the sparse representation is to find the sparse linear combination of few number of atoms for the given test signal. In this work, the sparse representation is solved by l_0 minimization. The algorithm OMP, a greedy technique, which helps to find an approximate l_0 solution by obtaining an optimal solution. It is known from the literature that OMP is an efficient, fast algorithm and it is really good at finding the sparsest solution. Moreover, the testing time includes only the time taken for calculating a single feature, finding sparse vector and classification. Hence, it performs faster compared to the machine learning algorithms.

4.3. How the proposed sparsity approach overcomes subject variability?

In the proposed approach, for each subject, a unique incoherent over-complete dictionary has been constructed. The rationale behind this approach is to address the issue that MI signals differ from one subject to another. The constructed dictionary holds the complete MI characteristics of a particular subject. The only requirement is that every user should be trained and the corresponding dictionary should be built prior to the use of a BCI system.

4.4. Threats to validity

The experiments and results reported in this paper are subjected to the following threats or assumption.

- All results mentioned are subject to the execution in the computing environment used.
- Results may be slightly different with different k-values in k-fold cross-validation.
- A limited database, namely dataset-IIIa from BCI Competition-III has been used. This is the dataset with only three subjects.
- No artifact has been removed from the EEG signals. The result may vary with EEG data with or without much noise due to a number of artefacts.
- Comparison of the proposed approach with the conventional classifiers using machine learning algorithms are based on the chosen parameters in our experiments, which may give another results, if tested with their optimal values.

5. Conclusion

Classification of motor imagery EEG data is a non-trivial task because of its high volume and dimension. The machine learning algorithms are computationally expensive and not so accurate. Further, as the machine learning algorithms are feature-based, they are not able to take the issues of subject variability, which is a pertinent issue in case of motor imagery data. The sparsity-based classification technique has been investigated in this work as an

alternative to machine learning algorithms for MI EEG signals. The proposed approach indeed outperforms the machine learning algorithms both in terms of computational cost as well as accuracy. Significantly, the subject-variability issue has been addressed satisfactorily. This observation is very important to develop any real-time BCI systems in many application domains, which are gaining popularity in recent times. Although the sparsity approach in this work has been intended only for motor imagery data, it can also be applied to other EEG datasets.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to thank Ministry of Human Resource Development (MHRD), Government of India for funding this study.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neucom.2019.08.037.

References

[1] J. Wolpaw, E.W. Wolpaw, Brain–Computer Interfaces: Principles and Practice, Oxford University Press, USA, 2012.
[2] B. He, B. Baxter, B.J. Edelman, C.C. Cline, W.Y. Wenjing, Noninvasive brain–computer interfaces based on sensorimotor rhythms, Proc. IEEE 103 (6) (2015) 907–925.
[3] A.S. Royer, A.J. Doud, M.L. Rose, B. He, EEG control of a virtual helicopter in 3-dimensional space using intelligent control strategies, IEEE Trans. Neural Syst. Rehabil. Eng. 18 (6) (2010) 581–589.
[4] D.J. McFarland, D.J. Krusienski, W.A. Sarnacki, J.R. Wolpaw, Emulation of computer mouse control with a noninvasive brain–computer interface, J. Neural Eng. 5 (2) (2008) 101.
[5] D. Huang, K. Qian, D.-Y. Fei, W. Jia, X. Chen, O. Bai, Electroencephalography (EEG)-based brain–computer interface (BCI): a 2-d virtual wheelchair control based on event-related desynchronization/synchronization and state control, IEEE Trans. Neural Syst. Rehabil. Eng. 20 (3) (2012) 379–388.
[6] C.J. Bell, P. Shenoy, R. Chalodhorn, R.P. Rao, Control of a humanoid robot by a noninvasive brain–computer interface in humans, J. Neural Eng. 5 (2) (2008) 214.
[7] D.J. McFarland, J.R. Wolpaw, Brain-computer interface operation of robotic and prosthetic devices, Comput. (Long Beach Calif) 41 (10) (2008).
[8] F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms, IEEE Trans. Biomed. Eng. 58 (2) (2011) 355–362.
[9] C.-Y. Chen, C.-W. Wu, C.-T. Lin, S.-A. Chen, A novel classification method for motor imagery based on brain–computer interface, in: International Joint Conference on Neural Networks (IJCNN), IEEE, 2014, pp. 4099–4102.
[10] Y. Zhang, G. Zhou, J. Jin, X. Wang, A. Cichocki, Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface, J. Neurosci. Methods 255 (2015) 85–91.
[11] Y. Shin, S. Lee, M. Ahn, H. Cho, S.C. Jun, H.-N. Lee, Simple adaptive sparse representation based classification schemes for EEG based brain–computer interface applications, Comput. Biol. Med. 66 (2015) 29–38.
[12] A.S. Aghaei, M.S. Mahanta, K.N. Plataniotis, Separable common spatio-spectral patterns for motor imagery BCI systems, IEEE Trans. Biomed. Eng. 63 (1) (2016) 15–29.
[13] A. Mahmood, R. Zainab, R.B. Ahmad, M. Saeed, A.M. Kamboh, Classification of multi-class motor imagery EEG using four band common spatial pattern, in: 39th IEEE International Conference on Engineering in Medicine and Biology Society (EMBC), 2017, pp. 1034–1037.
[14] D. Li, H. Zhang, M.S. Khan, F. Mi, A self-adaptive frequency selection common spatial pattern and least squares twin support vector machine for motor imagery electroencephalography recognition, Biomed. Signal Process. Control 41 (2018) 222–232.
[15] Q. Zheng, F. Zhu, J. Qin, P.-A. Heng, Multiclass support matrix machine for single trial EEG classification, Neurocomputing 275 (2018) 869–880.
[16] A. Schlögl, F. Lee, H. Bischof, G. Pfurtscheller, Characterization of four-class motor imagery EEG data for the BCI-competition 2005, J. Neural Eng. 2 (4) (2005) L14.
[17] C. Park, D. Looney, N. ur Rehman, A. Ahrabian, D.P. Mandic, Classification of motor imagery BCI using multivariate empirical mode decomposition, IEEE Trans. Neural Syst. Rehabil. Eng. 21 (1) (2013) 10–22.
[18] L.F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, D. Álvarez, R. Hornero, Adaptive stacked generalization for multiclass motor imagery-based brain computer interfaces, IEEE Trans. Neural Syst. Rehabil. Eng. 23 (4) (2015) 702–712.
[19] H. Baali, A. Khorshidtalab, M. Mesbah, M.J. Salami, A transform-based feature extraction approach for motor imagery tasks classification, IEEE J. Transl. Eng. Health Med. 3 (2015) 1–8.
[20] L. He, D. Hu, M. Wan, Y. Wen, K.M. von Deneen, M. Zhou, Common bayesian network for classification of EEG-based multiclass motor imagery BCI, IEEE Trans. Syst., Man, Cybern.: Syst. 46 (6) (2016) 843–854.
[21] M. Grosse-Wentrup, M. Buss, Multiclass common spatial patterns and information theoretic feature extraction, IEEE Trans. Biomed. Eng. 55 (8) (2008) 1991–2000.
[22] B. Blankertz, K.-R. Muller, D.J. Krusienski, G. Schalk, J.R. Wolpaw, A. Schlogl, G. Pfurtscheller, J.R. Millan, M. Schroder, N. Birbaumer, The BCI competition III: validating alternative approaches to actual BCI problems, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2) (2006) 153–159.
[23] L. He, B. Liu, D. Hu, Y. Wen, M. Wan, J. Long, Motor imagery eeg signals analysis based on bayesian network with gaussian distribution, Neurocomputing 188 (2016) 217–224.
[24] S. Kumar, A. Sharma, K. Mamun, T. Tsunoda, A deep learning approach for motor imagery EEG signal classification, in: 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), IEEE, 2016, pp. 34–39.
[25] Y. Shen, H. Lu, J. Jia, Classification of motor imagery EEG signals with deep learning models, in: International Conference on Intelligent Science and Big Data Engineering, Springer, 2017, pp. 181–190.
[26] N. Lu, T. Li, X. Ren, H. Miao, A deep learning scheme for motor imagery classification based on restricted boltzmann machines, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (6) (2017) 566–576.
[27] Z. Fan, M. Ni, Q. Zhu, E. Liu, Weighted sparse representation for face recognition, Neurocomputing 151 (2015) 304–309.
[28] P. Sharma, V. Abrol, A.K. Sao, Deep-sparse-representation-based features for speech recognition, IEEE/ACM Trans. Audio Speech Lang. Process. 25 (11) (2017) 2162–2175.
[29] L. Fang, C. Wang, S. Li, J.A. Benediktsson, Hyperspectral image classification via multiple-feature-based adaptive sparse representation, IEEE Trans. Instrum. Meas. 66 (7) (2017) 1646–1657.
[30] S.R. Sreeja, R.R. Sahay, D. Samanta, P. Mitra, Removal of eye blink artifacts from EEG signals using sparsity, IEEE J. Biomed. Health Inform. 22 (5) (2018) 1362–1372.
[31] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D.N. Metaxas, X.S. Zhou, Towards robust and effective shape modeling: sparse shape composition, Med. Image Anal. 16 (1) (2012) 265–277.
[32] L. Fang, S. Li, D. Cunefare, S. Farsiu, Segmentation based sparse reconstruction of optical coherence tomography images, IEEE Trans. Med. Imaging 36 (2) (2017) 407–421.
[33] M. Nejati, S. Samavi, S. Shirani, Multi-focus image fusion using dictionary-based sparse representation, Inf. Fusion 25 (2015) 72–84.
[34] B. Zhang, A. Perina, V. Murino, A. Del Bue, Sparse representation classification with manifold constraints transfer, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4557–4565.
[35] N. Ouzir, A. Basarab, H. Liebgott, B. Harbaoui, J.-Y. Tourneret, Motion estimation in echocardiography using sparse representation and dictionary learning, IEEE Trans. Image Process. 27 (1) (2018) 64–77.
[36] Y. Li, Z.L. Yu, N. Bi, Y. Xu, Z. Gu, S.-i. Amari, Sparse representation for brain signal processing: a tutorial on methods and applications, IEEE Signal Process. Mag. 31 (3) (2014) 96–106.
[37] Y. Shin, S. Lee, J. Lee, H.-N. Lee, Sparse representation-based classification scheme for motor imagery-based brain–computer interface systems, J. Neural Eng. 9 (5) (2012) 056002.
[38] N. Lu, T. Yin, Motor imagery classification via combinatory decomposition of ERP and ERSP using sparse nonnegative matrix factorization, J. Neurosci. Methods 249 (2015) 41–49.
[39] R. Ameri, A. Pouyan, V. Abolghasemi, Projective dictionary pair learning for EEG signal classification in brain computer interface applications, Neurocomputing 218 (2016) 382–389.
[40] S.R. Sreeja, J. Rabha, D. Samanta, P. Mitra, M. Sarma, Classification of motor imagery based EEG signals using sparsity approach, in: 9th International Conference on Intelligent Human Computer Interaction (IHCI), Springer, 2017, pp. 47–59.
[41] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, A. Cichocki, Sparse group representation model for motor imagery EEG classification, IEEE J. Biomed. Health Inform. 23 (2019) 631–641.
[42] Y. Zhang, C.S. Nam, G. Zhou, J. Jin, X. Wang, A. Cichocki, Temporally constrained sparse group spatial patterns for motor imagery BCI, IEEE Trans. Cybern. (99) (2018) 1–11.
[43] D. Wen, P. Jia, Q. Lian, Y. Zhou, C. Lu, Review of sparse representation-based classification methods on EEG signal processing for epilepsy detection, brain–computer interface and cognitive impairment, Front. Aging Neurosci. 8 (2016) 172.
[44] M.H. Alomari, E.A. Awada, A. Samaha, K. Alkamha, Wavelet-based feature extraction for the analysis of EEG signals associated with imagined fists and feet movements, Comput. Inf. Sci. 7 (2) (2014) 17.
[45] E.J. Candès, M.B. Wakin, An introduction to compressive sampling, IEEE Signal Process. Mag. 25 (2) (2008) 21–30.
[46] R. Gribonval, P. Vandergheynst, On the exponential convergence of matching

pursuits in quasi-incoherent dictionaries, IEEE Trans. Inf. Theory 52 (1) (2006) 255–261.
[47] T.T. Cai, L. Wang, Orthogonal matching pursuit for sparse signal recovery with noise, IEEE Trans. Inf. Theory 57 (7) (2011) 4680–4688.
[48] Z. Zhang, Y. Xu, J. Yang, X. Li, D. Zhang, A survey of sparse representation: algorithms and applications, IEEE Access 3 (2015) 490–530.
[49] R.G. Baraniuk, Compressive sensing [lecture notes], IEEE Signal Process. Mag. 24 (4) (2007) 118–121.
[50] M. Zibulevsky, M. Elad, L1-L2 Optimization in signal and image processing, IEEE Signal Process. Mag. 27 (3) (2010) 76–88.
[51] R. Tibshirani, Regression shrinkage and selection via the lasso: a retrospective, J. R. Stat. Soc.: Ser. B (Stat. Methodol.) 73 (3) (2011) 273–282.
[52] W.-J. Liang, T.-H. Chien, C.-S. Lu, Theoretical stopping criteria guided greedy algorithm for compressive cooperative spectrum sensing, Comput. Commun. 111 (2017) 165–175.
[53] G. Pfurtscheller, C. Neuper, Motor imagery activates primary sensorimotor area in humans, Neurosci. Lett. 239 (2–3) (1997) 65–68.
[54] R.S. Frackowiak, Human Brain Function, Academic press, 2004.
[55] G. Pfurtscheller, C. Brunner, A. Schlögl, F.L. Da Silva, Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks, Neuroimage 31 (1) (2006) 153–159.
[56] S.R. Sreeja, D. Samanta, P. Mitra, M. Sarma, et al., Motor imagery EEG signal processing and classification using machine learning approach, Jordan. J. Comput. Inf. Technol. (JJCIT) 4 (2) (2018) 80–93.
[57] A. Bashashati, M. Fatourechi, R.K. Ward, G.E. Birch, A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals, J. Neural Eng. 4 (2) (2007) R32.

S. R. Sreeja received her B.Tech degree in Information Technology and M.E degree in Computer Science Engineering from Anna University, Chennai, India, in 2010 and 2014, respectively. Presently, pursuing her Ph.D in the Department of Computer Science Engineering in Indian Institute of Technology, Kharagpur, India. Her current research interests are EEG signal processing, machine learning, sparse representation and brain-computer interface.

Debasis Samata received his Ph.D. degree in Computer Science and Engineering from IIT Kharagpur, India. He holds M.Tech. and B.Tech. degrees both in Computer Science and Engineering from Jadavpur University, Kolkata, India and Calcutta University, India, respectively. Presently, he is an Associate Professor in the Department of Computer Science and Engineering, IIT Kharagpur. His current research includes Human Computer Interaction, Brain Computing Interaction, Biometric-based System Security, and Data Analytics. For detail, please see http://cse.iitkgp.ac.in/~dsamanta/.