Motor Imagery EEG Signal Processing and Classification using Machine Learning Approach

Sreeja. S. R[1], Joytirmoy Rabha, K. Y. Nagarjuna, Debasis Samanta[2], Pabitra Mitra[3], Monalisa Sarma[4]
Indian Institute of Technology, Kharagpur
West Bengal, India-721302
[1]sreejasr@iitkgp.ac.in, [2]dsamanta@iitkgp.ac.in, [3]pabitra@cse.iitkgp.ernet.in, [4]monalisa@iitkgp.ac.in

***Abstract*—Motor imagery (MI) signals recorded via electroencephalography (EEG) is the most convenient basis for designing brain-computer interfaces (BCIs). As MI based BCI provides high degree of freedom, it helps motor disabled people to communicate with the device by performing sequence of MI tasks. But inter-subject variability, extracting user-specific features and increasing accuracy of the classifier is still a challenging task in MI based BCIs. In this work, we propose an approach to overcome the above mentioned issues. The proposed approach follows the pipeline such as channel selection, band-pass filter based CSP (common spatial pattern), feature extraction, feature selection using two different techniques and modeling using Gaussian Naïve Bayes (GNB) classifier. Since the optimal features are selected by feature selection techniques, it helps to overcome inter-subject variability and improves performance of GNB classifier. To the best of our knowledge, the proposed methodology has not been used for MI-based BCI applications. The proposed approach is validated using BCI competition III dataset IVa. The result of our proposed approach is compared with two conventional classifiers such as linear discriminant analysis (LDA) and support vector machine (SVM). The results prove that the proposed method provides an improved accuracy than LDA and SVM classifiers. The proposed method can be further developed to design a reliable and real-time MI-based BCI application.**

***Index Terms*—Motor Imagery, BCI, EEG, feature extraction, feature selection, machine learning.**

I. Introduction

BRAIN-COMPUTER Interfaces (BCIs) provides a direct connection between the human brain and a computer [1]. 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 [2]. The interpreted brain activities are direcly 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 functionl magnetic response imaging (fMRI) or magnetoencephalogram (MEG), but the most common BCIs are based on electroencephalogram (EEG). EEG-based BCIs facilitates many real-time applications due to its affordable cost and ease of use [3].

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 [4]. 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 [5]. These rhythms are readily measurable in both healthy and disabled people with neuromuscular injuries. Upon executing real or imaginary motor movements, it causes amplitude supression or enhancement of mu rhythm and these phenomena are called event-related desynchronization (ERD) and event-related synchronization (ERS), respectively [5]. Traditional BCIs rely on this neurophysiological phenomenon to determine whether the user is performing a motor task or not. As the dynamics of brain potentials associated with MI tasks can form spatio-temporal patterns, the *common spatial pattern* (CSP) [6] is a highly successful algorithm to extract relevant MI features. This algorithm is designed to capture the spatial projections of ERD/ERS in such a way that the power ratio differs greatly between two classes. Several variants of CSP have been devised such as common spatial spectral pattern (CSSP) [7], spectrally weighted common spatial pattern [8], iterative spatio-spectral patterns learning (ISSPL) [9], filter bank common spatial pattern [10], augmented complex common spatial pattern [11] and and separable common spatio-spectral patterns (SCSSP) [12].

Collecting EEG-based MI data is a tedious, time consuming process. Processing with entire EEG data delays the system and affects the accuracy of the classifier. It is also important to note that, for the same user the observed patterns differ from one day to another, or from session to session [1]. This inter-personal variability of EEG signals results in degraded performance of classifier. From the literature it is observed that many features like statistical [13], time-domain [14], frequency-domain [15], wavelet [16], auto-regressive coefficients [17] have been extracted from MI-based EEG signals. It is still a question whether the extracted features are subject-specific optimal features or not. Apart from this issue, for real-time applications the ongoing motor imagery events have to be detected and classified continuously into a control command as accurately and quickly as possible. The above issues motivate us to lay down our research objectives as follows: selecting channels considering the motor areas; addressing inter-personal variability; extracting the set of highly discriminant user-specific features; increasing the speed and accuracy of the classifier in MI-based BCI system.

As it is proven that the channels present near the active

978-1-5386-0527-1/17 $31.00 © 2017 IEEE
DOI 10.1109/ICTCS.2017.15

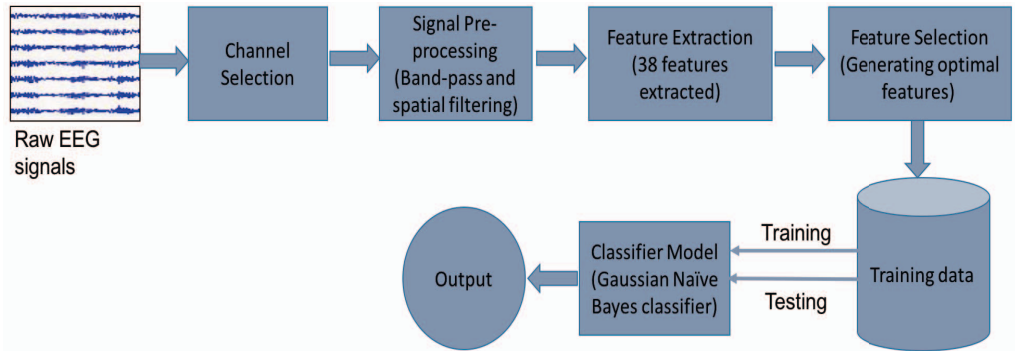

Fig. 1: Framework of the proposed approach

regions of the brain have more relevant information [13], it is best to consider those channels for further processing. To provide subject-specific optimal features, two different feature selection methodologies such as minimum redundancy maximum relevancy (mRMR) method and Lasso Regularization based feature selection method are studied and observed. In the same way, in literatures different classifiers have been applied to classify two class EEG-based MI tasks with different features [18]. In this work, Gaussian Naïve Bayes (GNB) classifier is modeled using the selected optimal features. To the best of our knowledge, band-pass filter based CSP, feature selection method and GNB model have not been used together for MI-based BCI applications.

The framework of the proposed EEG-based MI BCI system is shown in Fig. 1. The following aspects can be highlighted on the proposed EEG-based MI BCI system.

1) We selected only few electrodes present over the motor areas for processing.
2) Band-pass filter based CSP is applied to the selected EEG channels to spatially filter the signals.
3) The possible number of features are extracted from the spatially filtered data.
4) The most discriminant user-specific features using two different feature selection methods are observed.
5) The selected features are modeled using Gaussian Naïve Bayes classifier.
6) We compared the experimental results of the proposed method with two conventional classifiers, namely LDA and SVM, in terms of accuracy.

Our paper is organised as follows. In Section II, we present description of the data and the proposed technique in details. In Section III, the experimental results and performance evaluation are presented. Finally, conclusions and future work are outlined in Section IV.

II. 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, feature extraction, feature selection and classification of EEG-based MI data is discussed in detail.

A. Experimental Data

We used the publicly available dataset IVa from BCI competition III[1] 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 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.

B. EEG Signal Pre-processing

Channel Selection: 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 [19]. 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 [20]. 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.

Band-pass filtering: Since the original sampling rate of the EEG signal is 1000 Hz, it is downsampled to 100 Hz for further processing. Then the selected 30 channel EEG data are again passed through a band-pass filter between 7 - 30 Hz, as it is known from [5]. that mu (μ) and beta (β) rhythms lie within that frequency range. Then data segmentation is done where we used two second samples (200 samples) after the display of cue of each trial. Each segmented data is called as an epoch.

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

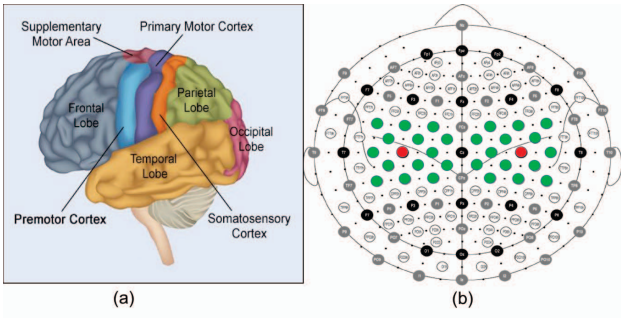

Fig. 2: (a) Motor cortex of the brain (b) Standard 10±20 system of electrode placement for 128 channel EEG system. The electrodes in green and red colour are selected for processing.

Spatial filtering: CSP is one of the most commonly used spatial filter in building MI based BCIs [6]–[12]. The signals which are segmented into two second time samples are spatially filtered using a CSP filter. CSP aims to find the linear transforms or spatial filters which maximizes the variance of one class while minimizing it for the other class. How CSP is applied to the given dataset is explained here.

Let $\mathbf{X_H}$ and $\mathbf{X_F}$ 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

$$\mathbf{X}_i^{CSP} = \mathbf{W}^T \mathbf{X}_i \tag{1}$$

where $i \in \{H, F\}$ 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:

$$\underset{w}{max} \left(\frac{\mathbf{w}^T C_H \mathbf{w}}{\mathbf{w}^T C_F \mathbf{w}} \right) \tag{2}$$

where $C_H = \mathbf{X_H X_H^T}$ and $C_F = \mathbf{X_F X_F^T}$. Inorder to make the computation easier to find $\mathbf{w}$, we computed $\mathbf{X_H}$ and $\mathbf{X_F}$ by taking the average of all epochs of each class. Equation (2) can be written as minimization problem as follows:

$$\underset{w}{min} \left(-\mathbf{w}^T C_H \mathbf{w} \right) \quad subject\ to \quad \mathbf{w}^T C_F \mathbf{w} = 1 \tag{3}$$

Solving the above equation using Lagrangian method, we finally have the resulting equation as:

$$C_H \mathbf{w} = \lambda C_F \mathbf{w} \tag{4}$$

Thus equation (2) becomes **eigenvalue decomposition problem**, where λ is the eigenvalue corresponds to the eigenvector $\mathbf{w}$, obtained by solving the following equation:

$$(C_H - \lambda C_F)\mathbf{w} = \mathbf{0} \tag{5}$$

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_H have the smallest eigenvalues for C_F. Since we used 30 EEG channels, we will have 30 eigenvalues and correspondingly 30 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 $\mathbf{W_{CSP}}$.

$$\mathbf{W_{CSP}} = [\mathbf{w_1}, \mathbf{w_2}, ..., \mathbf{w_m}, \mathbf{w_{c-m+1}}, ..., \mathbf{w_c}] \in \mathbb{R}^{2m \times c} \tag{6}$$

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

$$\begin{aligned} \mathbf{X}_H^{CSP} \in \mathbb{R}^{2m \times n} &:= \mathbf{W}_{CSP}^T \mathbf{X}_H \\ \mathbf{X}_F^{CSP} \in \mathbb{R}^{2m \times n} &:= \mathbf{W}_{CSP}^T \mathbf{X}_F \end{aligned} \tag{7}$$

The CSP filters can be plotted back to see the activations of various regions of the brain. Fig. 3 shows the scalp plot where the first 4 and last 4 magnitudes of the coefficients of the CSP filter is plotted. The dark red colour indicates the highest significance. The upper-left plot indicates the filter $\mathbf{w_1}$ and the down-right plot indicates the last filter $\mathbf{w_n}$.

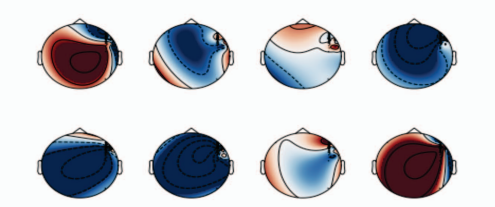

Fig. 3: Colormap of magnitutdes of the coefficients of CSP filter are projected on the scalp.

C. Feature extraction

The band-pass based spatially filtered signals are still of high dimension. To reduce the complexity of working with such high dimensional signals, we need to pull out some special features from the spatially filtered data. These features must maximize the discriminability between the two MI classes. A set of 38 features are extracted from each spatially filtered epoch. The extracted feature vectors are listed below.

1) Statistical Features: We extracted seven statistical features, namely, mean, median, standard deviation, skewness, kurtosis, maximum and minimum. These features describes the distribution of EEG signals in terms of amplitudes and moments [13].

2) Time-domain Features: Time-domain features [14] captures the temporal information of signals. As EEG is known to have a good temporal locality, we extracted a number of time-domain features, namely, hjorth parameters (ability, mobility and complexity), 1st difference mean and maximum, 2nd difference mean and maximum, mean and variance of vertex to vertex slope, mean and variance of vertex to vertex amplitudes, zero crossing, and coefficient of Variation.

3) Frequency-domain Features: The frequency-domain features [15] captures the frequency information of brain rhythms during motor imagery tasks. It is known and as stated earlier the frequency of motor imagery signals lies within 7-30 Hz.

We converted the time domain signals to frequency domain using Fourier transform and then selected two sub-bands of α (7-13 Hz) and β (13-30 Hz). For each of these sub-bands we calculated the band power and max-power.

4) Wavelet-based Features: Wavelet transform [16] is a spectral estimation technique in which any general function can be expressed as an infinite series of wavelets. The decomposition of the signal leads to a set of coefficients called wavelet coefficients. In our work we used the Discrete Wavelet Transform (DWT) which employs two functions, namely, scaling function and wavelet function. The DWT gives rise to two coefficients D_i and A_i which are the down sampled outputs of the high pass and low pass filters at each decomposition level. The extracted wavelet-based features are mean, standard deviation, energy and entropy of both D_i and A_i. We used coif1 wavelet as it gives best result among other wavelets [21].

5) Auto-regressive coefficients: Auto-regressive (AR) method [17] models the signal at any given time, as a weighted sum of signals at previous time and some noise. We implemented AR model of order 6 using the Burg's Algorithm and used the coefficients as features. Mathematically, it can be formulated as:

$$X(t) = a_1 X(t_1) + a_2 X(t_2) + ... + a_p X(t_p) + E_t \quad (8)$$

where, $X(t)$ is the measured signal at time t, E_t is the noise term and a_1 to a_p are the auto-regressive parameters.

Therefore, we extracted a set of N (38) features from each spatially filtered epoch i.e., $F = \{f_1, f_2, ..., f_N\}$ where $f_1 \in \mathbb{R}^{2k}$ and $F \in \mathbb{R}^{2k \times N}$. The extracted feature vectors are then normalized to reduce inter and intra-subject variability. Then, out of these normalized features, the discriminative subset of features have to be identified for a reliable classification.

D. Feature Selection

Feature selection approaches aims to select a small set of features S with dimension m, that is, $S = \{s_1, s_2, ..., s_m\}$ from a feature set $F = \{f_1, f_2, ..., f_N\}$, where $m \leq N$, and $S \subseteq F$. Reducing the number of irrelevant features will drastically improves the learning performance, lowers the computational complexity, and decreases the required storage. In this section, we have exercised two feature selection algorithms over the above normalized feature vectors.

Minimum-redundancy and Maximum-relevance (mRMR): As the name suggest this feature selection algorithm is based on selecting features with minimum redundancy and maximum relevancy depending on the mutual information values between various features [22]. Thus it involves selecting the feature S with the highest relevance to the target class C, based on mutual information, such that $I(S;C)$. Mathematically, it is defined as
Maximum relevance:

$$\max D, \qquad D = \frac{1}{|S|} \sum_{f_i \in S} I(f_i; C) \quad (9)$$

Minimum redundancy:

$$\min R, \qquad R = \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i, f_j) \quad (10)$$

where $I(f_i, f_j)$ is mutual information between the feature f_i and f_j, $|S|$ is the cardinality of the set S and C is the target class. The criteria of combining the above two constraints is called minimum-redundancy maximum-relevancy (mRMR) and it is given as

$$\max \Phi, \qquad \Phi = D - R \quad (11)$$

The features are sorted according to mRMR and the first three features are selected as optimal features. The selected features are given in Table I.

Lasso Regularization based Feature Selection: In regularization models, classifier induction and feature selection are simultaneously achieved by minimizing fitting errors and properly tuning penalities. The learned classifier w can have coefficients to be very small or to be zero. Feature selection is achieved by selecting the non-zero coefficients in w [23]. Mathematically, it is defined as

$$\hat{w} = \min_{w} \; c(w, X) + \alpha \, \|w\| \quad (12)$$

where $c(w, X)$ is the objective function of the classifier, α is the regularization parameter, and $\|w\|$ is a regularization (penalty) term. The model penalised with l_1 norm is called as Lasso regularization and defined as

$$\|w\| = \sum_{i=1}^{m} w_i \quad (13)$$

This model will have a sparse solution, such that it forces weak features to have zero as coefficients and they are excluded from the model. Thus Lasso (l_1) regularization inherently performs feature selection [24]. The selected features are listed in Table I.

In both feature selection methods, we selected the first three features. Then the common features between mRMR and lasso regularization are selected to trian the classification model, that is, *optimal features = features selected by (mRMR ∩ lasso regularization)*. Therefore, band-power (13-30 Hz) and wavelet energy of each of the spatially filtered epoch has been selected as the optimal features. These features are tuned by k-fold cross validation to create training and testing set.

TABLE I: Feature selection methods and the selected optimal features.

Feature Selection Method	Selected Features
Minimum-redundancy Maximum-relevancy (mRMR)	Band-power (13-30 Hz), wavelet energy, kurtosis
Lasso Regularization	Band-power (13-30 Hz), wavelet energy, AR with 6 coefficients,

E. Gaussian Naïve Bayes (GNB) Classifier

The Naïve Bayes theorem aims at assigning the class C_i to the feature vector by calculating the *a posteriori* probability of the feature vector [25]. Here $i \in \{H, F\}$, H denotes right-hand and F denotes right-foot MI class. Mathematically, it is defined as

$$p(C_i|S) = \frac{p(S|C_i) \times p(C_i)}{p(S)} \quad (14)$$

where C_i is the class, $S = \{s_1, s_2, ..., s_m\}$ is the set of selected optimal features. By assuming gaussian distribution, Naïve Bayes can be extended as gaussian Naïve Bayes. Gaussian distribution is easy to work with because only mean and variance need to be calculated from the taining data [26]. Let μ_{jH} and σ^2_{jH} be the mean and variance value of the feature vector s_j associated with class C_H. Then the class-conditional probability using gaussian normal distribution is defined as

$$p(S = s_j|C_H) = \frac{1}{\sqrt{2\pi\sigma^2_{jH}}} e^{-\frac{(s_j-\mu_{jH})^2}{2\sigma^2_{jH}}} \quad (15)$$

The prediction result provide the class and it is defined as

$$C_{pred} = \underset{i}{argmax} \;\; p\left(C_i \mid s_1, s_2, ..., s_m\right) \quad (16)$$

In this study, GNB classifier is used to classify two-class MI signals. The main advantage of this classifier is there is no parameter tuning like other classifiers. The parameters in GNB classifier is automatically calculated by maximum likelihood estimation (MLE) [25].

F. Performance measures

The performance of the proposed method was evaluated using the following measures.

1) Confusion matrix: The confusion matrix is a useful tool for analysing how frequently instances of class say X were correctly classified as class X. Having m classes, confusion matrix is a table of size $m \times m$. An entry at (i, j) indicates the number of instances of class i that were labeled by the classifier as class j. Here, right-hand should be classified as right-hand class. Thus the number of true positives (TP), false negatives (FN), false positives (FP) and ture negatives (TN) are obtained. For a classifier to have a good accuracy, ideally most of the diagonal entries (TP, TN) should have large values with the rest of entries being less or close to zero.

2) Accuracy: The accuracy is the ratio of total number of correctly classified samples to the total number of samples of all classes.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100 \quad (17)$$

3) κ score: The κ-score or κ-coefficient is a statistical measure that compares an observed accuracy (p_o) with an expected accuarcy (p_e). The κ-score is given by the following equation

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \quad (18)$$

$\kappa = 1$ indicates complete agreement between the MI BCI classes while $\kappa \leq 0$ means there is no agreement at all.

III. Experiments and Results

In this section, the implementation procedure and the experimental results of the proposed method using the dataset IVa of BCI competition III was explained. In this study, the codes were written in Python 2.7 making use of Scikit Learn[2], a popular machine learning library.

[2]http://scikit-learn.org

A. Performance of the proposed method

The given dataset had two MI classes, right-hand and right-foot to be classified. We experimented the proposed model with two different number of channels. One experiment considering all the 118 channles and the other considering only the 30 channels present over the motor cortex. The optimal feautres selected from the dataset after applying band-pass filter based CSP is given as input to the GNB model. Table II shows the k-fold cross-validation accuracy of the proposed method for each subject taking 118 and 30 electrodes respectively. Here, the value of k is taken as 10 as it is the common choice. From the values obtained it is observed that the reduced number of channels gives better accuracy than considering all the 118 channels.

TABLE II: Classification accuracy by k-fold cross-validation method for the proposed method.

Subjects	k-fold cross validation accuracy (%) (mean)	
	118 channels	30 channels
aa	92.42	94.23
al	90.60	92.36
av	90.58	93.75
aw	91.79	93.56
ay	94.45	96.98
Average	91.96	94.17

Since we got better accuracy with 30 channles we proceeded the experiment with the reduced number of channels. Inorder to make the evaluation easier, after selecting the optimal features, the training and testing data is created considering all the subject's data together. GNB classifier is build based on the training data and validated using the testing data. Table III shows the various performance measure achieved by the proposed method for various folds.

TABLE III: Values of TP, FP, TN, FN of confusion matrix, accuracy (%) and κ score of the proposed method of various folds.

k-folds	Confusion Matrix				Accuracy (%)	κ score
	TP	FN	FP	TN		
k=1	800	43	39	798	95.11	0.90
k=5	807	39	32	802	95.77	0.92
k=10	805	40	34	801	95.59	0.91
Average	**804**	**41**	**35**	**800**	**95.47**	**0.91**

B. Performance comparison with the conventional methods

The performance of the proposed approach is compared with two conventional calssifiers such as linear discriminant analysis (LDA) [27] and support vector machine (SVM) [28]. LDA and SVM are the most widely used classifiers in MI based BCI systems. The same optimal features selected to train GNB classifier are used to train these classifiers. The parameters for SVM classifier need to be choosen carefully to avoid under-fitting and over-fitting problems. The output obtained by LDA and SVM classifiers are evaluated and compared with the proposed GNB approach. The performance metrics were generated for three different folds and the average values obtained are shown in the Table IV. The result shows

that the proposed method provides an improved accuracy than the conventional methods.

TABLE IV: Comparison of performance metrics of the proposed method, LDA and SVM classifiers.

Methods	Confusion Matrix				Accuracy (%)	κ value
	TP	FN	FP	TN		
LDA	749	75	74	782	91.10	0.82
SVM	763	64	66	787	92.26	0.85
Proposed GNB	804	41	35	800	95.47	0.91

IV. Conclusion

In this work, we used a new combination of machine learning approach to classify two-class MI 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, it is important to note that the EEG signals produce variations among users at different sessions. This inter-subject variability is removed using two different feature selection techniques, namely mRMR and Lasso regularization. Our results proves that the performance of two-class MI-based BCI can be significantly improved using few number of channels and few feature vectors. This method also reduces the computational complexity significantly and increases the speed and accuracy of the classifier models. GNB classifier performed better than the conventional classifiers on following the above techniques. Hence, the proposed 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 proposed approach for classifying multi-class MI tasks which can be further used for communication purpose.

References

[1] J. Wolpaw and E. W. Wolpaw, *Brain-computer interfaces: principles and practice*. Oxford University Press, USA, 2012.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[3] S. R. Sreeja, V. Joshi, S. Samima, A. Saha, J. Rabha, B. S. Cheema, D. Samanta, and P. Mitra, "BCI augmented text entry mechanism for people with special needs," in *8th International Conference on Intelligent Human Computer Interaction, IHCI*, 2016, pp. 81–93.

[4] B. He, B. Baxter, B. J. Edelman, C. C. Cline, and W. Y. Wenjing, "Non-invasive brain-computer interfaces based on sensorimotor rhythms," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 907–925, 2015.

[5] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[6] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[7] S. Lemm, B. Blankertz, G. Curio, and K.-R. Muller, "Spatio-spectral filters for improving the classification of single trial eeg," *IEEE transactions on biomedical engineering*, vol. 52, no. 9, pp. 1541–1548, 2005.

[8] R. Tomioka, G. Dornhege, G. Nolte, B. Blankertz, K. Aihara, and K.-R. Müller, "Spectrally weighted common spatial pattern algorithm for single trial eeg classification," *Department of Mathematical Informatics, University of Tokyo, Tokyo, Japan, Tech. Rep*, vol. 40, 2006.

[9] W. Wu, X. Gao, B. Hong, and S. Gao, "Classifying single-trial eeg during motor imagery by iterative spatio-spectral patterns learning (isspl)," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 6, pp. 1733–1743, 2008.

[10] H. Higashi and T. Tanaka, "Simultaneous design of fir filter banks and spatial patterns for eeg signal classification," *IEEE transactions on biomedical engineering*, vol. 60, no. 4, pp. 1100–1110, 2013.

[11] C. Park, C. C. Took, and D. P. Mandic, "Augmented complex common spatial patterns for classification of noncircular eeg from motor imagery tasks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 1–10, 2014.

[12] A. S. Aghaei, M. S. Mahanta, and K. N. Plataniotis, "Separable common spatio-spectral patterns for motor imagery bci systems," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 1, pp. 15–29, 2016.

[13] Y. Li, P. P. Wen *et al.*, "Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain–computer interface," *Computer methods and programs in biomedicine*, vol. 113, no. 3, pp. 767–780, 2014.

[14] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, "Time domain parameters as a feature for eeg-based brain–computer interfaces," *Neural Networks*, vol. 22, no. 9, pp. 1313–1319, 2009.

[15] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 4, pp. 317–326, 2008.

[16] W.-Y. Hsu and Y.-N. Sun, "Eeg-based motor imagery analysis using weighted wavelet transform features," *Journal of neuroscience methods*, vol. 176, no. 2, pp. 310–318, 2009.

[17] M. Tavakolan, X. Yong, X. Zhang, and C. Menon, "Classification scheme for arm motor imagery," *Journal of medical and biological engineering*, vol. 36, no. 1, pp. 12–21, 2016.

[18] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 4, no. 2, p. R1, 2007.

[19] L. He, D. Hu, M. Wan, Y. Wen, K. M. von Deneen, and M. Zhou, "Common bayesian network for classification of eeg-based multiclass motor imagery bci," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 6, pp. 843–854, 2016.

[20] W.-K. Tam, K.-y. Tong, F. Meng, and S. Gao, "A minimal set of electrodes for motor imagery BCI to control an assistive device in chronic stroke subjects: a multi-session study," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, no. 6, pp. 617–627, 2011.

[21] M. H. Alomari, E. A. Awada, A. Samaha, and K. Alkamha, "Wavelet-based feature extraction for the analysis of eeg signals associated with imagined fists and feet movements," *Computer and Information Science*, vol. 7, no. 2, p. 17, 2014.

[22] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[23] V. Fonti and E. Belitser, "Feature selection using lasso," 2017.

[24] P. Zhao and B. Yu, "On model selection consistency of lasso," *Journal of Machine learning research*, vol. 7, no. Nov, pp. 2541–2563, 2006.

[25] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[26] K. S. Kassam, A. R. Markey, V. L. Cherkassky, G. Loewenstein, and M. A. Just, "Identifying emotions on the basis of neural activation," *PloS one*, vol. 8, no. 6, p. e66032, 2013.

[27] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust eeg single-trial analysis," *IEEE Signal processing magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[28] S. Siuly and Y. Li, "Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain–computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 4, pp. 526–538, 2012.