

Classification of Motor Imagery Based EEG Signals Using Sparsity Approach

S. R. Sreeja¹(✉), Joytirmoy Rabha¹, Debasis Samanta¹, Pabitra Mitra¹, and Monalisa Sarma²

¹ Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India
sreejasr@iitkgp.ac.in, joydan4123@gmail.com,
dsamanta@sit.iitkgp.ernet.in, pabitra@cse.iitkgp.ernet.in

² Subir Chowdhury School of Quality and Reliability, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India
monalisa@iitkgp.ac.in

Abstract. The advancement in brain-computer interface systems (BCIs) gives a new hope to people with special needs in restoring their independence. Since, BCIs using motor imagery (MI) rhythms provides high degree of freedom, it is been used for many real-time applications, especially for locked-in people. The available BCIs using MI-based EEG signals usually makes use of spatial filtering and powerful classification methods to attain better accuracy and performance. Inter-subject variability and speed of the classifier is still a issue in MI-based BCIs. To address the aforementioned issues, in this work, we propose a new classification method, spatial filtering based sparsity (SFS) approach for MI-based BCIs. The proposed method makes use of common spatial pattern (CSP) to spatially filter the MI signals. Then frequency bandpower and wavelet features from the spatially filtered signals are used to build two different over-complete dictionary matrix. This dictionary matrix helps to overcome the issue of inter-subject variability. Later, sparse representation based classification is carried out to classify the two-class MI signals. We analysed the performance of the proposed approach using publicly available MI dataset IVa from BCI competition III. The proposed SFS method provides better classification accuracy and runtime than the well-known support vector machine (SVM) and logistic regression (LR) classification methods. This SFS method can be further used to develop a real-time application for people with special needs.

Keywords: Electroencephalography (EEG)
Brain computer interface (BCI) · Motor imagery (MI)
Sparsity based classification · BCI for motor impaired users

1 Introduction

Brain-Computer Interface systems (BCIs) provides a direct connection between the human brain and a computer [20]. BCIs capture neural activities associated

with an external stimuli or mental tasks, without any involvement of nerves and muscles and provides an alternative non-muscular communication [21]. The interpreted brain activities are directly translated into sequence of commands to carry out specific tasks such as controlling wheel chairs, home appliances, robotic arms, speech synthesizer, computers and gaming applications. Although, brain activities can be measured through non-invasive devices such as functional magnetic response imaging (fMRI) or magnetoencephalogram (MEG), most common BCI are based on Electroencephalogram (EEG). EEG-based BCIs facilitates many real-time applications due to its affordable cost and ease of use [18].

EEG-based BCI systems are mostly build using visually evoked potentials (VEPs), event-related potentials (ERPs), slow cortical potentials (SCPs) and sensorimotor rhythms (SMR). Out of these potentials SMR based BCI provides high degrees of freedom in association with real and imaginary movements of hands, arms, feet and tongue [10]. The neural activities associated with SMR based motor imagery (MI) BCI are the so-called *mu* (7–13 Hz) and *beta* (13–30 Hz) rhythms [16]. These rhythms are readily measurable in both healthy and disabled people with neuromuscular injuries. Upon executing real or imaginary motor movements, it causes amplitude suppression or enhancement of *mu* rhythm and these phenomena are called event-related desynchronization (ERD) and event-related synchronization (ERS), respectively [16].

The available MI-based BCI systems makes use of spatial filtering and a powerful classification methods such as support vector machine (SVM) [17,18], logistic regression (LR) [13], linear discriminant analysis (LDA) [3] to attain good accuracy. These classifiers are computationally expensive and makes the BCI system delay. For real-time BCI applications, the ongoing MI events have to be detected and classified continuously into control commands as accurately and quickly. Otherwise, the BCI user especially motor impaired people may get irritated and bored. Moreover, for the same user, the observed MI patterns differ from one day to another, or from session to session [15]. This inter-personal variability of EEG signals also results in degraded performance of the classifier. The above issues motivates us to design a MI-based BCI system with enhanced accuracy, speed and no inter-subject variations for people with special needs.

With this purpose in hand, we propose a new spatial filtering based sparsity (SFS) approach in this paper to classify MI-based EEG signals for BCIs. In recent years, sparsity based classification has received a great deal of attention in image recognition [22] and speech recognition [9] field. In compressive sensing (CS), this sparsity idea was used and according to CS theory, any natural signal can be epitomized sparsely on definite constraints [5,8]. If the signal and an over-complete dictionary matrix is given, then the objective of the sparse representation is to compute the sparse coefficients, so that the signal can be represented as a sparse linear combination of atoms (columns) in dictionary [14]. If the dictionary matrix is designed from the best extracted feature of MI signal, it helps to overcome the issue of inter-personal and intra-personal variability, also enhances the processing speed and accuracy of the classifier.

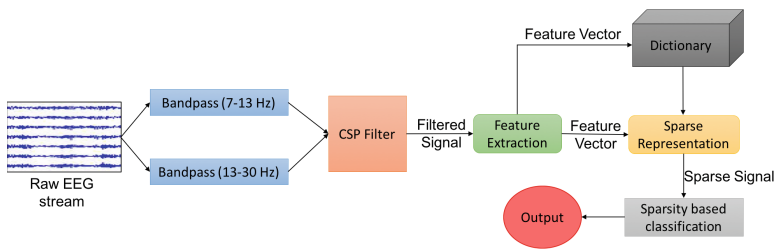


Fig. 1. Framework of the proposed SFS system.

The framework of the proposed system is shown in Fig. 1. In our proposed method, from 10–20 international system of EEG electrode placement, we considered only few channels located over motor areas for further processing. Later, the selected channels of EEG data are passed through a band-pass filter between 7–13 Hz and 13–30 Hz, as it is known from literature that most of MI signals lie within that frequency range. Then CSP is applied to spatially filter the signals and the features obtained from the filtered signals are used to build the columns (atoms) of dictionary matrix. This is an important phase in the proposed approach which is responsible for removing inter-personal variability and enhancement of classification accuracy. Later, sparsity based classification is carried out to discriminate the patterns of two-class MI signals. Furthermore, SFS method provides better accuracy and speed than the conventional support vector machine (SVM) and logistic regression (LR) classifier models.

Our paper is organised as follows. In Sect. 2, we present description of the data and the proposed technique in details. In Sect. 3, the experimental results and performance evaluation are presented. Finally, conclusions and future work are outlined in Sect. 4.

2 Data and Method

This section will describe the MI data used in this research and then the pipeline followed in the proposed method, that is, channel selection, pre-processing and spatial filtering based sparsity (SFS) classification of EEG-based MI data is discussed in detail.

2.1 Dataset Description

We used the publicly available dataset IVa from BCI competition III¹ to validate the proposed approach. The dataset consists of EEG recorded data from five healthy subjects (aa, al, av, aw, ay) who performed right-hand and right-foot MI tasks during each trial. According to the international 10–20 system, MI signals were recorded from 118 channels. For each subject, there were 140 trials for each

¹ <http://www.bbc.de/competition/iii>.

task, and therefore 280 trials totally. The measured EEG signal was filtered using a bandpass filter between 0.05–200 Hz. Then the signal was digitized at 1000 Hz with 16 bit accuracy and it is downsampled to 100 Hz for further processing.

2.2 Channel Selection and Preprocessing

The dataset consists of EEG recordings from 118 channels which is very large to process. As we are using the EEG signal of two class MI tasks (right-hand and right-foot), we extract the needed information from premotor cortex, supplementary motor cortex and primary motor cortex [11]. Therefore, from the 118 channels of EEG recording, 30 channels present over the motor cortex are considered for further processing. Moreover, removal of irrelevant channels helps to increase the robustness of classification system [19]. The selected channels are FC2, FC4, FC6, CFC2, CFC4, CFC6, C2, C4, C6, CCP2, CCP4, CCP6, CP2, CP4, CP6, FC5, FC3, FC1, CFC5, CFC3, CFC1, C5, C3, C1, CCP5, CCP3, CCP1, CP5, CP3 and CP1. The motor cortex and the areas of motor functions, the standard 10 ± 20 system of electrode placement of 128 channel EEG system and the electrodes selected for processing is shown in Fig. 2. The green and red circle indicates the selected channels and the red circle indicates the C3 and C4 channels on the left and right side of the scalp respectively.

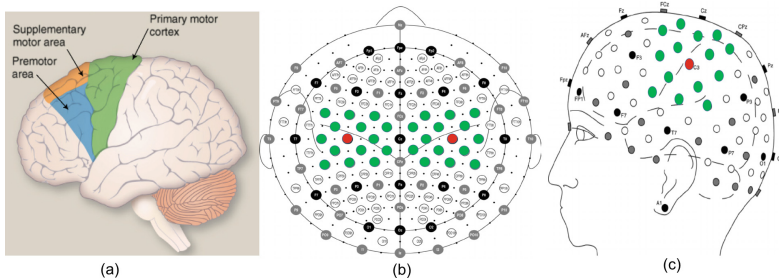


Fig. 2. (a) Motor cortex of the brain (b) Standard 10 ± 20 system of electrode placement for 128 channel EEG system. The electrodes in green and red colour are selected for processing (c) The anterior view of the scalp and the selected channels. (Color figure online)

From domain knowledge we know that, most brain activities related to motor imagery are within the frequency band of 7–30 Hz [16]. Bandpass filter can be used to extract the particular frequency band and also helps to filter out most of the high frequency noise. The bandpass filter can have as many sub-bands as one needed [12]. We have experimented with two sub-bands of 7–13 Hz and 13–30 Hz in the two-class MI signal classification problem. The choice of two sub-bands is due to the fact that mu (μ), beta (β) rhythms reside within those frequency bands. Then data segmentation is done where we used two second samples after the display of cue of each trial. Each segmentation is called as an *epoch*.

2.3 Proposed Spatial Filtering Based Sparsity Approach

The proposed spatial filtering based sparsity (SFS) approach follows three steps such as CSP filtering, design of dictionary matrix and sparsity based classification. A detailed explanation of each of these steps are given below.

CSP Filtering: Generally, for binary classification problems, CSP has been applied widely as it increases the variance of one class while it reduces the variance for the other class [1]. In this paper, how CSP filtering is applied for the given two-class MI-based EEG dataset is explained briefly. Let \mathbf{X}_1 and \mathbf{X}_2 be the two epochs of a multivariate signal related to right-hand and right-foot MI classes, respectively. They are both of size $(c \times n)$ where c is the number of channels (30) and n is the number of samples (100×2). We denote the CSP filter by

$$\mathbf{X}_i^{CSP} = \mathbf{W}^T \mathbf{X}_i \quad (1)$$

where i is the number of MI classes, \mathbf{X}_i^{CSP} is the spatially filtered signal, \mathbf{W} is the spatial filter matrix and $\mathbf{X}_i \in \mathbb{R}^{c \times n}$ is the input signal to the spatial filter. The objective of the CSP algorithm is to estimate the filter matrix \mathbf{W} . This can be achieved by finding the vector \mathbf{w} , the component of the spatial filter \mathbf{W} , by satisfying the following optimization problem:

$$\max_{\mathbf{w}} \left(\frac{\mathbf{w}^T C_1 \mathbf{w}}{\mathbf{w}^T C_2 \mathbf{w}} \right) \quad (2)$$

where $C_1 = \mathbf{X}_1 \mathbf{X}_1^T$ and $C_2 = \mathbf{X}_2 \mathbf{X}_2^T$. In order to make the computation easier to find \mathbf{w} , we computed \mathbf{X}_1 and \mathbf{X}_2 by taking the average of all epochs of each class. Solving the above equation using Lagrangian method, we finally have the resulting equation as:

$$C_1 \mathbf{w} = \lambda C_2 \mathbf{w} \quad (3)$$

Thus Eq. (2) becomes **eigenvalue decomposition problem**, where λ is the eigenvalue corresponds to the eigenvector \mathbf{w} . Here, \mathbf{w} maximizes the variance of right-hand class, while minimizing the variance of right-foot class. The eigenvectors with the largest eigenvalues for C_1 have the smallest eigenvalues for C_2 . Since we used 30 EEG channels, we will have 30 eigenvalues and eigenvectors. Therefore, CSP spatial filter \mathbf{W} will have 30 column vectors. From that, we select the first m and last m columns to use it as $2m$ CSP filter of \mathbf{W}_{CSP} .

$$\mathbf{W}_{CSP} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m, \mathbf{w}_{c-m+1}, \dots, \mathbf{w}_c] \in \mathbb{R}^{2m \times c} \quad (4)$$

Therefore, for the given two-class epochs of MI data, the CSP filtered signals are defined as follows:

$$\begin{aligned} \mathbf{X}_1^{CSP} &\in \mathbb{R}^{2m \times n} := \mathbf{W}_{CSP}^T \mathbf{X}_1 \\ \mathbf{X}_2^{CSP} &\in \mathbb{R}^{2m \times n} := \mathbf{W}_{CSP}^T \mathbf{X}_2 \end{aligned} \quad (5)$$

The above CSP filtering is simultaneously done for the filtered signals under the sub-bands of 7–13 Hz and 13–30 Hz.

Designing a Dictionary Matrix: The spatially filtered signals \mathbf{X}_1^{CSP} and \mathbf{X}_2^{CSP} are obtained for each epoch and for each sub-band. These spatially filtered signals are considered as the training signals in our experiment. Let the number of total training signals be N , considering each MI class i and each sub-band. Here, $i = 1$ for right-hand and $i = 2$ for right-foot class. The dictionary matrix can be designed with one type of feature or a combination of different features. In this work, we designed two types of dictionary matrix, one using frequency bandpower as feature and the other using wavelet transform energy as feature for each training signal. Initially, we experimented with many features like statistical, frequency-domain, wavelet-domain, entropy, auto-regressive coefficients, etc. But we found that bandpower and wavelet energy produces good differentiable between the two classes when it is plotted over the scalp. Figure 3 shows the spatial representation of bandpower and wavelet energy for two different MI classes. The Fig. 3(a) depicts that the bandpower of right-hand is scattered throughout the scalp while for right-foot the bandpower is high in the frontal region. In the same way, in Fig. 3(b) the wavelet energy is distributed all over the scalp for right-hand and only on a particular region for right-foot. Hence, these features are sufficiently good enough to discriminate the two MI classes.

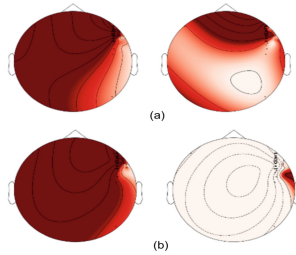


Fig. 3. Scalp plot of (a) bandpower of right-hand and right-foot MI respectively and (b) wavelet energy for right-hand and right-foot MI respectively.

From each row of the training signal, the second moment or the frequency bandpower and the wavelet energy using ‘coif1’ wavelet is calculated. This feature vector of each training signal forms the dictionary matrix. Concatenating the dictionary matrix of two-classes forms an over-complete dictionary. Since this dictionary matrix includes all the possible characteristics of the MI signals of the subjects, the inter-subject variability can be avoided. Figure 4 shows the dictionary constructed for the proposed approach. Thus, the dictionary matrix is defined as $\mathbf{D} := [\mathbf{D}_1; \mathbf{D}_2]$, where $D_i = [d_{i,1}, d_{i,2}, d_{i,3}, \dots, d_{i,N}]$. Each atom or column of the dictionary matrix is defined as $d_{i,j} \in \mathbb{R}^{2m \times 1}$, $j = 1, 2, \dots, N$, having $2m$ features. So, the dimension of the dictionary matrix \mathbf{D} using bandpower as feature will be $2m \times 4N$ and it is denoted as \mathbf{D}_{BP} and the same dimension remains on using wavelet energy as feature and it is denoted as \mathbf{D}_{WE} .

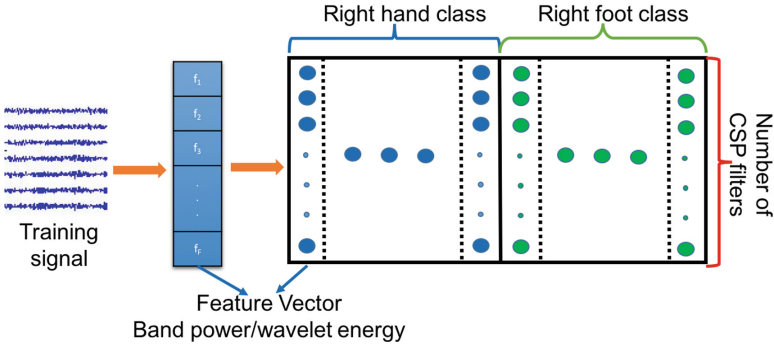


Fig. 4. Two-class dictionary designed for our proposed SFS approach. Each atom in the dictionary is obtained from the training signal of each class and each sub-band.

Sparse Representation: After the construction of dictionary matrix, we have our linear system of equations to get the sparse representation for the input test signal. The test signal is first converted into a feature vector $\mathbf{y} \in \mathbb{R}^{m \times 1}$, using the same way as the columns in dictionary \mathbf{D} is generated. So the input vector can be represented as a linear combination of few columns of \mathbf{D} and it is represented as:

$$\mathbf{y} = \sum_i s_{i,1} \mathbf{d}_{i,1} + s_{i,2} \mathbf{d}_{i,2} + \dots + s_{i,N} \mathbf{d}_{i,N} \quad (6)$$

where $s_{i,j} \in \mathbb{R}, j = 1, 2, \dots, N$ are the sparse coefficients and $i = (1, 2)$ for the two-class MI signals. In matrix form it can be represented as:

$$\mathbf{y} = \mathbf{D}\mathbf{s} \quad (7)$$

where $\mathbf{s} = [s_{i,1}, s_{i,2}, \dots, s_{i,N}]^T$. The objective of the sparse representation is to estimate the scalar coefficients, so that we can sparsely represent the test signal as a linear combination of few atoms of dictionary \mathbf{D} [14]. The sparse representation of an input signal \mathbf{y} can be obtained by performing l_0 norm minimization as follows:

$$\min_{\mathbf{s}} \|\mathbf{s}\|_0 \quad \text{subject to } \mathbf{y} = \mathbf{D}\mathbf{s} \quad (8)$$

l_0 norm optimization gives us the sparse representation but it is an NP-hard problem [2]. Therefore, a good alternative is the l_1 norm which can also be used to obtain sparsity. Recent development tells us that the representation obtained by l_1 norm optimization problem achieves the condition of sparsity and it can be solved in polynomial time [6, 7]. Thus the optimization problem in Eq. (8) becomes:

$$\min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{subject to } \mathbf{y} = \mathbf{D}\mathbf{s} \quad (9)$$

The orthogonal matching pursuit (OMP) is a greedy algorithm used to obtain sparse representation and is one of the oldest greedy algorithms [4]. It employs

the concept of orthogonalization to get orthogonal projections at each iteration and is known to converge in few iterations. For OMP to work in the desired way, all the feature vectors in dictionary \mathbf{D} should be normalized such that $\|\mathbf{D}_i(j)\| = 1$, where $i = (1, 2)$ are the classes and $j = 1, 2, \dots, N$. Using OMP we obtained the sparse representation \mathbf{s} , for the feature vector \mathbf{y} , which will be used further for classifying MI signals.

Sparsity Based Classification: After a successful minimization of sparse representation, the input vector \mathbf{y} will be approximated as a sparse vector which has the same size as the number of atoms in the dictionary \mathbf{D} . Each value of the sparse vector corresponds to the weight given to the corresponding atom of the dictionary. The dictionary is made of equal number of atoms for each class. If for example, there are 1400 atoms in the dictionary for a two-class MI, the first 700 values of the sparse signal tells us the linear relationship between the input vector and the first class i.e. right-hand MI class and so on. Hence, the results of the sparse representation can be used for classification by implying some simple classification rules in the sparse vector \mathbf{s} . In this work, we make use of two classification rules and it is termed as *classifier₁* and *classifier₂*. Mathematically, it is defined as follows:

$$Classifier_1(\mathbf{y}) = \underset{i=1,2}{\operatorname{argmax}} \max(\operatorname{Var}(\mathbf{s}_i)) \quad (10)$$

$$Classifier_2(\mathbf{y}) = \underset{i=1,2}{\operatorname{argmax}} \max(\operatorname{nonzero}(\mathbf{s}_i)) \quad (11)$$

where $\max()$ is a function that returns the maximum value of a vector, the function $\operatorname{Var}()$ is used to find the variance of data and $\operatorname{nonzero}()$ is used to find the number of sparse (non-zero) elements in a vector. The class i is determined, if it has maximum variance or maximum number of non-zero elements.

3 Experimental Results

The performance of the model in our experiment depends on the prediction performance of the classifier. A k -fold cross validation was performed on the dataset to split the entire data into k folds, from which $k - 1$ folds were used to build the dictionary and *one* fold for testing the model. Each fold was used for testing iteratively and the accuracies were calculated. Two different dictionaries were built: one with bandpower features $\mathbf{D}_{\mathbf{BP}}$ and the other with energies of a wavelet transform $\mathbf{D}_{\mathbf{WE}}$. Accuracy of a model based on training and testing test, is a good metric by itself to calculate the performance of the classifier.

3.1 Results of Sparsity Based Classification

We had right-hand and right-foot MI signals that needed to be classified. To illustrate how sparsity plays an important role in our classification, Fig. 5 shows

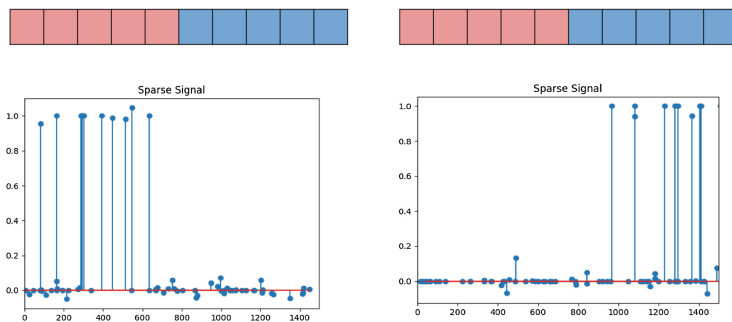


Fig. 5. Sparse representation \mathbf{s} obtained for the two sample test signals. Here, the left figure represents the sparse signal of right-hand class and the right figure for the right-foot class.

the sparse representation of two sample test signals belonging to two different classes using \mathbf{D}_{BP} as dictionary matrix. Here there are around 1400 atoms in the dictionary and so the first 700 elements corresponds to the first class and the rest for the second class. We can clearly see that the sparse representation is classifying the input signal with high accuracies. Table 1 shows the accuracies of each of the two classifiers in k -fold cross validation using the dictionaries \mathbf{D}_{BP} and \mathbf{D}_{WE} , respectively. The result shows that classifier_1 performs better than classifier_2 . It also shows us that the sparsity based classification using the dictionary \mathbf{D}_{WE} outperforms the band-power dictionary \mathbf{D}_{BP} . The normalized and non-normalized confusion matrices of each of the classifiers using dictionary \mathbf{D}_{WE} is given in the Fig. 6.

Table 1. k -fold cross validation accuracies for the classifiers using \mathbf{D}_{BP} and \mathbf{D}_{WE} dictionary.

k -folds		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$	Average
Classifier_1	\mathbf{D}_{BP}	94.64	94.64	96.42	90.47	94.64	91.66	92.85	92.85	93.45	92.26	93.38
	\mathbf{D}_{WE}	98.80	97.02	98.21	94.04	98.80	94.64	94.64	96.42	98.21	95.23	96.60
Classifier_2	\mathbf{D}_{BP}	94.64	95.83	92.26	91.07	93.45	89.88	93.45	91.07	95.64	89.28	92.65
	\mathbf{D}_{WE}	98.21	97.02	96.42	94.04	98.62	93.20	94.64	97.82	97.21	96.20	96.33

3.2 Comparison with SVM and LR

To evaluate the proposed SFS method, we compared our method using \mathbf{D}_{WE} as dictionary with the conventional SVM [17, 18] and LR [13] methods. As classifier_1 gives better accuracy than classifier_2 , it is used for comparison with the conventional methods. For real-time BCI applications, speed of the classifier is an important issue. Hence, CPU execution time is estimated for all the methods. All the classifier algorithms were performed using the same computer

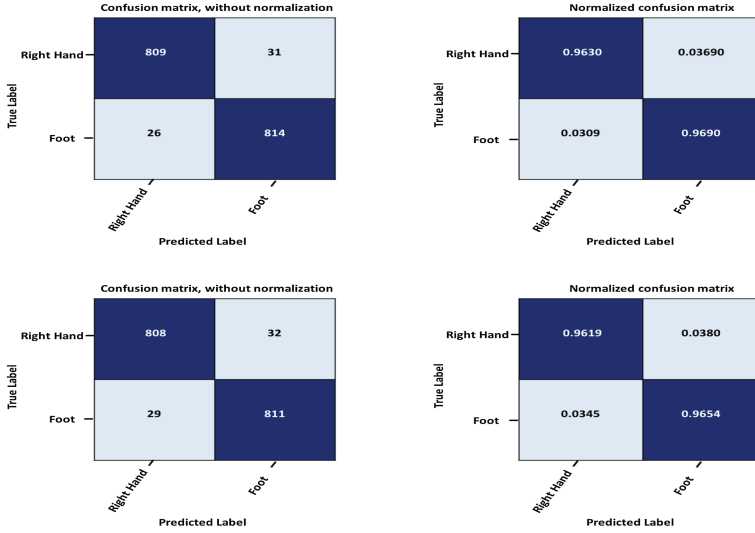


Fig. 6. Confusion matrix of $classifier_1$ and $classifier_2$ using the dictionary D_{WE} .

and same software Python 2.7, making use of Scikit Learn² machine learning package. The accuracies and the CPU execution time obtained for different folds for the proposed SFS method using $classifier_1$ and D_{WE} as dictionary, and the

Table 2. Comparison of k -fold cross-validation accuracy and CPU execution time of various folds for the proposed SFS approach, and the conventional SVM and LR classifier methods.

folds	k -fold cross-validation Accuracy (%)			CPU execution time (Seconds)		
	Proposed SFS	SVM	LR	Proposed SFS	SVM	LR
$k = 1$	98.80	94.22	93.74	24.02	30.62	31.00
$k = 2$	97.02	93.10	92.88	25.42	29.99	29.34
$k = 3$	98.21	93.46	91.79	23.00	28.23	29.21
$k = 4$	94.04	92.78	91.86	24.50	29.29	30.19
$k = 5$	98.80	94.60	94.44	24.32	30.83	28.00
$k = 6$	94.64	91.78	91.32	23.36	29.00	29.75
$k = 7$	94.64	91.90	91.46	24.17	29.35	28.86
$k = 8$	96.42	92.45	91.98	24.56	29.43	29.94
$k = 9$	98.21	93.62	92.85	23.32	29.29	30.32
$k = 10$	95.23	92.88	92.52	23.96	28.00	29.76
Average	96.60	93.08	92.48	24.06	29.40	29.64

² <http://scikit-learn.org>.

conventional SVM and LR are listed in Table 2. The average values obtained indicates that the proposed SFS method delivers high average classification accuracy and lesser execution time than the SVM and LR methods. Since the proposed method executes in lesser time with higher accuracy, it can be further used to build real-time MI-based BCI applications for motor disabled people.

4 Conclusion

In this work, we used a new spatial filtering based sparsity (SFS) approach to classify two-class MI-based EEG signals for BCI applications. Firstly, the EEG signal with 118 channels are of high-dimension. To reduce the computational complexity, constraints are applied on selecting channels. Secondly, to better discriminate the MI classes, two sub-bands of band-pass filter between 7–13 Hz and 13–30 Hz are applied to the selected number of channels followed by CSP filtering. Thirdly, it is important to note that EEG signals produce variations among users at different sessions. As SFS method requires a dictionary matrix, it is designed using the bandpower and wavelet features obtained from the spatially filtered signals. This dictionary matrix helps us to overcome the inter-subject variability problem. This method also reduces the computational complexity significantly and increases the speed and accuracy of the BCI system. Hence, the proposed SFS approach can be served to design a more robust and reliable MI-based real-time BCI applications like text-entry system, gaming, wheel-chair control, etc., for motor impaired people. Future work will focus on extending the sparsity approach for classifying multi-class MI tasks which can be further used for communication purpose.

References

1. Aghaei, A.S., Mahanta, M.S., Plataniotis, K.N.: Separable common spatio-spectral patterns for motor imagery BCI systems. *IEEE Trans. Biomed. Eng.* **63**(1), 15–29 (2016)
2. Baraniuk, R.G.: Compressive sensing [lecture notes]. *IEEE Signal Process. Mag.* **24**(4), 118–121 (2007)
3. Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Muller, K.R.: Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process. Mag.* **25**(1), 41–56 (2008)
4. Cai, T.T., Wang, L.: Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Trans. Inf. Theory* **57**(7), 4680–4688 (2011)
5. Candès, E.J., Wakin, M.B.: An introduction to compressive sampling. *IEEE Signal Process. Mag.* **25**(2), 21–30 (2008)
6. Candès, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted L1 minimization. *J. Fourier Anal. Appl.* **14**(5), 877–905 (2008)
7. Donoho, D.L.: For most large underdetermined systems of linear equations the minimal L1-norm solution is also the sparsest solution. *Commun. Pure Appl. Math.* **59**(6), 797–829 (2006)

8. Donoho, D.L., Tsaig, Y., Drori, I., Starck, J.L.: Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **58**(2), 1094–1121 (2012)
9. Gemmeke, J.F., Virtanen, T., Hurmalainen, A.: Exemplar-based sparse representations for noise robust automatic speech recognition. *IEEE Trans. Audio Speech Lang. Process.* **19**(7), 2067–2080 (2011)
10. He, B., Baxter, B., Edelman, B.J., Cline, C.C., Wenjing, W.Y.: Noninvasive brain-computer interfaces based on sensorimotor rhythms. *Proc. IEEE* **103**(6), 907–925 (2015)
11. He, L., Hu, D., Wan, M., Wen, Y., von Deneen, K.M., Zhou, M.: Common bayesian network for classification of EEG-based multiclass motor imagery BCI. *IEEE Trans. Syst. Man Cybern. Syst.* **46**(6), 843–854 (2016)
12. Higashi, H., Tanaka, T.: Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification. *IEEE Trans. Biomed. Eng.* **60**(4), 1100–1110 (2013)
13. Li, Y., Wen, P.P., et al.: Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain-computer interface. *Comput. Methods Programs Biomed.* **113**(3), 767–780 (2014)
14. Li, Y., Yu, Z.L., Bi, N., Xu, Y., Gu, Z., Amari, S.: Sparse representation for brain signal processing: a tutorial on methods and applications. *IEEE Signal Process. Mag.* **31**(3), 96–106 (2014)
15. Nicolas-Alonso, L.F., Corralejo, R., Gomez-Pilar, J., Álvarez, D., Hornero, R.: Adaptive stacked generalization for multiclass motor imagery-based brain computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* **23**(4), 702–712 (2015)
16. Pfurtscheller, G., Neuper, C.: Motor imagery and direct brain-computer communication. *Proc. IEEE* **89**(7), 1123–1134 (2001)
17. Siuly, S., Li, Y.: Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* **20**(4), 526–538 (2012)
18. Sreeja, S., Joshi, V., Samima, S., Saha, A., Rabha, J., Cheema, B.S., Samanta, D., Mitra, P.: BCI augmented text entry mechanism for people with special needs. In: Basu, A., Das, S., Horain, P., Bhattacharya, S. (eds.) *IHCI 2016. LNCS*, vol. 10127, pp. 81–93. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52503-7_7
19. Tam, W.K., Tong, K., Meng, F., Gao, S.: A minimal set of electrodes for motor imagery BCI to control an assistive device in chronic stroke subjects a multi-session study. *IEEE Trans. Neural Syst. Rehabil. Eng.* **19**(6), 617–627 (2011)
20. Wolpaw, J., Wolpaw, E.W.: *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press, Oxford (2012)
21. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* **113**(6), 767–791 (2002)
22. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

