Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 04, No. 02, August 2018.

MOTOR IMAGERY EEG SIGNAL PROCESSING AND CLASSIFICATION USING MACHINE LEARNING APPROACH

S. R. Sreeja[1], Debasis Samanta[1], Pabitra Mitra[1] and Monalisa Sarma[2]

(Received: 6-Dec.-2017, Revised: 5-Mar.-2018, Accepted: 10-Mar.-2018)

ABSTRACT

Typically, people with severe motor disabilities have limited opportunities to socialize. Brain-Computer Interfaces (BCIs) can be seen as a hope of restoring freedom to immobilized individuals. Motor imagery (MI) signals recorded via electroencephalograms (EEGs) are the most convenient basis for designing BCIs as they provide a high degree of freedom. MI-based BCIs help motor disabled people to interact with any real-time BCI applications by performing a sequence of MI tasks. But, inter-subject variability, extracting user-specific features and increasing accuracy of the classifier are still a challenging task in MI-based BCIs. In this work, we propose an approach to overcome the above-mentioned issues. The proposed approach considers channel selection, band-pass filter based common spatial pattern, feature extraction, feature selection and modeling using Gaussian Naïve Bayes (GNB) classifier. Since the optimal features are selected by feature selection techniques, they help overcome inter-subject variability and improve 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 has been validated using BCI competition III dataset IVa. The result of our approach has been compared with those of two classifiers; namely, Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM). The results prove that the proposed method provides an improved accuracy over LDA and SVM classifiers. The proposed method can be further developed to design reliable and real-time MI-based BCI applications.

KEYWORDS

Motor imagery, Brain computer interface, Electroencephalography, Feature extraction, Feature selection, Machine learning.

1. INTRODUCTION

Brain-Computer Interfaces (BCIs) provide a direct connection between the human brain and a computer [1]. BCIs capture neural activities associated with external stimuli or mental tasks, without any involvement of nerves and muscles and provide an alternative non-muscular communication [2]. The interpreted brain activities are directly translated into a sequence of commands to carry out specific tasks, such as controlling wheel chairs, home appliances, robotic arms, speech synthesizers, computers and gaming applications. Brain activities can be measured through invasive and non-invasive devices. So far, a wide research has been carried out on human behaviour and task classification using invasive techniques, such as electrocorticography (ECoG) [3] and Local Field Potentials (LFPs) [4]-[5]. In ECoG, the electrodes are placed directly on the exposed surface of the brain to record electrical activity and LFP refers to the electrical field recorded using a small-sized electrode in the extracellular space of brain tissue. These techniques involve surgery and are risky. A non-invasive BCI uses brain activities recorded from an electroencephalogram (EEG), functional Magnetic Response Image (fMRI) or magnetoencephalogram (MEG), …etc. Among the available non-invasive devices, EEG-based BCIs facilitate many real-time applications, as they satisfy convenience criteria (non-intrusive, non-obtrusive and simple) and effectiveness criteria (sensitive, efficient and compatible) [6].

This paper is an extended version of a short paper that was presented at the International Conference "New Trends in Computing Sciences (ICTCS) 2017", 11-13 October 2017, Amman, Jordan.

1. S. R. Sreeja, Debasis Samanta and Pabitra Mitra are with Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, West Bengal, India. Emails: sreejasr@iitkgp.ac.in, dsamanta@iitkgp.ac.in, pabitra@cse.iitkgp.ernet.in.
2. Monalisa Sarma is with Subir Chowdhury School of Quality and Reliability, Indian Institute of Technology Kharagpur, West Bengal, India. Email: monalisa@iitkgp.ac.in.

EEG-based BCI systems are mostly built using Visually Evoked Potentials (VEPs), Event-Related Potentials (ERPs) [7], Slow Cortical Potentials (SCPs) and Sensori-Motor Rhythms (SMRs). Out of these potentials, SMR-based BCI provides a high degree of freedom in association with real and imaginary movements of hands, arms, feet and tongue [8]. 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 [9]. These rhythms are readily measurable in both healthy and disabled people with neuromuscular injuries. Upon executing real or imaginary motor movements, amplitude suppression or enhancement of μ rhythm is caused and these phenomena are called Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS), respectively [9]. Traditional BCIs rely on these neurophysiological phenomena 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) [10] 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) [11], spectrally weighted common spatial pattern [12], Iterative Spatio-Spectral Pattern Learning (ISSPL) [13], Filter Bank Common Spatial Pattern (FBCSP) [14], augmented complex common spatial pattern [15], Separable Common Spatio-Spectral Pattern (SCSSP) [16] and self-adaptive CSP [17].

Collecting EEG-based MI data is a tedious and 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 the classifier. From the literature, it is observed that many features, like statistical [18], time-domain [19], frequency-domain [20], wavelet [21], auto-regressive coefficients [22] 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 regions of the brain have more relevant information [18], it is best to consider those channels for further processing. Few research on feature optimization for MI-based BCI [23] and Steady-State VEP (SSVEP)-based BCI [24] has been carried out recently. In this work, 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. In the same way, in literature, different classifiers have been applied to classify EEG-based MI tasks with different features [25]. Recent research on classification and pattern recognition shows that a Bayesian classifier produces enhanced results than the existing classifiers [26]– [29]. Hence, in this work, Gaussian Naïve Bayes (GNB) classifier is used and modeled using the selected optimal features.

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

1) The channels present over the motor areas are selected 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 is 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 those of two classifiers; namely, LDA and SVM, in terms of accuracy and time.

In order to make our proposed method more suitable for real-time classification, constraints are

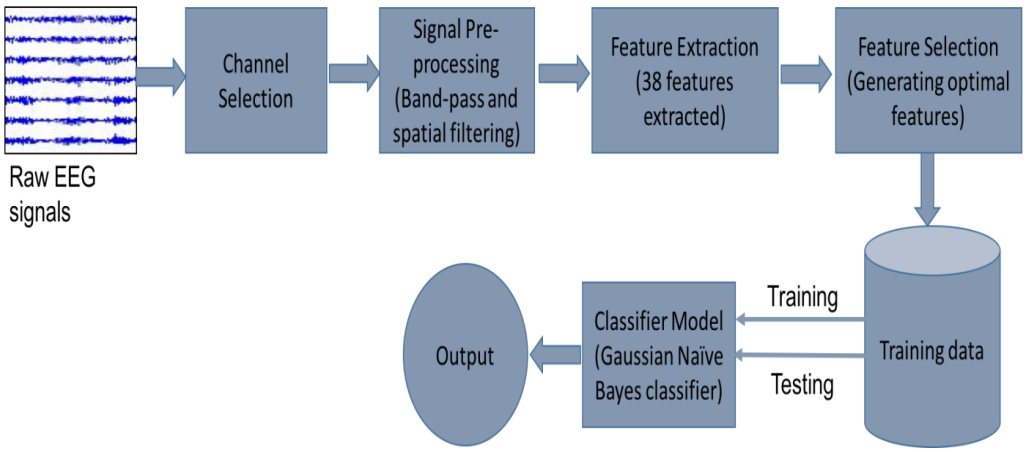

Figure 1. Framework of the proposed approach.

applied on selecting channels and on selecting features. Normally, EEG signals suffer from inter-and intra-subject variability. In our proposed method, the selected features are normalized using z-score normalization, which makes the features lie in a certain range and reduces the variability between different sessions and subjects. Unlike other classifiers, in GNB classifier, there is no parameter tuning. Finally, only with two feature sets, the GNB classifier classifies the different MI tasks accurately and quickly. The pipeline followed in our work, like channel selection, band-pass filter-based CSP, feature selection and GNB model, proves to be a better method for real-time MI-based BCI applications.

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

2. DATA AND METHOD

This section will describe the MI data used in this research and then the steps followed in the proposed method; namely, channel selection, pre-processing, feature extraction, feature selection and classification of EEG-based MI data in detail.

2.1 Experimental Data

We used the publicly available dataset IVa from BCI competition III [30] to validate the proposed approach. The dataset consists of EEG-recorded data from five healthy subjects (aa, al, av, aw and 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. The visual cue for 1 trial lasted for 3.5 seconds and the time paradigm of a single trial is shown in Figure 2. For each subject, there were 140

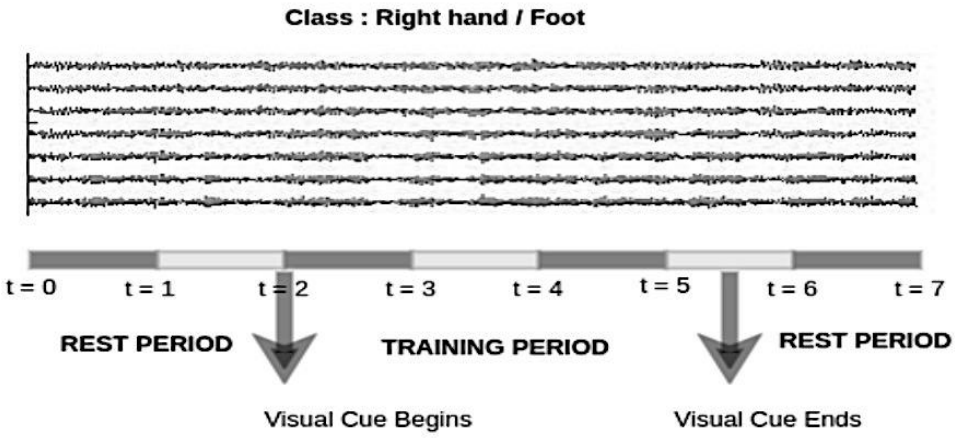

Figure 2. Time taken for a single trial.

trials for each task and therefore 280 trials totally. The measured EEG signal was filtered using a bandpass filter between 0.05 and 200 Hz. Then the signal was digitized at 1000 Hz with 16-bit accuracy.

2.2 EEG Signal Pre-processing

2.2.1 Channel Selection

The dataset consists of EEG recordings from 118 channels, which means that it 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 [31]. 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 increase the robustness of classification system [32]. 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 shown in Figure 3. 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.

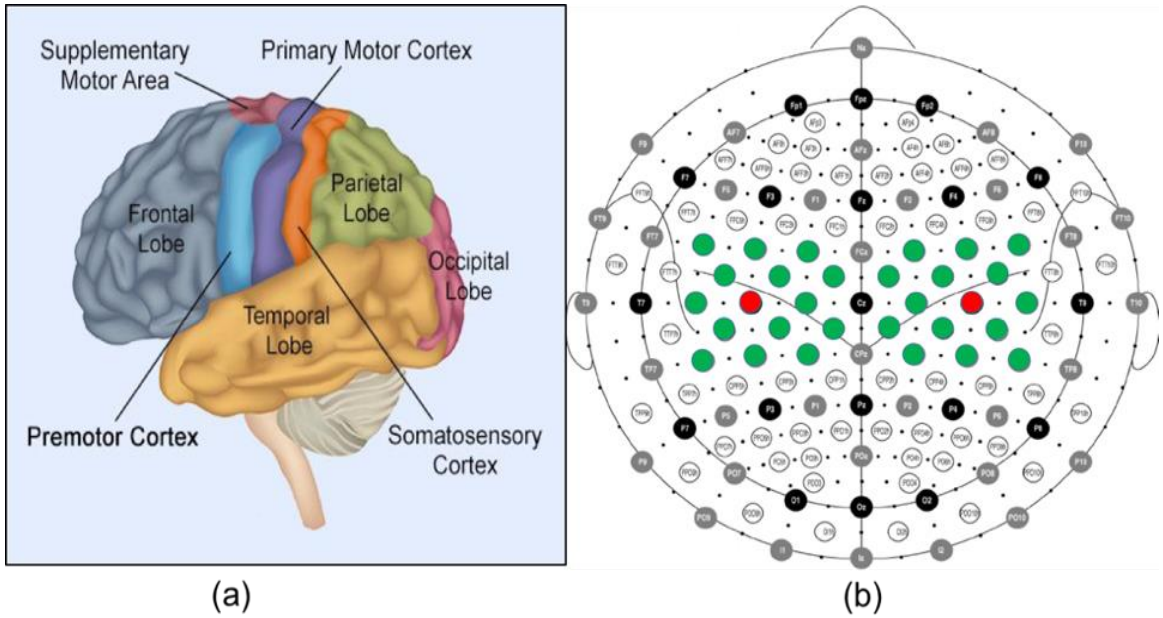

Figure 3. (a) Motor cortex of the brain (b) Standard 10±20 system of electrode placement for 128-channel EEG system. The electrodes in green and red color are selected for processing.

2.2.2 Bandpass Filtering

Since the original sampling rate of the EEG signal is 1000 Hz, it is down sampled to 100 Hz for further processing. Then, the selected 30-channel EEG data is again passed through a band-pass filter between 7 and 30 Hz, as it is known from [9] that mu (μ) and beta (β) rhythms lie within that frequency range. Then, data segmentation is done, where we used three-second data (300 samples) after the display of cue of each trial, assuming that the subject might be moving the right hand or right foot on an average of 3 seconds. Each segmented data is called an epoch.

2.2.3 Spatial Filtering

CSP is one of the most commonly used spatial filters in building MI-based BCIs [10]–[17]. 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.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 04, No. 02, August 2018.

Let $\boldsymbol{X}_H$ and $\boldsymbol{X}_F$ 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:

$$\boldsymbol{X}_i^{\boldsymbol{CSP}} = \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{X}_{\boldsymbol{i}} \tag{1}$$

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

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

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

$$\min_{w}(-w^T C_H w) \quad \text{Subject to} \quad w^T C_F w = 1 \tag{3}$$

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

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

Thus, Equation (2) becomes an eigenvalue decomposition problem, where λ is the eigenvalue which corresponds to the eigenvector w, obtained by solving the following equation:

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

Here, 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 $\boldsymbol{W}$ will have 30 column vectors. From that, we select the first m and last m columns to use as $2m$ CSP filter of $\boldsymbol{W}_{CSP}$.

$$\boldsymbol{W}_{CSP} = [w_1, w_2, \dots, w_m, w_{c-m+1}, \dots, w_c] \in \mathbb{R}^{2m \times c} \tag{6}$$

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

$$\boldsymbol{X}_H^{CSP} \in \mathbb{R}^{2m \times c} \coloneqq \boldsymbol{W}_{CSP}^T\boldsymbol{X}_H$$

$$\boldsymbol{X}_F^{CSP} \in \mathbb{R}^{2m \times c} \coloneqq \boldsymbol{W}_{CSP}^T\boldsymbol{X}_F \tag{7}$$

The CSP filters can be plotted back to see the activations of various regions of the brain. Figure 4 shows the scalp plot, where the first 4 and last 4 magnitudes of the coefficients of the CSP filter are plotted. The dark red colour indicates the highest significance. The upper-left plot indicates the filter w_1 and the down-right plot indicates the last filter w_n.

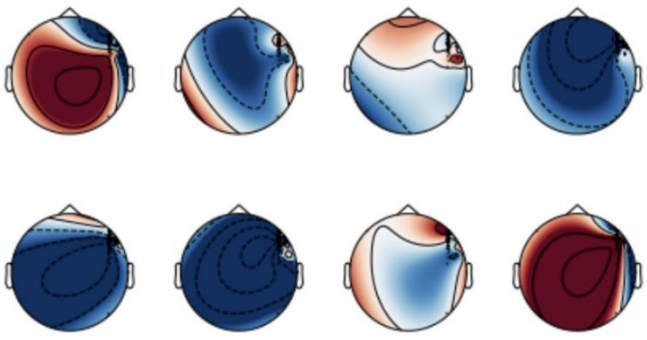

Figure 4. Colormap of magnitudes of the coefficients of CSP filter projected on the scalp.

2.3 Feature Extraction

The spatially filtered signals $\boldsymbol{X}_i^{CSP}$ are obtained for each epoch and 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.

2.3.1 Statistical Features

We extracted seven statistical features; namely, mean, median, standard deviation, skewness, kurtosis, maximum and minimum, as stated in Table 1. These features describe the distribution of EEG signals in terms of amplitudes and moments [18].

Table 1. Statistical features and their description.

Parameters	Description
Mean	Mean value of the signal, $\mu_{X_i^{CSP}} = \frac{1}{N}\sum_{k=0}^{N-1} x_k$
Median	Median (middle) value of the signal
Standard deviation	Standard deviation of the signal, $\sigma_{X_i^{CSP}} = \sqrt{\frac{1}{N-1}\sum_{k=0}^{N-1}(x_k - \mu_{X_i^{CSP}})^2}$
Skewness	Asymmetry value of the signal, $S_{X_i^{CSP}} = \frac{\frac{1}{N}\sum_{k=0}^{N-1}(x_k - \mu_{X_i^{CSP}})^3}{\sigma_{X_i^{CSP}}{}^3}$
Kurtosis	Flatness measure of the signal, $K_{X_i^{CSP}} = \frac{\frac{1}{N}\sum_{k=0}^{N-1}(x_k - \mu_{X_i^{CSP}})^4}{(\sigma_{X_i^{CSP}}{}^2)^2} - 3$
Maximum	Maximum positive amplitude
Minimum	Minimum negative amplitude

2.3.2 Time-domain Features

Time-domain features [19] capture the temporal information of signals. As EEG is known to have a good temporal locality, we extracted a number of time-domain features, as listed in Table 2; namely, hjorth parameters (activity, 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.

Table 2. Time-domain features and their description.

Parameters	Description
Activity	Mean power/variance ($\sigma^2_{X_i^{CSP}}$)
Mobility	$\left(\frac{\sigma'_{X_i^{CSP}}}{\sigma_{X_i^{CSP}}}\right)$, where $\sigma'_{X_i^{CSP}}$ is the standard deviation of first derivative
Complexity	$\left(\frac{\sigma''_{X_i^{CSP}}}{\sigma'_{X_i^{CSP}}} / \frac{\sigma'_{X_i^{CSP}}}{\sigma_{X_i^{CSP}}}\right)$, where $\sigma''_{X_i^{CSP}}$ is the stand. devi. of second derivative
1st Diff. Mean and Max.	Mean and maximum value of the first derivative of the signal
2nd Diff. Mean and Max.	Mean and maximum value of the second derivative of the signal
Mean V-V slope	Mean of vertex to vertex (peak-peak) slope
Variance V-V slope	Variance of vertex to vertex (peak-peak) slope
Mean V-V amplitudes	Mean of vertex to vertex (peak-peak) amplitudes
Variance V-V amplitudes	Variance of vertex to vertex (peak-peak) amplitudes
Zero crossing	Number of times the signal crossing zero
Coeff. of variation	Ratio of standard deviation to the mean.

2.3.3 Frequency-domain features

The frequency-domain features [20] capture the frequency information of brain rhythms during motor imagery tasks. It is known as stated earlier that the frequency of motor imagery signals lies within 7-30 Hz. If the time-domain signal is converted into frequency domain, all frequency related features can be estimated. Power spectral density (PSD) is used to estimate the frequency content of the signal. In our work, PSD is computed using Welch method, where averaging a periodogram spectrum is carried out on overlapping data segments. The spatially filtered signal (300 samples) is divided into small segments and assumed to be stationary. A window function, such as rectangular, hamming, hanning, …etc., is typically chosen for this purpose. For computing PSD, we used Hanning window and took the window size as 100 with 50% overlapping. From the obtained array of frequencies, we 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 its ratios (α/β and β/α). PSD obtained for the given spatially filtered signal is shown in Figure 5.

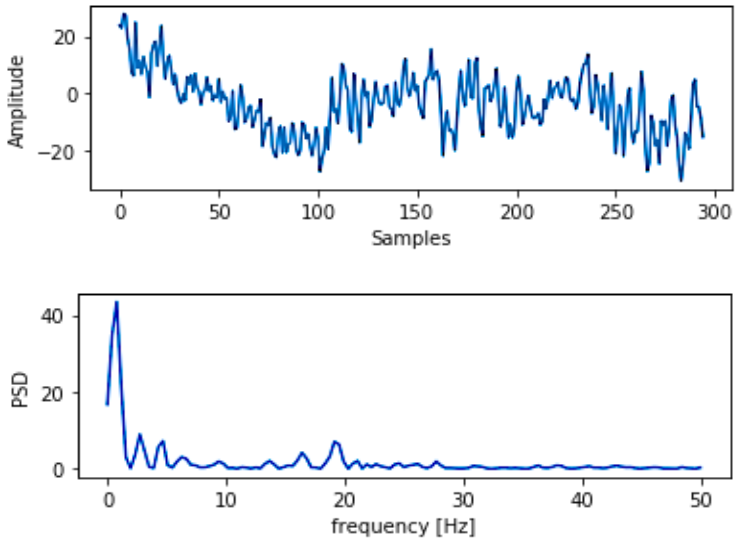

Figure 5. Spatially filtered signal of an epoch and its PSD.

2.3.4 Wavelet-based Features

Wavelet transform [21] 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. Features may be computed for any D_i and A_i corresponding to important MI-based EEG frequency bands. We used *coif1* wavelet with level-1 decomposition, as it gives best result among other wavelets [33]. Here, D_i corresponds to beta-band (13-30 Hz) and A_i corresponds to alpha-band (7-13 Hz). The extracted wavelet-based features are mean, standard deviation, energy and entropy of both D_i and A_i.

2.3.5 Auto-regressive Coefficients

Auto-regressive (AR) method [22] 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:

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

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., $F = \{f_1, f_2, \dots, f_N\}$, where $f_1 \in \mathbb{R}^{2k}$ and $F \in \mathbb{R}^{2k \times N}$.

2.3.6 Feature Normalization

The extracted feature vectors are then normalized to a common range to reduce inter-and intra-subject variability. We used z-score normalization [34], so that the mean value of the signal is zero and the standard deviation is one. Mathematically, this is defined as:

$$z = \frac{x-\mu}{\sigma} \tag{9}$$

where μ is the mean and σ is the standard deviation. Then, out of these normalized features, the discriminative subset of features has to be identified for a reliable classification.

2.4 Feature Selection

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

2.4.1 Minimum-redundancy and maximum-relevance (mRMR)

As the name suggests, this feature selection algorithm is based on selecting features with minimum redundancy and maximum relevance depending on the mutual information values between various features [35]. 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, this is defined as:

Maximum relevance:

$$\max D, \qquad D = \frac{1}{|S|}\sum_{f_i \in S} I(f_i; C) \tag{10}$$

Minimum redundancy:

$$\min R, \qquad R = \frac{1}{|S|^2}\sum_{f_i, f_j \in S} I(f_i; f_j) \tag{11}$$

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 criterion of combining the above two constraints is called minimum-redundancy maximum-relevancy (mRMR) and is given as:

$$\max \Phi, \qquad \Phi = D - R \tag{12}$$

The features are sorted according to mRMR and the first three features are selected as optimal features. The selected features are given in Table 1.

2.4.2 Lasso Regularization-based Feature Selection

In regularization models, classifier induction and feature selection are simultaneously achieved by minimizing fitting errors and properly tuning penalties. The learned classifier w can have coefficients to be very small or zero. Feature selection is achieved by selecting the non-zero coefficients in w [36]. Mathematically, this is defined as:

$$\hat{w} = \min_{w} c(w, X) + \alpha\|w\| \tag{13}$$

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 penalized with l_1 norm is called Lasso regularization and is defined as:

$$\|w\| = \sum_{i=1}^{m} w_i \tag{14}$$

This model will have a sparse solution, such that it forces weak features to have zero coefficients and be excluded from the model. Thus, Lasso (l_1) regularization inherently performs feature selection [37].

The selected features are listed in Table 3.

Table 3. 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 of D_i, kurtosis
Lasso Regularization	Band-power (13-30 Hz), wavelet energy of D_i, AR with 6 coefficients

In both feature selection methods, we selected three features as mentioned in Table 3. Then, the common features between mRMR and Lasso regularization are selected to train the classification model. Thus, in our work, we consider optimal features = features selected by (mRMR ∩ Lasso regularization). Therefore, band-power (13-30 Hz) and wavelet energy of each of the spatially filtered epochs have been selected as the optimal features. We tried to represent the obtained optimal features using scalp plot to find the difference between the MI classes visually. We found that band-power and wavelet energy produce good differentiation between the two classes, as shown in Figure 6. These features are tuned by k-fold cross-validation to create training and testing set.

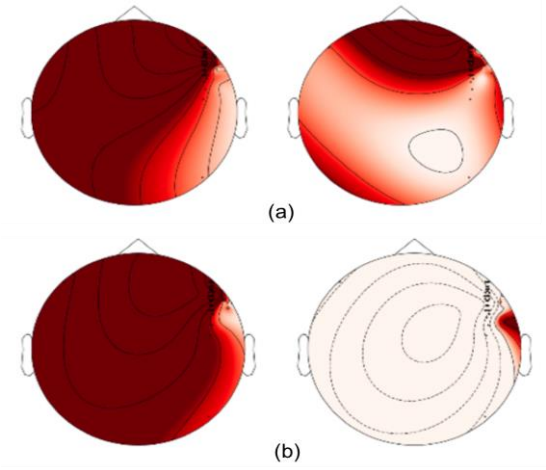

Figure 6. Scalp plot of (a) band-power of right-hand and right-foot MI and (b) wavelet energy for right-hand and right-foot MI.

2.5 Gaussian Naïve Bayes (GNB) Classifier

The Naïve Bayes theorem aims at assigning the class C_i to the feature vector by calculating *a posteriori* probability of the feature vector [38]. Here, $i \in \{H, F\}$, H denotes right-hand and F denotes right-foot MI class. Mathematically, this is defined as:

$$p(C_i|S) = \frac{p(S|C_i) \times p(C_i)}{p(S)} \tag{15}$$

where C_i is the class, $S = \{s_1, s_2, \dots, s_m\}$ 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 training data [39]. Let μ_{jH} and ${\sigma^2}_{jH}$ 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:

$$p(S = s_j|C_H) = \frac{1}{\sqrt{2\pi{\sigma^2}_{jH}}} e^{-\frac{(s_j-\mu_{jH})^2}{2{\sigma^2}_{jH}}} \tag{16}$$

The prediction result provides the class and is defined as:

$$C_{pred} = \underset{i}{\text{argmax}} \quad p(C_i|s_1, s_2, \dots, s_m) \tag{17}$$

In this study, GNB classifier is used to classify two-class MI signals. The main advantage of this

classifier is that there is no parameter tuning like in other classifiers. The parameters in GNB classifier are automatically calculated by maximum likelihood estimation (MLE) [38].

2.6 Performance Measures

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

2.6.1 Confusion Matrix

The confusion matrix is a useful tool for analyzing how frequently instances of a class (say X) were correctly classified as class X instances. 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 instances. Here, right-hand instances should be classified as belonging to the right-hand class. Thus, the numbers of true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN) are obtained. For a classifier to have 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.

2.6.2 Accuracy

The accuracy is the ratio of total number of correctly classified samples to the total number of samples of all classes.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{18}$$

2.6.3 κ Score

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

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{19}$$

where, κ = 1 indicates complete agreement between the MI BCI classes, while κ ≤ 0 means that there is no agreement at all.

3. Experiments, Results and Discussion

In this section, the implementation procedure and the experimental results of the proposed method using the dataset IVa of BCI competition III are explained. In this study, the codes were written in Python 2.7, making use of Scikit Learn [40]; a popular machine learning library.

3.1 Performance of the Proposed Method

The given dataset had two MI classes; right-hand and right-foot, to be classified. We experimented with the proposed model using two different numbers of channels. One experiment considered all the 118 channels and the other considered only the 30 channels present over the motor cortex. The optimal features selected from the dataset after applying band-pass filter-based CSP are given as input to the GNB model. Table 4 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 from 1 to 10; the average and the standard deviation values obtained for all the 10 folds are given in Table 4.

From the values obtained, it is observed that the reduced number of channels gives better accuracy than considering all the 118 channels.

Since we got better accuracy with 30 channels, we proceeded with the experiment with the reduced number of channels. In order 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 built based on the training data and validated using the testing data. Table 5 shows the various performance measures achieved by the proposed method for all the 10 folds.

Table 4. Classification accuracy by *k*-fold cross-validation method for the proposed method.

Subjects	***k*-fold cross-validation accuracy (%) (mean ± std)**	
	118 channels	**30 channels**
aa	93.72±2.81	95.16±1.95
al	89.49±5.73	91.88±4.39
av	91.03±7.22	93.85±4.87
aw	90.48±6.45	94.01±5.62
ay	93.86±3.76	96.65±2.32
Average	**91.72±5.19**	**94.31±3.83**

Table 5. Values of TP, FP, TN and FN of confusion matrix, accuracy (%) and κ score of the proposed method for all the 10 folds.

***k*-folds**	**Confusion Matrix**				**Accuracy (%)**	**κ score**
	TP	**FN**	**FP**	**TN**		
k=1	800	43	39	798	95.12	0.90
k=2	807	39	32	802	95.77	0.92
k=3	805	40	34	801	95.60	0.91
k=4	798	41	40	801	95.18	0.90
k=5	803	40	38	799	95.36	0.91
k=6	799	38	40	803	95.36	0.91
k=7	810	36	32	802	95.95	0.92
k=8	805	41	35	799	95.48	0.91
k=9	808	36	32	804	95.95	0.92
k=10	812	33	28	807	96.37	0.93
Average ± std	**804.7 ± 4.72**	**38.7 ±2.98**	**35 ± 4.11**	**801.6 ±2.67**	**95.61 ± 0.40**	**0.91 ± 0.01**

3.2 Performance Comparison with LDA and SVM Classifiers

The performance of the proposed approach is compared with those of two classifiers; namely, linear discriminant analysis (LDA) [41] and support vector machine (SVM) [42]. 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 chosen carefully to avoid under-fitting and over-fitting problems. The outputs obtained by LDA and SVM classifiers are evaluated and compared with the proposed GNB approach. The performance metrics were generated for all the 10 folds. The average and the standard deviation values obtained are shown in Table 6. The result shows that the proposed method provides an improved accuracy over the LDA and SVM classifiers. The runtime of each classifier is noted down, where the values obtained show that the proposed method takes few milliseconds lesser than LDA and SVM methods.

Table 6. Comparison of performance metrics of the proposed method, LDA and SVM classifiers for all the 10 folds (Mean ± Standard deviation**).**

Methods	**Confusion Matrix**				**Accuracy (%)**	**κ score**	**Runtime (in milliseconds)**
	TP	**FN**	**FP**	**TN**			
LDA	752 ± 6.56	71 ± 7.71	71 ± 6.83	786 ± 4.13	91.55 ± 0.75	0.83 ± 0.03	126 ± 4
SVM	778 ± 5.68	54 ± 7.43	56 ± 5.32	792 ± 4.05	93.45 ± 0.63	0.87 ± 0.02	120 ± 5
GNB	**804.7 ± 4.72**	**38.7 ±2.98**	**35 ± 4.11**	**801.6 ±2.67**	**95.61 ± 0.40**	**0.91 ± 0.01**	**118** ± 4

3.3 Discussion

The proposed approach does not require any artifact or noise removal. The band-pass and spatial filtering used in the method itself removes high-and low-frequency artifacts; therefore, there is no need for any explicit artifact removal methods. The results presented in Table 4 prove that the proposed method produces better accuracy with a minimum number of channels. The existing approaches for MI-based EEG data classification make use of high-dimensional feature vectors. The

accuracy achieved may be comparable, but it makes the BCI system slow. The feature selection method used in our work determines the ideal feature set and makes the BCI system work faster. The main advantage of GNB classifier is that there is no parameter tuning. The results listed in Table 6 demonstrate that the GNB classifier with only two feature vectors produces good accuracy and compiles faster than other classifiers. Henceforth, our method proves to be simpler, faster and more accurate for MI-based BCI applications.

4. CONCLUSIONS

In this work, we used a new combination of machine learning approach to classify two-class MI signals for BCI applications. Firstly, the EEG signals with 118 channels are of high dimension. To reduce 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 prove that the performance of two-class MI-based BCI can be significantly improved using a few channels and a few feature vectors. This method also reduces computational complexity significantly and increases the speed and accuracy of the classifier models. GNB classifier performed better than the LDA and SVM classifiers. Hence, the proposed approach can be used to design 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 purposes.

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] J. Calderon, Y. Yang, C. Inman, J. Willie and G. Berman, "Decoding Human Behavior from Complex Neural Interactions," Bulletin of the American Physical Society, 2018.

[4] H. M. Golshan, A. O. Hebb, S. J. Hanrahan, J. Nedrud and M. H. Mahoor, "An FFT-based Synchronization Approach to Recognize Human Behaviors Using STN-LFP Signal," IEEE Int'l Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 979–983, 2017.

[5] H. M. Golshan, A. O. Hebb, S. J. Hanrahan, J. Nedrud and M. H. Mahoor, "A Hierarchical Structure for Human Behavior Classification Using STN Local Field Potentials," Journal of Neuroscience Methods, vol.293, pp.254–263, 2018.

[6] 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," The 8th Int'l Conference on Intelligent Human Computer Interaction (IHCI), pp. 81-93, 2016.

[7] B. Blankertz, S. Lemm, M. Treder, S. Haufe and K. R. Muller, "Single-trial Analysis and Classification of ERP Components- A Tutorial," NeuroImage, vol. 56, no. 2, pp. 814–825, 2011.

[8] B. He, B. Baxter, B. J. Edelman, C. C. Cline and W. Y. Wenjing, "Non-invasive Brain-Computer Interfaces Based on Sensorimotor Rhythms," Proc. of the IEEE, vol. 103, no. 6, pp. 907–925, 2015.

[9] G. Pfurtscheller and C. Neuper, "Motor Imagery and Direct Brain-Computer Communication," Proceedings of the IEEE, vol. 89, no. 7, pp. 1123–1134, 2001.

[10] H. Ramoser, J. Muller-Gerking and G. Pfurtscheller, "Optimal Spatial Filtering of SingleTtrial EEG during Imagined Hand Movement," IEEE Trans. Rehabilitation Engg., vol. 8, no. 4, pp. 441–446, 2000.

[11] S. Lemm, B. Blankertz, G. Curio and K. R. Muller, "Spatio-spectral Filters for Improving the Classification of Single Trial EEG," IEEE Trans. Biomedical Engg., vol. 52, no. 9, pp. 1541–1548, 2005.

[12] R. Tomioka, G. Dornhege, G. Nolte, B. Blankertz, K. Aihara and K. R. Muller, "Spectrally weighted common spatial pattern algorithm for single trial EEG classification," Department of Mathematical

Informatics, University of Tokyo, Japan, Tech. Rep., vol. 40, 2006.

[13] W. Wu, X. Gao, B. Hong and S. Gao, "Classifying Single-trial EEG during Motor Imagery by Iterative Spatio-spectral Patterns Learning (ISSPL)," IEEE Trans. Biomedical Engg., vol. 55, no. 6, pp.1733–1743, 2008.

[14] H. Higashi and T. Tanaka, "Simultaneous Design of FIR Filter Banks and Spatial Patterns for EEG Signal Classification," IEEE Trans. on Biomedical Engg., vol. 60, no. 4, pp. 1100–1110, 2013.

[15] C. Park, C. C. Took and D. P. Mandic, "Augmented Complex Common Spatial Patterns for Classification of Noncircular EEG from Motor Imagery Tasks," IEEE Trans. Neural Systems and Rehabilitation Engg., vol. 22, no. 1, pp. 1–10, 2014.

[16] A. S. Aghaei, M. S. Mahanta and K. N. Plataniotis, "Separable Common Spatio-spectral Patterns for Motor Imagery BCI Systems," IEEE Trans. on Biomedical Engg., vol. 63, no. 1, pp. 15–29, 2016.

[17] D. Li, H. Zhang, M. S. Khan and F. Mi, "A Self-Adaptive Frequency Selection Common Spatial Pattern and Least Squares Twin Support Vector Machine for Motor Imagery Electroencephalography Recognition," Biomedical Signal Processing and Control, vol. 41, pp. 222–232, 2018.

[18] 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.

[19] C. Vidaurre, N. Kramer, B. Blankertz and A. Schlogl, "Time Domain Parameters as a Feature for EEG-Based Brain–Computer Interfaces," Neural Networks, vol. 22, no. 9, pp. 1313–1319, 2009.

[20] 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 Trans. Neural Systems and Rehabilitation Engg., vol. 16, no. 4, pp. 317–326, 2008.

[21] 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.

[22] M. Tavakolan, X. Yong, X. Zhang and C. Menon, "Classification Scheme for Arm Motor Imagery," Journal of Medical and Biological Engg., vol. 36, no. 1, pp. 12–21, 2016.

[23] Y. Zhang, G. Zhou, J. Jin, X. Wang and A. Cichocki, "Optimizing Spatial Patterns with Sparse Filter Bands for Motor Imagery-based Brain–Computer Interface," Journal of Neuroscience Methods, vol. 255, pp. 85–91, 2015.

[24] H. Wang, Y. Zhang, N. R. Waytowich, D. J. Krusienski, G. Zhou, J. Jin, X. Wang and A. Cichocki, "Discriminative feature extraction *via* multivariate linear regression for SSVEP-based BCI," IEEE Trans. Neural Systems and Rehabilitation Engg., vol. 24, no. 5, pp. 532–541, 2016.

[25] F. Lotte, M. Congedo, A. L´ecuyer, F. Lamarche and B. Arnaldi, "A Review of Classification Algorithms for EEG-based Brain–Computer Interfaces," Jo. of Neural Engg., vol. 4, no. 2, p. R1, 2007.

[26] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang and A. Cichocki, "Sparse Bayesian Classification of EEG for Brain–Computer Interface," IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 11, pp. 2256–2267, 2016.

[27] V. P. Oikonomou, A. Maronidis, G. Liaros, S. Nikolopoulos and I. Kompatsiaris, "Sparse Bayesian Learning for Subject Independent Classification with Application to SSVEP-BCI," The 8th Int'l IEEE/EMBS Conf. Neural Engineering (NER), pp. 600–604, 2017.

[28] Y. Zhang, G. Zhou, J. Jin, Y. Zhang, X. Wang and A. Cichocki, "Sparse Bayesian Multiway Canonical Correlation Analysis for EEG Pattern Recognition," Neurocomputing, vol. 225, pp. 103–110, 2017.

[29] Y. Zhang, Y. Wang, J. Jin and X. Wang, "Sparse Bayesian Learning for Obtaining Sparsity of EEG Frequency Bands-based Feature Vectors in Motor Imagery Classification," Int'l Journal of Neural Systems, vol. 27, no. 02, p. 1650032, 2017.

[30] "BCI Competition III", [Online], Available: http://www.bbci.de/competition/iii

[31] L. He, D. Hu, M. Wan, Y. Wen, K. M. von Deneen and M. Zhou, "Common Bayesian Network for Classification of EEG-based Multiclass Motor Imagery BCI," IEEE Trans. Systems, Man and Cybernetics: Systems, vol. 46, no. 6, pp. 843–854, 2016.

[32] 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 Trans. Neural

Systems and Rehabilitation Engg., vol. 19, no. 6, pp. 617–627, 2011.

[33] 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.

[34] X. Zhang, L. Yao, D. Zhang, X. Wang, Q. Z. Sheng and T. Gu, "Multi-person Brain Activity Recognition *via* Comprehensive EEG Signal Analysis," arXiv preprint arXiv:1709.09077, 2017.

[35] H. Peng, F. Long and C. Ding, "Feature Selection Based on Mutual Information Criteria of Max-dependency, Max-relevance and Min-redundancy," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1226–1238, 2005.

[36] J. Gui, Z. Sun, S. Ji, D. Tao and T. Tan, "Feature Selection Based on Structured Sparsity: A Comprehensive Study," IEEE Trans. Neural Networks and Learning Systems, vol. 28, no. 7, pp. 1490–1507, 2017.

[37] P. Zhao and B. Yu, "On Model Selection Consistency of Lasso," Journal of Machine Learning Research, vol. 7, no. Nov., pp. 2541–2563, 2006.

[38] I. H. Witten, E. Frank, M. A. Hall and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2016.

[39] 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.

[40] Scikit-learn, "Machine Learning in Python", [Online], Available: http://scikit-learn.org/

[41] 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.

[42] 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 Trans. Neural Systems and Rehabilitation Engg., vol. 20, no. 4, pp. 526–538, 2012.

ملخص البحث:

في أكثر الأحوال، تكون لدى الأشخاص المعاقين حركياً صعوبات في التواصل مع المجتمع. ويمكن للبينيّات التي تربط الدماغ والحاسوب أن تعيد شيئاً من الأمل لهؤلاء. وتشكل إشارات التصوير الحركي المسجلة بواسطة المخططات الدماغية الكهربائية الأساس الأكثر ملاءمة لتصميم بينيّات الدماغ-الحاسوب، لأنها توفر درجة أعلى من الحرية. وتساعد بينيّات الدماغ-الحاسوب القائمة على التصوير الحركي الأشخاص على التفاعل مع أي من تطبيقاتها عبر أداء سلسلة من مهمات التصوير الحركي. إلا أن هناك عدداً من التحديات، منها: اختلاف ملاءمتها من شخص لآخر، واستخلاص خصائصها المعيّنة للمستخدم، وزيادة دقتها. وتقترح هذه الدراسة طريقة للتغلب على المشكلات المذكورة، حيث نقترح طريقة تأخذ بعين الاعتبار: اختيار القنوات، والنمط المكاني المشترك القائم على استخدام مرشِّح تمرير نطاق ترددي، واستخلاص واختيار الخصائص، والنمذجة باستخدام مصنِّف بايز الغاوسي البسيط. ونظراً لأنه يتم اختيار الخصائص الأمثل عن طريق تقنيات اختيار الخصائص، فإن ذلك يساعد في التغلب على مسألة اختلاف ملاءمة البينيّات من شخص لآخر ويؤدي الى تحسين أداء المصنِّف. وتم التحقق من نجاعة الطريقة المقترحة باستخدام مجموعة البيانات (IVa)، كما جرت مقارنة المصنِّف المستخدم في هذه الدراسة بمصنِّفين آخرين هما: (LDA و SVM). وأثبتت النتائج تفوق المصنِّف المقترح من حيث الدقة. كذلك يمكن تطوير الطريقة المقترحة مستقبلاً لتصميم تطبيقات موثوقة وفي الزمن الحقيقي لبينيّات الدماغ-الحاسوب المستندة الى التصوير الحركي.