Detection of motor imagery EEG signals employing Naïve Bayes based learning process

Detection of motor imagery EEG signals employing Naïve Bayes based learning process

Siuly, Hua Wang and Yanchun Zhang
Centre for Applied Informatics, College of Engineering & Science
Victoria University, Melbourne, Australia
siuly.siuly@vu.edu.au ; hua.wang@vu.edu.au; yanchun.zhang@vu.edu.au

Abstract. The objective of this study is to develop a reliable and robust analysis system that can automatically detect motor imagery (MI) based electroencephalogram (EEG) signals for the development of brain-computer interface (BCI) systems. The detection of MI tasks provides an important basis for designing a communication way between brain and computer in creating devices for people with motor disabilities. This paper presents a synthesis approach based on optimum allocation system and Naive Bayes (NB) algorithm for detecting mental states based on EEG signals. In this study, an optimal allocation (OA) is introduced to discover the most effective representatives with minimal variability from a large number of MI based EEG data and the NB classifier is employed on the extracted features for discriminating the MI signals. The feasibility and effectiveness of the proposed method is demonstrated by analyzing the results and its success on two public benchmark datasets. The results indicate that the proposed approach outperforms the most recently reported five methods and achieves 0.64%-20.90% improvement on average accuracy. The performances of this proposed approach implies that it can be reliably used to detect EEG based MI activity and can be a promising avenue for EEG based BCI applications.

Key-words. Motor imagery; EEG; Optimum allocation; Naïve Bayes; Brain computer interface.

1. Introduction

Motor imagery (MI) is one of the most frequently used mental strategies in brain-computer interface (BCI) applications [1, 2, 3] for severe motor disabled patients and rehabilitation [4, 5, 6, 7]. MI is a common mental task in which subjects are instructed to imagine themselves performing a specific motor action (such as a hand or foot movement) without an overt motor output [8] and each task is treated as a MI class. There are various acquisition techniques for capturing MI brain activities such as, electroencephalography (EEG), electrocorticography (ECoG), Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI), and Magnetoencephalography (MEG). Among these techniques, EEG is the most studied measure of potential for non-invasive BCI designs, mainly due to its excellent temporal resolution, non-invasiveness, usability, and low set-up costs [9, 10, 11, 12]. The BCI is a communication system that provides a direct communication pathway for

transmitting messages from the human brain to computers by analyzing the brain's mental activities [13,14]. Obtaining this communication channel between brain and computer is very essential in creating devices for disabled people (who cannot move their hands, legs or other limbs), and for taking the interaction between human and machine to another level.

In the BCI development, users produce EEG signals of different brain activity for different MI tasks that will be identified by a system and are then translated into commands. These commands will be used as feedback for motor disabled patients to communicate with the external environments. If the MI tasks are reliably distinguished through detecting typical patterns in EEG data, a motor disabled people could communicate with a device by composing sequences of these mental states. Thus, a MI-based BCI provides a promising control and communication means to people suffering from motor disabilities. Therefore, the detection of MI tasks is very essential for the BCI development to generate control signals. In most current MI based BCIs, the detection algorithms are carried out mainly in two stages: feature extraction and feature detection [15]. A successful EEG-based BCI system mainly depends on whether the extracted features are able to differentiate MI-oriented EEG patterns. How to improve the recognition performance of MI signals is still a vital issue for the development of BCI systems. The goal of the study is to develop an approach for detecting different MI EEG signals improving the classification performance. The present study proposes a methodology where an optimum allocation scheme is developed for feature extraction stage and a probabilistic classifier, Naïve Bayes (NB) is employed for detecting the obtained features. In order to verify the effectiveness of the proposed approach, we compare it with the five most recently reported methods that are discussed in Section 2.

There are strong grounds of using an optimum allocation technique for getting representative sample from each group of a category of MI data in this study. An optimum allocation technique is developed to allocate numbers of sample units into different groups with a minimum variation, providing the most precision. This method is applicable when a dataset is heterogeneous and very large in size. When measuring an EEG, a large amount of data with different categories is obtained over a time period and this huge amount data are not directly usable in BCI applications. Then it is required to divide the dataset into several groups to make homogeneity within group according to their specific characteristics and used to select representative samples from the groups, such that those samples reflect the entire data. Thus this study intends to develop an optimum allocation technique based sampling to select representative sample points from every time group instead of random sampling. In the optimum allocation based sampling, sample points are selected from each group considering variability of the observations but the random sampling does not consider variability. To describe the original patterns of EEG signals more representatively, the variability consideration is the most important thing to provide the highest precision of a sample for the least cost during the selection of sample points from a group. In this study, a sample is defined as a subset (or small part) of observations from a group.

Through optimum allocation procedure, a sample selected from a group of a particular class of MI data is called an 'optimal allocated sample' denoted as Opt.S and all of the optimal allocated samples together for that MI category is called AOS set as described in details in Section 3. In order to achieve the representative characteristics, eleven statistical features are extracted from the AOS set as discussed in details in Section 3.2.2. These features represent the characteristics of the original MI EEG data without redundancy. The extracted features are then used as the inputs to the Naïve Bayes (NB) probabilistic model for the classification of MI EEG signals. To the best of our knowledge, such an optimum allocation based NB approach has not been used on the MI data for detection of MI task in BCI so far. The reason of choosing of the NB method as a classifier for this study is due to the simplicity of its structure, and the speed of the learning algorithm it employs [16, 17]. Another advantage is that small amount of bad data or "noise" does not perturb the results by much.

The proposed approach is evaluated on two datasets, IVa and IVb of BCI Competition III [18, 19], where both sets contain MI EEG recorded data. A popular *k*-fold cross validation method (*k*=10) is used to assess the performance of the proposed method for reducing the experimental time and the number of experiments in the MI tasks EEG signal classification. This cross-validation procedure is applied to control over-fitting of the data. The performance of the proposed approach is also compared with five most recent reported methods. The study results from both datasets demonstrate that our proposed algorithm produces a promising performance for the detection of MI EEG signals. Experimental results also show that the proposed approach outperforms the other five most recently reported methods with respect to the classification performance for dataset IVa.

The rest of the paper is organized as follows: Section 2 presents a review of the existing methods of the detection of MI EEG signals. Section 3 describes methodology that is proposed in this study. This section also discusses the performance evaluation methods. The experimental results and discussions are provided in Section 4. Finally Section 5 draws the conclusions of the study.

2. Related work

Due to the rapidly growing interest in the MI-based BCIs, numerous methods have been reported by different researchers for the detection of MI EEG signals. In this section, we have provided a brief description of most recently reported five methods, which were implemented on dataset IVa of BCI Competition III.

Suk and Lee in [20] introduced a bayesian spatiospectral filter optimization (BSSFO) based bayesian framework for discriminative feature extraction for motor imagery classification in an EEG-based BCI in which the class-discriminative frequency bands and the corresponding spatial filters are optimized by means of the probabilistic and information-theoretic approaches. In that work, the problem of simultaneous spatiospectral filter optimization is formulated as the estimation of an unknown posterior probability density function (pdf) that represents the probability that a single-trial

EEG of predefined mental tasks can be discriminated in a state. In order to estimate the posterior pdf, they proposed a particle-based approximation method by extending a factored-sampling technique with a diffusion process. The method achieved overall 75.46% of classification accuracy. The weakness of their method is that the classification performances are very low that are not enough for comparison with the existing methods.

Zhang et al. [21] devised an approach based on z-score linear discriminant analysis (Z-LDA), which introduces a different decision boundary definition strategy to handle with the heteroscedastic class distributions. They employed common spatial pattern (CSP) to estimate the spatial projection matrix, which projects the EEG signal from original sensor space to a surrogate sensor space in the feature extraction stage. Finally they used the obtained features to the Z-LDA and then compared with LDA, support vector machine (SVM), nonparametric discriminant analysis (NDA) and heteroscedastic LDA (HLDA). Although the CSP is a popular method in BCI applications, it is very sensitive to noise, and often over-fits with small training sets. The overall accuracy performance was 81.1% for their proposed Z-LAD approach that is not a sufficient amount.

Siuly and Li [22] developed a scheme based on cross-correlation and least square support vector machine (LS-SVM) for the detection of two-class MI signals. That study employed a cross-correlation technique for feature extraction and a least square support vector machine (LS-SVM) for classifying the obtained features. The effectiveness of the proposed classifier was verified replacing the LS-SVM classifier by a logistic regression (LR) classifier and a kernel logistic regression (KLR) classifier, separately, with the same extracted features. Experimental results showed the superiority of the LS-SVM classifier compared to the LR and KLR classifiers. Their method achieved the overall classification accuracy of 95.72%. This method may not be suitable if data is very large in size as the cross-correlation technique takes more time in execution.

Siuly et al. [23] reported a clustering technique-based LS-SVM for the classification of EEG signals. They developed a clustering technique for feature extraction and the obtained features were used to the LS-SVM as the inputs for recognition of EEG signals. It employed the 10-fold cross-validation method to evaluate the performance .The average accuracy was 88.32%. The weakness of that method was that they did not select the parameters optimally through any technique. They manually selected the parameters for the LS-SVM method.

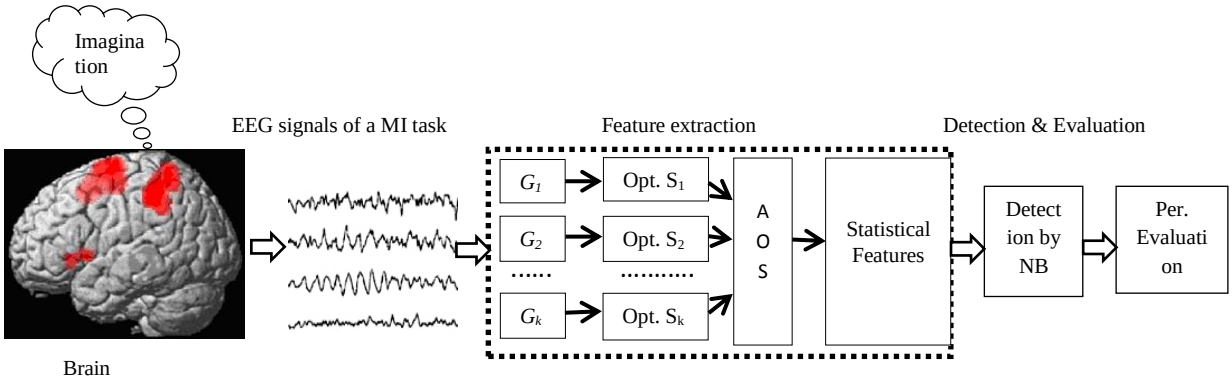
Lu et al. [24] introduced a regularized common spatial patterns (R-CSP) algorithm by incorporating the principle of generating learning for EEG signal classification. That study used two regularization parameters in regularizing the covariance estimates and these parameters were not selected optimally through a technique. They obtained an average accuracy rate of 74.2% for all subjects. It was reported that the algorithm was particularly effective in small sample settings.

Addressing aforementioned problems, this paper proposes an automatic approach based on optimum allocation and Naïve Bayes (NB) classifier which can discriminate two-class MI tasks for the development of BCI systems. In the proposed algorithm, an optimum allocation based technique is

developed to obtain the most representative sample points from each group of a MI category considering minimum variability and a NB classifier is applied to detect the MI tasks for the application of BCI systems. The proposed method will be suitable for applicable for any large size of EEG data.

3. Methodology

The proposed approach aims to develop a methodology for the detection of MI based EEG signals for the application in BCI systems that can work in automatic way. The developed scheme in this study is labelled in four stages as described in Fig. 1. The first stage is the data acquisitions, the second stage is feature extraction, third is detection and the final stage is performance evaluation. These stages are discussed in the following sections.



Note: G_1=Group 1; G_2=Group 2; G_k=Group k; Opt.S_1=Optimal allocated sample 1; Opt.S_2= Optimal allocated sample 2: Opt.S_k= Optimal allocated sample k; AOS= All of the Optimal allocated samples together from the groups of a class.

Fig.1. Diagram for the proposed methodology for detection of MI EEG signals

3.1. Signal acquisitions

In this study, we used two datasets, IVa and IVb from BCI Competition III [18,19], which was provided by Fraunhofer FIRST, Intelligent Data Analysis Group (Klaus-Robert Müller, Benjamin Blankertz), and Campus Benjamin Franklin of the Charité - University Medicine Berlin, Department of Neurology, Neurophysics Group (Gabriel Curio).

Dataset IVa [18, 19] comprises EEG recordings of five healthy subjects (namely, **aa**, **al**, **av**, **aw** and **ay**) with a sampling frequency of 100 Hz. In each trial, a visual cue was shown for 3.5 s and the subjects performed left hand, right hand, or right foot motor imaginary, but cues for only the classes of right hand (RH) and right foot (RF) were provided for the competition [19]. The subjects sat in comfortable chairs with their arms resting on armrests. This data set contains MI EEG data from the four initial sessions without feedback. The EEG signals were recorded from 118 electrodes

according to the international 10/20 system. There were 280 trials for each subject, namely 140 trials for each task per subject. A training set and a testing set consisted of different sizes for each subject. Among 280 trials, 168, 224, 84, 56 and 28 trials compose the training set for subjects, **aa**, **al**, **av**, **aw**, **ay**, respectively, and the remaining trials compose the testing set. This study uses the down-sampled data at 100 Hz where the original sampling rate is 1000 Hz.

Dataset IVb [18, 19] was collected from one healthy male subject. He sat in a comfortable chair with arms resting on armrests. This data set has the data from the seven initial sessions without feedback. The EEG data consisted of two classes: left hand (LH) and right foot (RF) MI. Signals were recorded from 118 channels in 210 trials. 118 EEG channels were measured at the positions of the international 10/20 system. Signals were band-pass filtered between 0.05 and 200 Hz and digitized at 1000 Hz with 16 bit (0.1 μV) accuracy. They provided a version of the data that was down-sampled at 100 Hz, which is used in this research.

3.2. Feature extraction

This study develops an optimum allocation based approach for feature extraction to find a suitable representation of the original EEG recordings. The extracted features provide the inter-class discrimination information for detecting different categories or different classes (e.g. right hand movement; right foot movement) of MI tasks. The proposed optimum allocation based approach consists of the following steps as described below.

3.2.1. Data partition

In this step, the full data of EEG signals for each category (e.g. right hand movement) of MI tasks is partitioned into various groups to properly account for possible stationarities as signal processing methods require stationarity of signals. Although an overall EEG signal may not be stationary, usually smaller windows, or parts of those signals will exhibit stationarity. The partitions of the observations are performed with respect to a specific time period. The time period is determined viewing the signals periodic patterns in each class. Each partition is called 'group' in this work where the groups for the data of a particular MI task are denoted as $G_1, G_2, \ldots, G_k$ as shown in Fig.1. The number of observation of k groups are denoted as $N_1, N_2, \ldots, N_k$, respectively. It is worthy to mention that the groups must be non-overlapping.

Based on the data structure, we segment the recorded EEG signals of every MI task in each subject into seven ($k=7$) groups such as $G_1, G_2, \ldots, G_7$ for dataset IVa and into ten ($k=10$) groups such as $G_1, G_2, \ldots, G_{10}$ for dataset IVb. For the RH class of dataset IVa, we get the number of observation for each of seven groups as 11627 that means $N_1 = N_2 = \ldots = N_7 = 11627$, while the RH class holds 81389 data points of 118 dimensions. For the RF class of the same dataset, we acquire the sizes of each group is 15689 that means $N_1 = N_2 = \ldots = N_7 = 15689$ while the RF consists of 109823

observations of the same dimension. For dataset IVb, we get 9743 data points in each of the groups for the LH class e.g. $N_1 = N_2 = \ldots = N_{10} = 9743$ and 11065 data points in each of the groups of the RF class, e.g. $N_1 = N_2 = \ldots = N_{10} = 11065$ while the LH class and RF class hold 97430 and 110652 data points of 118 dimensions, respectively.

3.2.2. Determination of optimal allocated sample (Opt.S) size and then selection of the Opt.S by the optimum allocation

This step aims to select a representative sample from every group of a MI task in each subject considering minimum variance. Generally in a random sample section, variability is not considered within a group which is most important thing to provide precision of sample. This study develops an approach called optimum allocation, which is used to determine the number of observations to be selected from different groups considering minimum variability among the values. If the variability within a group is large, the size of a sample from that a group is also large. On the other hand, if the variability of the observations within a group is small, the sample size will be small in that group. Furthermore, this optimum allocation is also used to find out how a given total sample size for an entire dataset of each MI task in a subject, denoted as n, should be allocated among the k groups with the smallest possible variability. In this study, the observation of EEG signals of each MI class (e.g. movement of right hand) is considered as a population.

Suppose, x_{ijl} is the value of the l^{th} observation of the j^{th} channel in the i^{th} group in a sample. Here $i=1, 2,\ldots, k$; $j=1, 2,\ldots, h$; $l=1,2,\ldots, n_i$, where n_i is the sample size of the i^{th} group which is determined by the optimum allocation approach. X_{ijl} is the corresponding value in the population where $l=1, 2,\ldots, Ni$. The precision of each group largely depends on the choice of the sample size. In order to find out the variability of the mean in this process, we assume that the samples are drawn independently from different group, and the sample mean is an unbiased estimator of the population mean $\bar{X}$. The variance of the sample mean $\bar{x}$, $V(\bar{x}) = E[\bar{x} - E(\bar{x})]^2 = E[\bar{x} - \bar{X}]^2$

$$where\ \bar{x} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{h}\sum_{l=1}^{n_i} x_{ijl}}{n_1 h + n_2 h + \ldots + n_k h} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{h} n_i \bar{x}_{ij}}{nh}, \text{ and } \quad \bar{X} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{h}\sum_{l=1}^{N_i} X_{ijl}}{n_1 h + n_2 h + \ldots + n_k h} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{h} N_i \bar{X}_{ij}}{nh}$$

$$\Rightarrow V(\bar{x}) = E[\frac{1}{h}\sum_{i=1}^{k}\sum_{j=1}^{h}\frac{N_i}{N}(\bar{x}_{ij} - \bar{X}_{ij})]^2;$$

Assuming that the sampling fraction is the same in all groups, e.g. $\frac{n_i}{n} = \frac{N_i}{N}$ $where\ n = n_1 + n_2 + .. + n_k\ and\ N = N_1 + N_2 + .. + N_K$

$$\Rightarrow V(\bar{x}) = \frac{1}{h^2}[\sum_{i=1}^{k}\sum_{j=1}^{h}\frac{{N_i}^2}{N^2}E(\bar{x}_{ij} - \bar{X}_{ij})]^2 = \frac{1}{h^2}\sum_{i=1}^{k}\sum_{j=1}^{h}\frac{{N_i}^2}{N^2}V(\bar{x}_{ij})) \qquad (1)$$

Here, $\bar{x}_{ij}$ is the mean of a simple random sample in the j^{th} channel of the i^{th} group whose variance is

$$V(\bar{x}_{ij}) = \frac{(N_i - n_i)}{N_i} \frac{{s_{ij}}^2}{n_i} \text{ by [26]}$$

Putting the value of $V(\bar{x}_{ij})$ into equation (1), we obtain,

$$V(\bar{x}) = \frac{1}{h^2} \sum_{i=1}^{k} \sum_{j=1}^{h} \frac{{N_i}^2}{N^2} \frac{(N_i - n_i)}{N_i} \frac{{s_{ij}}^2}{n_i} \tag{2}$$

Where N_i is the size of the i^{th} group; n_i is the required sample taken from the i^{th} group; ${s_{ij}}^2$ is the standard deviation of the j^{th} channel in the i^{th} group; and n is the total sample size in the stratification process.

Now let us see how a given total sample size, n, should be allocated among different group so that the estimator, $\bar{x}$, will have the smallest possible variability. The question is to determine n_1, n_2, ...,n_k for minimizing, $V(\bar{x})$, subject to the constraint that the total size n equals $n = n_1 + n_2 + .. + n_k$. This is equal to minimizing the function

$$\phi = V(\bar{x}) + \lambda(\sum_{i=1}^{k} n_i - n) = \frac{1}{h^2} \sum_{i=1}^{k} \sum_{j=1}^{h} \frac{{N_i}^2}{N^2} \frac{(N_i - n_i)}{N_i} \frac{{s_{ij}}^2}{n_i} + \lambda(\sum_{i=1}^{k} n_i - n) \tag{3}$$

For n_i, λ is an unknown Langrange's multiplier. For the extreme case of the function, we have $\frac{\delta\phi}{\delta n_i} = 0$ *and* $\frac{\delta^2\phi}{\delta {n^2}_i} > 0$. By differentiating the function ϕ with respect to n_i and equating the derivation to zero, we have,

$$\frac{\delta\phi}{\delta n_i} = -\frac{1}{h^2} \sum_{i=1}^{k} \sum_{j=1}^{h} \frac{{N_i}^2}{N^2} \frac{(N_i - n_i)}{N_i} \frac{{s_{ij}}^2}{n_i} + \sum_{i=1}^{k} \lambda = 0$$

$$\Rightarrow n_i = \frac{N_i}{Nh\sqrt{\lambda}} \sqrt{\sum_{j=1}^{h} {s_{ij}}^2} \tag{4}$$

Summing up the both sides of equation (4), we have $\sqrt{\lambda} = \frac{\sum_{i=1}^{k} (N_i \sqrt{\sum_{j=1}^{h} {s_{ij}}^2})}{hNn}$ and putting the value of $\sqrt{\lambda}$ into equation (5), we get:

$$n_i = \frac{N_i \sqrt{\sum_{j=1}^{h} {s_{ij}}^2}}{\sum_{i=1}^{k} (N_i \sqrt{\sum_{j=1}^{h} {s_{ij}}^2})} \times n \tag{5}$$

Thus equation (5) is derived to calculate the best sample size for the i^{th} group solving a set of equations by optimum allocation. Using equation (5), a sample selected from a group of a MI task in a subject is called 'optimum allocated sample' denoted as Opt.S. The all of the Opt.S (s) from the groups of a MI task together makes a matrix called AOS as described in Fig.1. For example: if we select three Opt.S from three groups of a MI class with the sizes 10, 12, 11,respectively, then the sizes of AOS will be 33. In equation (5), total sample size, n is determined by using equations (6) in [26, 27]

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{PS}} \tag{6}$$

Here, $n_0 = \frac{z^2 \times p \times q}{d^2}$ where n_0 means the initial sample size, z is the standard normal variate (Z-value) for the desired confidence level; p is the assumed proportion in the dataset estimated to have a particular characteristic; q=1-p and d is the margin of errors or the desired level of precision; and PS denote population size which consider the total number of data points in a class. The total sample size, n can be calculated by using a survey software called '*Sample size calculator* ' that is available in online, http://www.surveysystem.com/sscalc.htm.

Generally, the sum of all Opt.S sizes from all groups in a MI class should be approximately equal to the total sample size (n) of that class (for example, $n = n_1 + n_2 + .. + n_k$) as all groups come from individual MI class . Sometimes, the calculated n may be a bit larger than the given n due to the rounding figure of the calculated sample size. In this research, we get Z=2.58 considering 99% confidence level; d=0.01 for 99-100% confidence interval. If the estimator p is not known, 0.50 (50%)

Table 1: Calculated sample size by the optimum allocation approach for dataset IVa

Groups	Sizes	Obtained sizes of the Opt.S in each of the seven groups of every two class	
		RH	RF
G_1	n_1	3786	1702
	N_1	11627	15689
G_2	n_2	1895	1473
	N_2	11627	15689
G_3	n_3	1674	2360
	N_3	11627	15689
G_4	n_4	1567	3945
	N_4	11627	15689
G_5	n_5	1344	2429
	N_5	11627	15689
G_6	n_6	2150	1476
	N_6	11627	15689
G_7	n_7	1401	1067
	N_7	11627	15689
AOS	Total, n	13817	14452
	Total N	81389	109823

is used as it produces the largest sample size. The larger the sample size, the more sure we can be, that their answers truly reflect the whole data. Thus, we consider p=0.50 so that the sample size is the maximum and q=1-p=.50 (50%). By equation (6), for dataset IVa, we obtain, n=13816 for the RH class with data size of 81389 and n=14451 for the RF class with the data size of 109823. For dataset IVb, we have n=14213 for the LH class with the size of 97430 and n=14466 for the RF class with the size of 110652.

The sizes of the Opt.S (n_i) for each group of every MI class in every subject for dataset IVa and IVb are calculated by equation (5) as presented in Tables 1 and 2, respectively. As the number of data points in each of the five subjects of dataset IVa is same, the calculated sample sizes for each group of every class in Table 1 are applicable for every subject. As shown in both tables, the sample sizes are not equal in every group in a class, due to different variability of the observations in different groups. Using the obtained sample size of each group (displayed in Tables 1 and 2), we select sample from every group of each class in both datasets. As mentioned before, a selected sample from a group is called Opt.S and all Opt.S in a class for a subject are integrated together denoted as AOS set of that class. For example, as shown in Table1, for the RH class of dataset IVa, we obtain seven Opt.S for each subject with the sizes of 3786, 1895, 1674, 1567, 1344, 2150, 1401 (e.g. n_1=*3786*, n_2=*1895*, n_3=*1674*, n_4=*1567*,n_5=*1344*, n_6=*2150 and* n_7=*1401)*, while they are 1702, 1473, 2360, 3945, 2429, 1476 and 1067 for the RF class. Thus the AOS set for the RH class and the RF class in every subject

Table 2: Calculated sample size by the optimum allocation approach for one subject for dataset IVb

Groups	Sizes	Obtained sizes of the Opt.S in each of the seven groups of every two class	
		LH	RF
G_1	n_1	1219	1506
	N_1	9743	11065
G_2	n_2	1052	1312
	N_2	9743	11065
G_3	n_3	2445	924
	N_3	9743	11065
G_4	n_4	1024	2141
	N_4	9743	11065
G_5	n_5	1031	1473
	N_5	9743	11065
G_6	n_6	1922	1705
	N_6	9743	11065
G_7	n_7	1291	2218
	N_7	9743	11065
G_8	n_8	1625	869
	N_8	9743	11065
G_9	n_9	1922	949
	N_9	9743	11065
G_{10}	n_{10}	1291	1371
	N_{10}	9743	11065
AOS	Total, n	14822	14468
	Total N	97430	110652

of dataset IVa consists of 13817 and 14452 observations, respectively as displayed in Table1. Again, for dataset IVb, it can be seen in Table 2 that the size of the AOS set for the LH class and the RF class are 14822 and 14468, respectively. Then these AOS sets are used to extract representative characteristics. Note that in both datasets, the dimension of each AOS set for every subject is 118.

3.2.3. Statistical feature extraction

Choosing good discriminating features is the key to any successful pattern recognition system. It is usually hard for a BCI system to extract a suitable feature set which distils the required inter-class discrimination information in a manner that is robust to various contaminants and distortions. This study considers eleven statistical features: *mean, median, mode, standard deviation, maximum, minimum, first quartile (Q_1), third quartile (Q_3) (75^{th} percentile), inter-quartile range (IQR), skewness and kurtosis*. These features are calculated from each AOS set of every class to achieve representative characteristics that ideally contain all important information of the original signal patterns. The reasons of considering those features are described here. *Mean* corresponds to the centre of a set of values while *median* is the middle most observation. *Mode* is the value in the data set that occurs most often. In a tabular form, the *mode* is the value with the highest frequency. *Mean* and *median* are the measures irrespective of data are discrete or continuous. However, the *mode* is most suitable for discrete data but is tricky for continuous case. The *mode* for a continuous probability distribution is defined as the peak of its histogram or density function. *Mean*, *median* and *mode* are the most used features that can describe almost all distributions with a reasonable degree of accuracy [22, 28, and 29] and provide a fairly good idea about the nature of the data. *Standard deviation* gives information about the spread of data on how close the entire set of data is to the average value in the distribution. *Maximum and minimum* values are used to describe the range of observations in the distribution. Q_1 and Q_3, measure how the data is distributed in the two sides of the median. *IQR* is difference between Q_3 and Q_1 that is used in measuring the spread of a data set that excludes most outliers. *Skewness* describes the shape of a distribution that characterizes the degree of asymmetry of a distribution around its mean [30]. *Kurtosis* measures of whether the data are peaked or flat relative to a normal distribution.

In this step, we calculate a feature sets of eleven features from each AOS set in each class from a subject in both datasets. From every AOS set of each MI class, we acquire a feature vector set of size 118 with11 dimensions. Thus we obtain a vector set of size 236 with 11 dimensions for two-class MI data of every subject in datasets, IVa and IVb. In each subject, the obtained feature vector set is divided into a training set and a testing set using the *10*-fold cross validation approach as discussed in Section 3.4. The training set is applied to train a classifier and the testing vectors are used to verify the accuracy and the effectiveness of the classifiers for discriminating MI tasks. In our experiments, the proposed method is trained on one single subject in the both datasets, separately, as the MI based

EEG signals are naturally highly subject-specific depending on physical and mental tasks. In this research we present all experimental results from the testing set.

3.3. Detection

This study employs Naive Bayes (NB) classifier to detect two-class MI tasks for the application of BCI systems as it provides a flexible way for dealing with any number of attributes or classes and fastest learning algorithm that examines all its training input. The NB [16, 17, 31, 32] is a straightforward and frequently used probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. The NB classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. Depending on the precise nature of the probability model, the NB classifier can be trained very efficiently in a supervised learning setting. In practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood. In this classifier, each class with highest post-probability is addressed as the resulting class.

Suppose, $X=\{X_1, X_2, X_3,.....,X_n\}$ is a feature vector set that contains C_k (k=1,2,..m) classes data to be classified into. Each class has a probability $P(C_k)$ that represents the prior probability of detecting a feature into C_k and the values of $P(C_k)$ can be estimated from the training dataset. For the n feature values of X, the goal of classification is clearly to find the conditional probability $P(C_k| x_1, x_2, x_3,.....,x_n)$. By Bayes's rule, this probability is equivalent to

$$P(C_k| X_1, X_2, X_3,.....,X_n)= \frac{P(C_k)P(X_1, X_2, X_3,.....,X_n | C_k)}{\sum P(C_k)P(X_1, X_2, X_3,.....,X_n | C_k)} \quad (7)$$

Using the chain rule for the reaped application of conditional probability, we have,

$$P(C_k, X_1, X_2, X_3,.....,X_n)=P(C_k).P(X_1, X_2, X_3,.....,X_n|C_k)$$
$$=P(C_k).P(X_1| C_k).P(X_2|C_k,X_1).P(X_3|C_k,X_1,X_2)....P(X_n|C_k,X_1,X_{2,}...X_{n-1}) \quad (8)$$

For the joint probability and for the independent assumption of Naïve Bayes theorem, we get

$$P(C_k, X_1, X_2, X_3,.....,X_n)=P(C_k).P(X_1| C_k).P(X_2|C_k).P(X_3|C_k)....P(X_n|C_k)= P(C_k)\prod_{i=1}^{n} P(X_i | C_k) \quad (9)$$

Thus from equation (7) we have,

$$P(C_k| X_1, X_2, X_3,.....,X_n)= \frac{P(C_k)\prod_{i=1}^{n} P(X_i | C_k)}{\sum_{j=1}^{k} P(C_j)\prod_{i=1}^{n} P(X_i | C_j)} \quad (10)$$

Equation (10) is the fundamental equation of the NB classifier. If we are interested only in the most probable value of C_k, then we have the NB classification rule

$$C_k \leftarrow \underset{C_k}{arg\,max} \frac{P(C_k)\prod_{i=1}^{n} P(X_i \mid C_k)}{\sum_{j=1}^{k} P(C_j)\prod_{i=1}^{n} P(X_i \mid C_j)} \tag{11}$$

which simplifies to the following because the denominator does not depend on C_k

$$C_k \leftarrow \underset{C_k}{\arg\max} P(C_k)\prod_{i=1}^{n} P(X_i \mid C_k) \tag{12}$$

Thus the NB classifier combines this model with a decision rule. The decision rule for the NB classifier is defined as below:

$$classify(X_{1,} X_{2,}....,X_n) = \underset{C_k}{\arg\max}\, p(C_k)\prod_{i=1}^{n} P(X_i \mid C_k) \tag{13}$$

In this wok, we use the obtained feature vector set as the input in equation (13). In the training stage, $P(X_i|C_k)$ is estimated with respect to the training data. In the testing stage, based on the posterior probability $P(C_k|X_i)$, a decision whether a test sample belongs to a class C_k is made. For dataset IVa, C_k (k=1, 2) is treated as RH =-1 and RF =+1 and for the dataset IVb, C_k (k=1, 2) is considered as LH =-1 and RF =+1. Thus in this research, we achieve the detection results of each fold for each subject from the both datasets.

3.4. Performance evaluation

This study uses 10-fold cross-validation [22, 23, 33] process to assess the performance of the proposed approach. In 10-fold cross-validation procedure, a data set is partitioned into 10 mutually exclusive subsets (folds) of approximately equal size and the method is repeated 10 times. Each time, one of the folds is used as a test set and the other nine folds are put together to form a training set. In this research, the stability of the performance of the proposed method is assessed based on different statistical measures, such as accuracy, true positive rate or sensitivity and true negative rate or specificity as described in equations (14), (15) and (16) [34, 35, 36]. These statistical measures are calculated from each of the 10 folds. The overall performance of the proposed method is computed averaging the accuracy values across all 10 trials.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \tag{14}$$

Overall performance= $\frac{1}{10}\sum_{k=1}^{10} \text{accuracy}^k$ where accuracy^k is the accuracy of k^{th} fold (k=1,2,…..10).

$$\text{True positive rate (TPR)=Sensitivity=} \frac{TP}{TP + FN} \tag{15}$$

$$\text{True negative rate (TNR)=Specificity=} \frac{TN}{TN + FP} \tag{16}$$

True positive (T P): patterns correctly predicted as pertaining to the positive class

True negative (TN): patterns correctly predicted as belonging to the negative class
False positive (FP): patterns predicted as positive which come from negative class
False negative (FN): patterns predicted as negative whose true class is positive

In this research, for every subject of dataset IVa, we consider right hand (RH) class as positive class and right foot (RF) class as negative class. In dataset IVb, we consider left hand (LH) as positive class and right foot (RF) as negative class.

4. Experiments, results and discussions

This section presents experimental outcomes of the proposed optimum allocation based NB approach for two datasets: IVa and IVb of BCI Competition III and also provides a comparison of the present method with five recent reported methods for dataset IVa. As we did not find any research reports for the dataset IVb in the literature, we could not compare the experimental results with other methods. In this research, the experiments of the proposed method are performed on one single subject in the both datasets, separately, as the MI based EEG signals are naturally highly subject-specific depending on physical and mental tasks. All of the experimental works of this research are executed in MATLAB (version 7.14, R2012a). In this paper, all experimental results are presented based on the testing set.

4.1. Results for BCI III: Dataset IVa

Table 3 presents the accuracy for each of the 10 folds and the overall performances over the ten folds for dataset IVa. The overall performances for each subject are reported in terms of mean ± standard deviation (SD) of the accuracy over the ten folds. As shown in Table 3, most of the accuracy values for each of folds are close to 100. The overall performances for subjects, **aa**, **al**, **av**, **aw** and **ay** are 97.92%, 97.88%, 98.26%, 94.47% and 93.26%, respectively and the average of the performances for all subject is 96.36%. Table 3 also reports that there is no significant variation of the accuracies among the different folds which indicates the stability of the proposed method.

Table 3: Experimental outcomes for the proposed approach for dataset IVa of BCI Competition III

Subject	Accuracy for each of the 10 folds (%)										Overall performance
	1	2	3	4	5	6	7	8	9	10	
aa	83.33	100	100	100	95.83	100	100	100	100	100	97.92±5.29
al	95.83	95.83	100	100	100	95.83	95.83	100	95.45	100	97.88±2.24
av	95.83	100	95.83	100	100	100	100	100	100	90.91	98.26 ±3.11
aw	95.83	95.83	100	83.33	100	91.67	100	91.67	95.45	90.91	94.47 ±5.26
ay	79.17	95.83	95.83	87.5	100	100	91.67	91.67	95.45	95.45	93.26 ±6.24
Total											96.36±2.32

Fig.2 presents the pattern of the true positive rate (TPR) for each subject of dataset IVa. Here the TPR means the correctly detection rate for the RH class for this dataset. This figure shows the individual TPR against each of the 10-folds for the five subjects, **aa**, **al**, **av**, **aw** and **ay**. As can be seen

in Fig.2, the most of the values of the TPR for the proposed approach is close to 100 for each of the folds of each subject and the variations of the TPR among the 10-folds for each subject is not substantial that indicate the proposed approach is fairly stable.

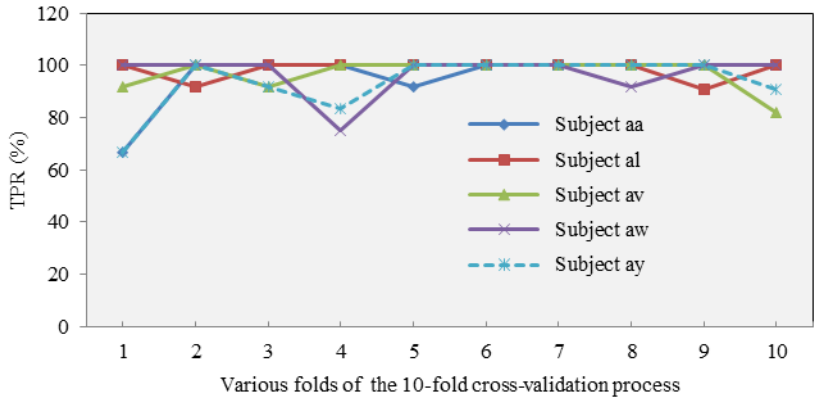


Fig.2. Patterns of the true positive rate (TPR) for each subject of dataset IVa

The contour of the true negative rate (TNR) for each of the five subjects is provided in Fig.3. For dataset IVa, the TNR refers to the correctly detection rate for the RF class. This figure displays the separate TNR for each of the ten folds of each of the five subjects. From Fig. 3, it is observed that the TNR in most of the folds of each subject is approximately 100% and there is no

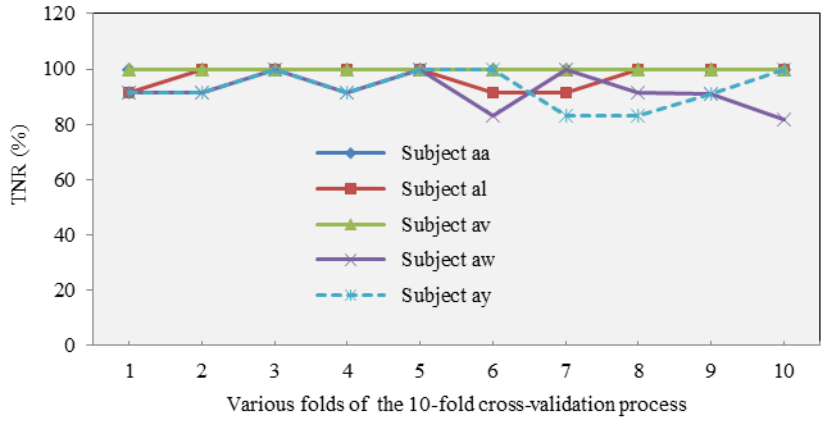


Fig.3. Patterns of the true negative rate (TNR) for each subject of dataset IVa

significant variation of the TNR among the ten folds of each subject. This indicates that the proposed method is reliable and robust. Along with Table 3, Fig.2 and Fig.3, it can be concluded that, although there is bit variability in performance over the subjects, generally the proposed approach provides higher performance for all of the subjects and it is consistent and fairly stable.

4.2. Results for BCI III: Dataset IVb

The experimental outcomes for the proposed optimum allocation based NB approach for dataset IVb are presented in Table. 4. As mentioned in Section 3.1, this dataset holds data for one male subject. This table displays individual accuracy rate of each of the ten folds of that subject and the overall performance of the proposed method in terms of mean ± standard deviation of the accuracy over ten

Table 4. Experimental outcomes for the proposed approach for dataset IVb of BCI Competition III

Subject	Accuracy for each of the 10 folds (%)										Overall performance
	1	2	3	4	5	6	7	8	9	10	
One healthy male	79.17	83.33	91.67	91.67	95.83	95.83	100	95.83	100	86.36	**91.97±7.02**

folds. As shown in the table, the method provides higher accuracy values for most of the folds and the variation among the different folds is not significant. The overall performance for this dataset is 91.97% and the standard deviation is 7.02%.

Fig.4 shows the pattern of TPR of each of the ten folds for a healthy male subject of dataset IVb. Here the TPR means the correctly detection rate for the LH class of this dataset. It can be seen from this figure that most of the values of the TPR lies in approximately 95 to 100. There is no significant difference of the TPR values among the ten folds that indicate the consistency of the proposed method.

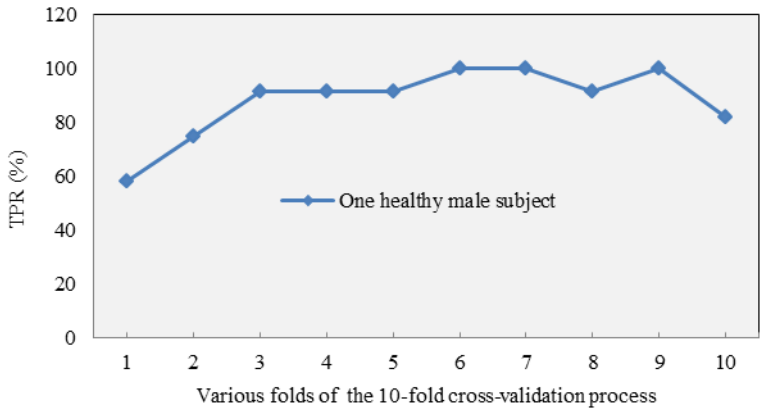


Fig. 4. Pattern of the true positive rate (TPR) for dataset IVb

The shape of the TNR of each of the ten folds for one healthy subject is illustrated in Fig.5. Here the TNR refers the correctly detection rate of the RF class for the dataset IVb. This figure demonstrates that the most of the values of the TNR is close to 100 and the variation among the TNR values of the ten folds is not substantial. This proves the reliability of the proposed approach. Thus, it is obvious from Table 4, Fig.4 and Fig.5 for dataset IVb that the proposed algorithm produces a good performance both individually and overall.

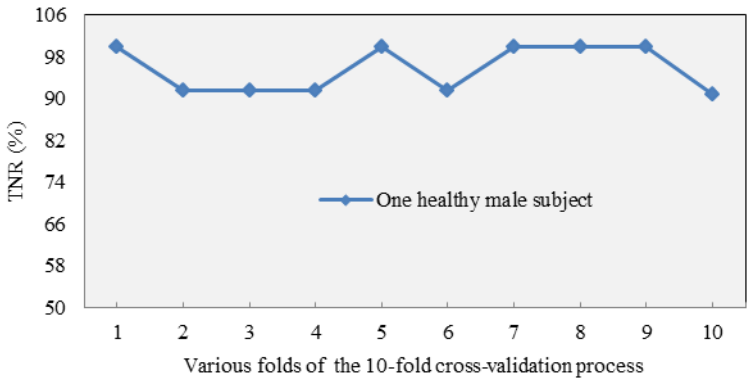


Fig.5. Pattern of the true negative rate (TNR) for dataset IVb

4.3. Comparison to previous work

To further examine the efficiency, this section provides a report for comparison between the proposed approach and five recently developed methods for dataset IVa. These five methods are already discussed in Section 2. As mentioned before, we cannot present the comparison results for dataset IVb as there were no reported research results available in literature. Table 5 presents the comparison results of the performance for the proposed method and the five existing algorithms for dataset IVa. This table shows the classification performance for each of the five subjects as well as the overall mean and the SD of the performances for all subjects. As shown in Table 5, the proposed method yields the most excellent performance of 97.92% and 93.26% for subjects, **aa** and **ay**, respectively. The performance of the proposed method for subject **av** is also very high (98.26%) that close to the highest performance (98.75%). The CC based LS-SVM [22] and Z-LDA method [21] provides better result for subjects **al** and **aw**, respectively. The highest mean for all five subjects is obtained by the proposed approach which is 96.36% and the SD value is the lowest (2.32%) as well. Therefore, it can be stated that generally the proposed approach significantly outperforms the five existing methods.

Table 5. A comparison report over five most recent reported methods for dataset IVa

Authors	Methods	Detection performance (%)						
		aa	**al**	**av**	**aw**	**ay**	Mean	SD
Proposed approach	OA & NB based approach	97.92	97.88	98.26	94.47	93.26	**96.36**	**2.32**
Suk and Lee [20]	BSSFO	79.46	94.64	57.65	91.96	53.57	75.46	19.06
Zhang et al. [21]	Z-LDA	77.7	100.0	68.4	99.60	59.9	81.1	18.2
Siuly and Li [22]	CC based LS-SVM	97.88	99.17	98.75	93.43	89.36	95.72	4.35
Siuly et al. [23]	CT based LS-SVM	92.63	84.99	90.77	86.50	86.73	88.32	3.22
Lu et al. [24]	R-CSP with aggregation	76.80	98.20	74.50	92.90	77.00	83.90	10.86

Further looking at the performance comparison in Table 5, it is noted that the proposed algorithm is ranked first in terms of the overall performance (96.36%), while the CC based LS-SVM method [22] comes second position (95.72%) and the CT based LS-SVM algorithm [23] is third (88.32%). The Bayesian spatio-spectral filter optimization algorithm [20] is the last (75.46%). The

results indicate that the proposed method achieves up to 20.90% improvements overall the five existing methods for dataset IVa of BCI competition III.

5. Conclusions

Although MI activities has emerged as the most useful for real-life BCIs, there are still some problems that make it a challenge to detect EEG signals of MI activities for the application of BCIs. In this paper, we propose an automatic approach that interprets how EEG signals are organised to detect different categories of MI tasks. Our proposed approach develops an optimum allocation based algorithm to determine representatives sample points from every group of the original data considering the minimum variation within each group. Then eleven statistical features are extracted from a group of samples points for a particular MI activity. After that, a probabilistic model, NB classifier is employed to detect different MI tasks based on extracted features. In our experiments on two public databases, IVa and IVb of BCI Competition III, the proposed method outperforms the state-of-the-art methods in terms of overall detection performance. The adoption of the optimum allocation technique with the NB resulted in an improvement of performance up to 20.90% compared to other five reported methods. The performance also show that two-class MI based EEG signals can be reliably identified using the proposed approach and this may be a promising avenue for robust EEG based BCI applications.

Acknowledgments

This work is partially supported by the Australian Research Council (ARC) Linkage Project (LP100200682) and Discovery Project (DP140100841).

References

[1] Pfurtscheller, G. and Neuper, C. (2001) 'Motor imagery and direct brain-computer communication' (Invited paper), *Proc. IEEE (Special Issue) Neural Eng. Merging Eng. Neurosci.*, Vol. 89, 1123–1134.

[2] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., and Vaughan, T.M. (2002) 'Brain-computer interfaces for communication and control', *Clin. Neurophysiol.*, Vol. 113, 767–791.

[3] Faller, J., Vidaurre, C., SolisEscalante, T., Neuper, C. andS cherer, R. (2012) 'Auto calibration and recurrent adaptation: towards a plug and play online ERD-BCI.',*IEEE Trans. Neural Syst. Rehabil. Eng.* Vol. 20, 313–319.

[4] Neuper, C., Müller, G., Kübler, A., Birbaumer, N. and Pfurtscheller, G. (2003) 'Clinical application of an EEG-based brain-computer interface, a case study in a patient with severe motor impairment', *Clin. Neurophysiol.* Vol.114, 399–409.

[5] Long, J, Li, Y. and Yu, Z. (2010) 'A semi-supervised support vector machine approach for parameter setting in motor imagery-based brain computer interfaces', *Cogn Neurodyn*., Vol. 4, 207–216.

[6] Birbaumer, N., Murguialday, A.R. and Cohen, L. (2008) 'Brain-computer interface in paralysis', *Curr.Opin. Neurol.* Vol. 21, 634–638.

[7] Kaiser,V., Daly, I., Pichiorri, F., Mattia, D., Müller-Putz, G.R. and Neuper, C. (2012) 'Relationship between electrical brain responses to motor imagery and motor impairment in stroke', *Stroke* 43, 2735–2740.

[8] Alkadhi, H., Brugger, P., Boender- maker, S.H., Crelier, G., Curt, A., Hepp-Reymond, M.C., et al. (2005) 'What disconnection tells about motor imagery: evidence from paraplegic patients', *Cereb. Cortex,* Vol. 15, 131–140.

[9] Pfurtscheller G., Neuper C, Muller G.R., Obermaier B., Krausz G., Schlogl A., Scherer R., Graimann B., Keinrath C., Skliris D., Wortz M., Supp G., Schrank C. (2003) 'Graz-BCI: state of the art and clinical applications', *IEEE Trans Neural Syst Rehab Eng*., Vol. 11, no. 2,1–4.

[10] Kauhanen, L., Nykopp, T., Lehtonen, J., Jylanki, P., Heikkonen, J., Rantanen, P., Alaranta, H. and Sams, M. (2006) 'EEG and MEG brain-computer interface for tetraplegic patients', *IEEE Trans. Neural Syst. Rehabil. Eng.* Vol. 14, 190–193.

[11] Kronegg, J., Chanel, G., Voloshynovskiy, S. and Pun, T. (2007) 'EEG-based synchronized braincomputer interfaces: A model for optimizing the number of mental tasks', *IEEE Trans. Neural Syst. Rehabil. Eng.*, Vol. 15**,** 50–58.

[12] Ahangi, A., Karamnejad, M., Mohammadi, N., Ebrahimpour, R., and Bagheri, N. (2013) 'Multiple classifier system for EEG signal classification with application to brain–computer interfaces', *Neural Comput & Applic.,Vol.* 23,1319–1327.

[13] Pfurtscheller, G., Muller-putz, G. R. , Schlogl, A., Graimann, B., Scherer R. and Leeb, R. (2006) '15 years of BCI research at Graz university of technology: current projects', *IEEE Trans. Neur. Syst. Rehab. Eng.,* Vol. 14, no. 2, 205-210.

[14] McFarland, D. J., Sarnacki W. A. and Wolpaw, J. R. (2003) 'Brain-computer interface (BCI) operation: optimizing information transfer rates', *Biol. Psycho.*, Vol. 63, pp. 237-251.

[15] Mason, S.G., Birch, G.E. (2003) 'A general framework for brain–computer interface design', *IEEE Trans Neural Syst Rehab Eng. Vol.*11, no.1, 70–85.

[16] Mitchel, T., *Machine Learning*, McGraw-Hill Science, 1997.

[17] Wiggins, M., Saad, A., Litt, B. and Vachtsevanos, G. (2011) 'Evolving a Bayesian Classifier for ECG-based Age classification in Medical Applications', *Appl Soft Comput*, Vol. 8, no. 1, 599-608

[18] BCI competition III, http://www.bbci.de/competition/iii

[19] Blankertz, B., Muller, K. R., Krusierski, D. J., schalk, G., wolpaw, J. R., Schlgl, A., Pfurtscheller,

G. and Birbaumer, N. (2006) 'The BCI competition III: validating alternative approaches to actual BCI problems', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 14, no. 2, 153-159.

[20] Suk, H. and Lee, S.W. (2013) 'A Novel Bayesian framework for Discriminative Feature Extraction in Brain-Computer Interfaces', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, no. 2, 286-299.

[21] Zhang, R., Xu, P., Guo, L., Zhang, Y., Li, P. and Yao, D. (2013) 'Z-Score Linear Discriminant Analysis for EEG Based Brain-Computer Interfaces', *PLoS ONE*, Vol. 8, no. 9, e74433.

[22] Siuly and Li, Y. (2012) '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, 526-538.

[23] Siuly, Y. Li, and Wen P. (2011) 'Clustering technique-based least square support vector machine for EEG signal classification', *Computer Methods and Programs in Biomedicine*, Vol. 104, 358-372.

[24] Lu, H., Eng, H.L. , Guan, C., Plataniotis, K.N. and Venetsanopoulos, A.N. (2010) 'Regularized common spatial patterns with aggregation for EEG classification in small-sample setting', *IEEE Transactions on Biomedical Engineering*, Vol. 57, 2936-2945.

[25] Siuly and Y. Li, (2014) 'A novel statistical framework for multiclass EEG signal classification', *Engineering Applications of Artificial Intelligence*, Vol. 34,154–167.

[26] Cochran, W.G., *Sampling Techniques*, Wiley, New York, 1977.

[27] Islam, M.N., *An Introduction to Sampling Methods: Theory and Applications*, Book World, Dhaka, 2007.

[28] Islam, M. N., *An introduction to statistics and probability*, 3rd ed., Mullick & brothers, Dhaka New Market, Dhaka-1205, pp. 160-161, 2004.

[29] De Veaux, R. D., Velleman, P.F. and Bock, D.E., *Intro Stats*, 3rd ed., Pearson Addison Wesley, Boston, 2008.

[30] Siuly, Y. Li and P. Wen, (2014) '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, 767-780.

[31] Richard, D.G.S., Duda, O., Hart, P.E., *Pattern classification*, 2nd edn. Wiley, New York, 2000.

[32] Bhattacharyya, S. et al. (2011) 'Performance Analysis of Left/Right Hand Movement Classification from EEG Signal by Intelligent Algorithms', *Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB) IEEE Symposium*, 2011.

[33] Siuly, Li, Y. and Wen, P. (2014) 'Comparisons between Motor Area EEG and all-Channels EEG for Two Algorithms in Motor Imagery Task Classification', *Biomedical Engineering: Applications, Basis and Communications (BME)*, Vol. 26, no. 3, 1450040 (10 pages).

[34] Siuly, Li, Y. and Wen, P., (2011) 'EEG signal classification based on simple random sampling technique with least square support vector machines', *International journal of Biomedical*

Engineering and Technology, Vol. 7, no. 4, 390-409.

[35] Gu, Q., Zhu, L. and Cai, Z. (2009) 'Evaluation measures of the classification performance of imbalanced data sets', *ISICA 2009, CCIS 51*, pp.461-471.

[36] Siuly, Li, Y. and Wen, P. (2013) 'Detection of Motor Imagery Tasks through CC-LR Algorithm in Brain Computer Interface', *International Journal of Bioinformatics Research and Applications*, Vol. 9, no. 2, 156-172.