Improving the Separability of Motor Imagery EEG Signals Using a Cross Correlation-Based Least Square Support Vector Machine for Brain–Computer Interface

Siuly Siuly and Yan Li, *Member, IEEE*

***Abstract*—Although brain–computer interface (BCI) techniques have been developing quickly in recent decades, there still exist a number of unsolved problems, such as improvement of motor imagery (MI) signal classification. In this paper, we propose a hybrid algorithm to improve the classification success rate of MI-based electroencephalogram (EEG) signals in BCIs. The proposed scheme develops a novel cross-correlation based feature extractor, which is aided with a least square support vector machine (LS-SVM) for two-class MI signals recognition. To verify the effectiveness of the proposed classifier, we replace the LS-SVM classifier by a logistic regression classifier and a kernel logistic regression classifier, separately, with the same features extracted from the cross-correlation technique for the classification. The proposed approach is tested on datasets, IVa and IVb of BCI Competition III. The performances of those methods are evaluated with classification accuracy through a 10-fold cross-validation procedure. We also assess the performance of the proposed method by comparing it with eight recently reported algorithms. Experimental results on the two datasets show that the proposed LS-SVM classifier provides an improvement compared to the logistic regression and kernel logistic regression classifiers. The results also indicate that the proposed approach outperforms the most recently reported eight methods and achieves a 7.40% improvement over the best results of the other eight studies.**

***Index Terms*—Brain–computer interface (BCI), cross-correlation technique, electroencephalogram (EEG), feature extraction, kernel logistic regression, least square support vector machine (LS-SVM), logistic regression, motor imagery (MI).**

I. Introduction

RECENTLY, electroencephalogram (EEG)-based brain–computer interfaces (BCIs) have become popular in the study of brain science, neural engineering and rehabilitation [1]. BCIs are devices that utilize the nonmuscular channels of the brain (e.g., EEG) for communication and control. A BCI provides a direct communication interface between a brain and an external device in which a message sent by an individual cannot pass through the brain's normal output pathways but is detected through brain signals. Brain activities produce electrical signals detectable on the scalp on the cortical surface, or within the brain. BCIs translate these signals into outputs that allow total lock-in patients suffering from brain or spinal cord injury to communicate without the participation of peripheral nerves and muscles via *thoughts* alone [2]. BCIs convert human intentions or thoughts into control signals to establish a direct communication channel between the human brain and output devices, as presented in Fig. 1. This figure depicts the basic structure of BCI technologies, on how a brain signal of a thought is passing on to a BCI system, and how the BCI system processes those signals into a control signal for a user application.

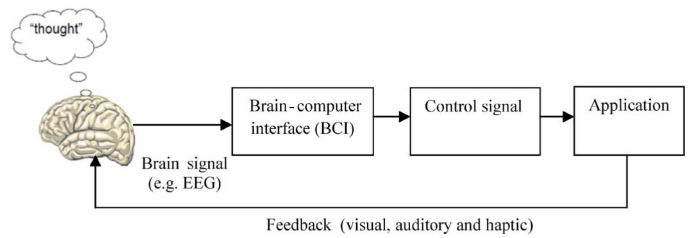

Fig. 1. Fundamental structure of brain–computer interface (BCI).

A class of EEG-based BCIs which rely on motor imagery (MI) of users are of particular interest to the BCI community since this type of BCI has a relatively robust communication performance and can help to understand the underlying mechanism of MI [3]. 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. There are various acquisition techniques for capturing MI brain activities. Among these techniques, EEG is the most studied measure of potential for noninvasive BCI designs, mainly due to its excellent temporal resolution, noninvasiveness, usability, and low set-up costs [4]–[6].

As shown in Fig. 1, a BCI system works through EEG brain signals. A big challenge, therefore, is for BCI systems to correctly and efficiently identify different EEG signals of different MI tasks using appropriate classification algorithms. In most current MI based BCIs, machine learning algorithms are carried out in two stages: feature extraction and feature classification

Manuscript received March 01, 2011; revised June 11, 2011 and November 22, 2011; accepted January 04, 2012. Date of publication January 23, 2012; date of current version July 03, 2012.

The authors are with the Centre for Systems Biology, Department of Mathematics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: siuly@usq.edu.au; siuly_1976@yahoo.com; liyan@usq.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNSRE.2012.2184838

1534-4320/$31.00 © 2012 IEEE

[7]. Because of many factors, such as low topographical resolutions and high noise levels, it has been a key issue to extract effective features from the EEG signals and perform the classification [8]. Although many methods have been reported for feature extraction as well as classification to produce impressive results in BCI applications as discussed in Section II, nevertheless a MI-based BCI system is not satisfactory due to the lack of classification accuracy.

The goal of the study is to improve the classification accuracy of MI data for BCI systems and also to investigate whether a cross-correlation technique is an appropriate method for feature extraction in an EEG-based MI data. For this purpose, the present study proposes a novel algorithm where a cross-correlation technique is developed for feature extraction and a least square support vector machine (LS-SVM) is employed for classifying the obtained features. The performance of the LS-SVM classifier is compared with a popular logistic regression classifier and a kernel logistic regression classifier for the same feature vector set. In order to further verify the effectiveness of the proposed cross-correlation-based LS-SVM algorithm, we also compare it with the eight most recently reported methods in the literature.

There are strong grounds of using a cross-correlation technique for feature extraction from MI EEG data in this study. The recorded multichannel EEG signals are highly correlated and the different signals from different scalp sites do not provide the same amount of discriminative information [9]. These signals are also typically very noisy and are not directly usable in BCI applications. Cross-correlation is a very powerful technique to identify the relationship between the EEG signals from two different electrodes and is also able to provide the discriminative information about those signals. This technique produces a new signal called cross-correlogram (mathematically called cross correlation sequence) using two signals. If the two EEG signals have the same rhythm, the peak of the cross-correlogram curve will appear in the centre. In addition, the cross correlation can diminish noise from the EEG signals by means of correlation calculation because of the characteristics of signal periodicity. Hence the cross-correlogram is a nearly noise-free signal that can provide more signal information compared to the original signal [10]. The process also takes into consideration any potential phase differences between the two signals via the inclusion of a lead or lag term [11]. Thus, a cross-correlation technique works better for the feature extraction from the MI EEG data. In order to reduce the dimensionality of the cross-correlation sequences, six statistical features, *mean*, *median*, *mode*, *standard deviation*, *maximum* and *minimum* values, are extracted from each sequence as discussed in details in Section III-B3. These features represent the characteristics of the original MI EEG data without redundancy. The extracted features are then used as the inputs to the LS-SVM and also to the logistic regression and the kernel logistic regression classifiers. To the best of our knowledge, such a cross-correlation technique has not been used on the MI data for feature extraction so far.

The LS-SVM is a robust intelligent technique for classification in BCI applications. It has the advantage over other techniques of converging to a global optimum, not to a local optimum that depends on the initialization or parameters affecting the rate of convergence. The computation of the LS-SVM is faster compared with other machine learning techniques because there are fewer random parameters and only the support vectors are used in the generalization process [12]. In spite of its advantages, this method has not been applied for the MI tasks classification in BCI development except the clustering technique based algorithm [13]. But, in the algorithm [13], the hyper parameters of the LS-SVM were not selected optimally through a technique. It is well known that the parameters of the LS-SVM play an important role in affecting the classification performance. To obtain more reliable results, this study, instead of manual selection, employs a LS-SVM for the MI EEG signal classification where the hyper parameters of the LS-SVM are chosen optimally using a two-step grid search algorithm.

The proposed approach is evaluated on datasets, IVa and IVb of BCI Competition III, where both sets contain MI EEG recorded data. In this paper, a 10-fold cross-validation method is used for assessing the performance of the proposed method. This procedure divides the feature vector sets into 10 approximately equal-sized distinct partitions. One partition is then used for testing, whilst other partitions are used for training the model. To further improve the estimate, the procedure is repeated ten times and all accuracy rates over these 10 runs are averaged. The average accuracy over the ten runs obtained from the test data is taken as the performance evaluation criteria in this study. Experimental results show that the proposed LS-SVM classifier achieves a better performance compared to the logistic regression and kernel logistic regression for the same cross-correlated features. The results also demonstrate that the proposed approach outperforms the other eight most recently reported methods with respect to the classification performance for dataset IVa.

The rest of the paper is organized as follows. Section II presents a review of the existing classification techniques. Section III, after briefly describing the MI EEG datasets used in this study, presents the description of the proposed approach and the performance evaluation method. The experimental results and discussions are provided in Section IV. Finally Section V draws the conclusions of the study.

II. Review of the Existing Classification Techniques

A variety of methods have been reported by different researchers using dataset IVa of BCI Competition III. A brief discussion of some recent work is provided below. So far there is no classification results reported for dataset IVb.

Yong *et al.* [14] reported a sparse spatial filter optimization for EEG channel reduction in the brain computer interface, where the spatial filter was used to project the signals and the variance of the projected signals was the only feature used in the linear discriminant analysis (LDA) as the input for the classification. They reported a classification accuracy of 57.5% for subject $\mathbf{aa}$, 86.9% for subject $\mathbf{al}$, 54.4% for subject $\mathbf{av}$, 84.4% for subject $\mathbf{aw}$ and 84.3% for subject $\mathbf{ay}$, and the average accuracy was 73.5% using a 10-fold cross validation. But the major limitation of that study was that they manually selected their regularization parameter.

Lu *et al.* [15] 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. The reported classification accuracy rates were 69.6%, 83.9%, 64.3%, 70.5%, 82.5% for subject $\mathbf{aa}$, $\mathbf{al}$, $\mathbf{av}$, $\mathbf{aw}$, $\mathbf{ay}$, respectively. 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.

Lotte *et al.* [16] proposed four methods representing a family of a theoretical framework based on regularized common spatial patterns (RCSP). Their proposed methods are regularized CSP with selected subjects (SSRCSP) [16], CSP with Tikhonov regularization (TRCSP) [16], CSP with weighted Tikhonov regularization (WTRCSP) [16] and spatially regularization (SRCSP) [16]. It was reported that their methods can perform efficiently subject-to-subject transfer for classifying MI data in BCIs. In particular, the TRCSP and WTRCSP algorithms are better than the other two algorithms. The average classification success rate reached at 73.56% for the SSRCSP, 77.98% for the TRCSP, 75.93% for the WTRCSP and 78.63% for the SRCSP. All their four algorithms are based on a common spatial patterns (CSP) method. Although the CSP is a popular method in BCI applications, it is very sensitive to noise, and often over-fits with small training sets.

Lu *et al.* [17] introduced a regularization and aggregation technique with CSP for EEG signal classification in a small sample setting (SSS). To tackle the problem of regularization parameter determination, a number of R-CSPs were aggregated to give an ensemble-based solution. The parameter determination problem was solved through a cross-validation procedure. The cross-validation method was employed to determine the regularization parameters of the R-CSP for the EEG signal classification in SSS. The obtained classification accuracy rates were 76.8%, 98.2%, 74.5%, 92.9%, and 77.0% for subject $\mathbf{aa}$, $\mathbf{al}$, $\mathbf{av}$, $\mathbf{aw}$, $\mathbf{ay}$, respectively, for experiment III. The overall accuracy performance was 83.9%.

Most recently, Siuly *et al.* [13] reported a clustering technique-based least square support vector machine algorithm (LS-SVM) for EEG signal classification. They developed a clustering technique for feature extraction and the obtained features were used to the LS-SVM as the inputs for classification. It employed the 10-fold cross-validation method to evaluate the performance and achieved the classification accuracy of 92.63% for subject $\mathbf{aa}$, 84.99% for subject $\mathbf{al}$, 90.77% for subject $\mathbf{av}$, 86.50% for subject $\mathbf{aw}$ and 86.73% for subject $\mathbf{ay}$. The average accuracy performance 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.

From the discussion of the literature, it is observed that most of the reported methods are limited in their success and effective only in a small sample setting. In most of the cases, the methods did not select their parameters using a suitable technique while the parameters significantly affect the classification performance. To overcome these problems, this paper proposes a new approach which can discriminate two-class MI tasks for the development of BCI systems. In the proposed algorithm, a cross-correlation technique is developed for feature extraction and a LS-SVM is applied to classify the obtained features. To the best of our knowledge, the cross-correlation technique and the LS-SVM have not been used together before for the MI task recognition in BCI applications. This study employs a two-step grid search algorithm for selecting the optimal combinations of the parameters for the LS-SVM.

III. Data and Method

In this section, first the data acquisition used in this research is described. Next, the proposed method and its implementation on the experimental data are presented. Finally, the performance evaluation procedure is discussed for the proposed algorithm.

A. Data Acquisition

This study uses two publicly available datasets, IVa and IVb provided in the BCI Competition III [18], [19] to test the effectiveness of the proposed method. Both datasets consist of EEG recording data collected during MI tasks.

Dataset IVa [18], [19] was recorded from five healthy subjects (labelled $\mathbf{aa}$, $\mathbf{al}$, $\mathbf{av}$, $\mathbf{aw}$, $\mathbf{ay}$) who performed right-hand (class 1) and right-foot (class 2) MI tasks. 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. During each trial, the subject was required to perform either of the two (right hand and right foot) MI tasks for 3.5 s. A training set and a testing set consisted of different sizes for each subject. Among 280 trials, 168, 224, 84, 56, and 28 trials composed the training set for subject $\mathbf{aa}$, $\mathbf{al}$, $\mathbf{av}$, $\mathbf{aw}$, $\mathbf{ay}$, respectively, and the remaining trials composed the test 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 and right-foot MI. Signals were recorded from 118 channels in 210 trials. 118 EEG channels were measured at the positions of the extended 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.

B. Proposed Method

The present study develops an algorithm that can automatically classify two categories of MI EEG signals in BCI systems. The proposed cross-correlation-based LS-SVM scheme for the MI signals classification is illustrated in Fig. 2. The approach employs a cross-correlation technique to extract representative features from the original signals, and then the extracted features are used as the inputs to the LS-SVM classifier.

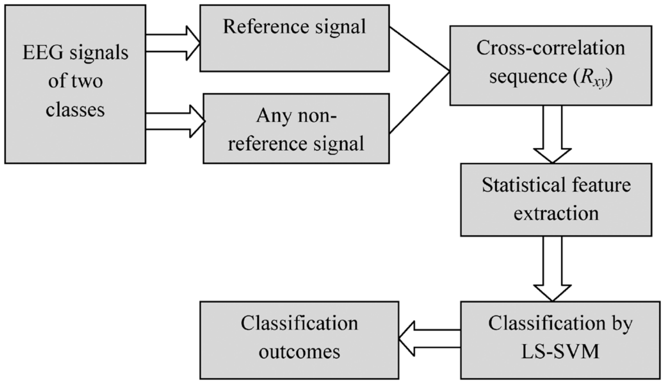

Fig. 2. Block diagram of the proposed technique for the MI EEG signal classification in BCIs development.

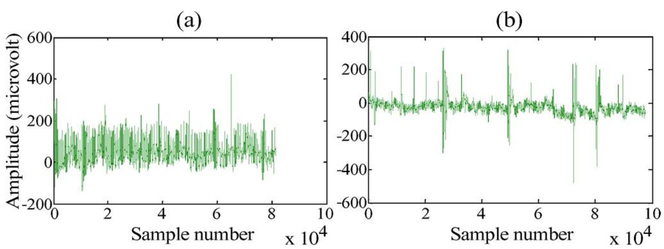

Fig. 3. Typical reference signals: (a) Dataset IVa and (b) Dataset IVb.

In order to evaluate the performance of the LS-SVM classifier, we test a logistic regression classifier and a kernel logistic regression classifier, separately. They also employ the same features extracted from the cross-correlation method as the inputs. The block diagram of the proposed method in Fig. 2 depicts the procedure for the MI EEG signal classification as described in the following steps.

Reference Signal Selection: One signal is selected as a reference signal among all channel signals in one subject of the two-class MI tasks. A reference signal should be noiseless as the signal with the noise will be incoherent with anything in the reference. In this work, any other signal, that is not a reference signal, is treated as a non-reference signal.

In this study, we use two datasets, IVa and IVb of BCI Competition III. Dataset IVa consists of MI tasks EEG signals from the right-hand class and the right-foot class. Dataset IVb consists of MI tasks EEG signals of the left-hand class and the right-foot class. Both datasets contain 118 channel data in each class of a subject. We consider the electrode position Fp1 in the international 10/20 system as the reference signal for the cross-correlation technique for all the subjects.

For dataset IVa, the Fp1 is selected from the right-hand class while it is from the right-foot class for dataset IVb. Fig. 3(a) and (b) shows the typical reference signals of subject $\mathbf{aa}$ for dataset IVa and dataset IVb, respectively.

Computation of a Cross-Correlation Sequence: A cross-correlation sequence, denoted by "R_{xy}," is calculated recursively using a reference signal and any one of other nonreference signals, using the cross-correlation technique as shown in Fig. 2. In this study, the following (1) of the cross-correlation method [20], [21] is used to compute a cross-correlation sequence, where $x[i]$ is considered as the reference signal and $y[i]$ is regarded as any other nonreference signal in one subject of the two-class MI EEG data

$$\begin{aligned} R_{xy}[m] = & \sum_{i=0}^{N-|m|-1} x[i]y[i-m]; \\ & m = -(N-1), -(N-2), \\ & \ldots 0, 1, 2, 3 \ldots (N-2), (N-1). \end{aligned} \tag{1}$$

Here, N $(N > 1)$ is the number of sample points, m represents time-shift parameters known as lag, and R_{xy} is the cross-correlated sequence. As each of the signals, $x[i]$ and $y[i]$, consists of N finite number of samples, the resultant cross-correlation sequence has $(2N-1)$ samples. The graphical presentation of a cross-correlation sequence is called a cross-correlogram. The reference signal of a class is cross-correlated with the data of the remaining signals of this class and the data of all signals of another class. If we have two classes of EEG signals, and class 1 has n signals and class 2 has m signals, and a reference signal is chosen from class 1, then a total of $(n-1)$ cross-correlation sequences are obtained from class 1 and a total of m cross-correlation sequences from class 2.

As there are 118 signals in each of the two classes of a subject in datasets, IVa and IVb, in each subject of both datasets, the reference signal is cross-correlated with the data from the remaining 117 signals of the reference signal class. This reference signal is also cross-correlated with the data of all 118 signals of the nonreference signal class. Thus, for each subject, a total of 117 cross-correlation sequences/cross-correlograms are obtained from the reference signal class and 118 from the nonreference signal class. For example, in subject $\mathbf{aa}$, the signal of the Fp1 channel is a reference signal, which comes from the right-hand class and this reference signal is cross-correlated with the data from the remaining 117 signals of the right-hand class. In the right-foot class of subject $\mathbf{aa}$, this reference signal is also cross-correlated with the data of all 118 signals of this class. Thus, for subject $\mathbf{aa}$, a total of 117 cross-correlation sequences/cross-correlograms are obtained from the right-hand class and 118 from the right-foot class. The same process is followed for subjects $\mathbf{al}$, $\mathbf{av}$, $\mathbf{aw}$ and $\mathbf{ay}$ and the subject of dataset IVb in this study.

Fig. 4 presents typical signals of the right-hand and the right-foot MI data for subject $\mathbf{aa}$of dataset IVa. The typical cross-correlograms for the right-hand and the right-foot MI signals of the same subject are also shown in Fig. 4. The cross-correlogram of the right-hand signal is obtained using the reference signal and the right-hand MI signal and the cross-correlogram of the right-foot signal is acquired using the reference signal and the right-foot MI signal as depicted in Fig. 4.

Fig. 5 shows typical signals of dataset IVb for the right-foot MI and the left-hand MI. This figure also presents typical results of the cross-correlation for the right-foot MI signal and the left-hand MI signal. As shown in Fig. 5, the cross-correlogram of the right-foot MI signal is obtained using the reference signal and the right-foot MI signal and the left-hand cross-correlogram is generated by the reference signal and the left-hand MI signal.

It is known that if two curves have exactly the same shape, this means, they are highly cross-correlated with each other and cross-correlation is around 1. From Figs. 4 and 5, one can see

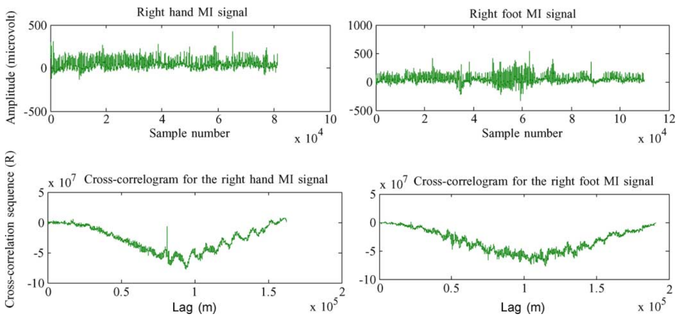

Fig. 4. Typical right-hand and right-foot MI signals and their respective cross-correlograms for subject **aa** in dataset IVa.

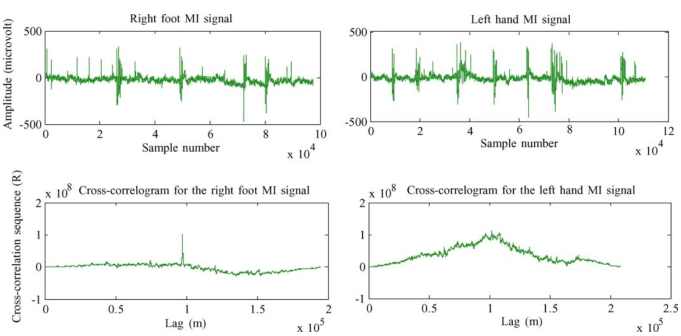

Fig. 5. Typical right-foot and left-hand MI signals and their respective cross-correlograms for dataset IVb.

that the shapes of the two curves are not exactly same, which indicates the statistical independency. That means, there is more of a chance to achieve better separation.

As mentioned earlier, if each of $x[i]$ and $y[i]$ signals has a finite number of samples N, the resulting cross-correlation sequence has $(2N-1)$ samples. Hence, in each of Figs. 3, 4, and 5, the scale for each signal is shown over the range of 0 to 10×10^4 samples and the scale for the corresponding cross-correlogram is shown over the range of 0 to 2×10^5 samples. The cross-correlogram signals convey greater signal information and consist of low level noises compared to the original signal. It is worthy to mention that a cross-correlogram contains information about the frequencies which are common to both waveforms, one of which is usually the signal and the other a reference wave [11]. Six statistical features, *mean*, *median*, *mode*, *standard deviation*, *maximum* and *minimum* are extracted from each cross-correlogram as discussed in the following section.

Statistical Feature Extraction: To reduce the dimensions of the cross-correlation sequences, this study considers six statistical features, *mean*, *median*, *mode*, *standard deviation*, *maximum* and *minimum* as the feature representatives ideally containing all important information of the original signal patterns. These features are calculated from each cross-correlation sequence or cross-correlogram to create feature vector sets. The six traits of the cross-correlation sequences are found to serve as important indicators of the neurological state of the subjects [10], [22].

There are strong reasons for the considerations of these six quantitative descriptors in this research. *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 and 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 [23], [24]. These three features give a fairly good idea about the nature of the data (shows the "middle value"), especially when combined with measurements on how the data is distributed. *Standard deviation* describes how observations in a distribution are spread out around a typical value (mean). The *standard deviation* is the average distance between the actual data and the mean. This feature gives information about the spread of data how close the entire set of data is to the average value in the distribution. Data sets with a small *standard deviation* have tightly grouped, precise data. Data sets with large *standard deviations* have data spread out over a wide range of values. *Maximum and minimum* values are used to describe the range of observations in the distribution. Therefore, *mean, median, mode, standard deviation, maximum and minimum* values are considered as the most valuable parameters for representing the characteristics of the MI EEG signals and thus for representing the brain activities as a whole in this paper.

In this study, we obtain 117 cross-correlation sequences from the reference signal class and 118 from the nonreference signal class for a subject in both datasets. We calculate the mentioned six features from each cross-correlation sequence. For example, subject **aa** contains 117 cross-correlation sequences for the right-hand class (reference signal class) and 118 cross-correlation sequences for the right-foot class (nonreference signal class). As we calculate the six features from each cross-correlation sequence, so we obtain 117 feature vectors of six dimensions from the right-hand class and 118 features vectors of the same dimensions from the right-foot class for subject **aa**. Thus we acquire a total of 235 feature vectors with six dimensions for this subject. We follow the same process for the other subjects in both datasets. We use MATLAB "mean," "median," "std," "max," "min" function for calculating *mean*, *median*, *standard deviation*, *maximum* and *minimum* values, respectively, from each cross-correlation sequence. In the *mode* calculation, we compute a histogram from a cross-correlation sequence and then the peak of the histogram is considered as an estimate of the *mode* for that cross-correlation sequence. These feature vector sets are divided into a training set and a testing set using a 10-fold cross validation method, which is discussed in Section III-C. These feature vectors are inputs for the LS-SVM and also for the logistic regression and the kernel logistic regression classifiers in the classification stage.

Classification: This study employs the LS-SVM with radial basis function (RBF) kernel as a classifier to distinguish the features obtained from the cross-correlation technique. The decision function of the LS-SVM in (2) is derived directly from solving a set of linear equations [25]–[27]. The details of the LS-SVM algorithm could be found in reference [27]

$$y(x) = sign\left(\sum_{i=1}^{n} y_i \alpha_i K(x, x_i) + b\right) \tag{2}$$

where x_i is the ith input feature vector of d dimensions, y_i (which is either $+1$ or -1) is the class label of x_i, n is the number of feature vectors, b is the bias term, α_i $(i = 1, 2, \ldots, n)$ denote Lagrange multipliers called support values, and $K(x, x_i)$ is the RBF kernel defined [27], [28] as $K(x, x_i) = \exp(-(\|x - x_i\|)^2/2\sigma^2)$.

In this study, the obtained training feature vector of six dimensions are used as the input in (2) to train the LS-SVM classifier and the testing feature vector sets are employed to verify the performance and the effectiveness of the trained LS-SVM for the classification of two-class of EEG signals in the both datasets. For dataset IVa, y_i is treated as $\mathrm{right\ foot} = +1$ and $\mathrm{right\ hand} = -1$ and for the dataset IVb, y_i is considered as $\mathrm{right\ foot} = +1$ and $\mathrm{left\ hand} = -1$. As the result of the LS-SVM lies largely on the choice of a kernel, the RBF kernel is chosen after many trials. The two important parameters (γ, σ^2) of the LS-SVM are selected by a two-step grid search technique for getting reliable performance of the method, that is discussed in Section IV-A. The solution of (2) provides the prediction results that directly assign the samples with a label $+1$ or -1 to identify that which category it belongs to.

To compare the performance of the proposed LS-SVM classifier, we employ the logistic regression classifier instead of the LS-SVM for the same feature sets as its inputs. In this study, the logistic regression [29], [30] in (3) is applied for the classification of the MI features. The detailed description of the logistic regression is available in [30]

$$\mathrm{P}(y = 1|x_1, x_2, \ldots, x_n) = \pi = \frac{e^{\beta_0 + \sum_{i=1}^{n} \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^{n} \beta_i x_i}} \tag{3}$$

where n is the number of feature vectors and $x_1, x_2, \ldots, x_n$ are feature vectors, and y is its class label, either 0 or 1. Here $x_1, x_2, \ldots, x_n$ are treated as independent/predictor variables and y is a dependent variable. Under the logistic regression framework, the probability of the dependent variable y, when y belongs to class 1, is defined in (3) as [31]. Here π is a conditional probability of the form $\mathrm{P}(y = 1|x_1, x_2, \ldots, x_n)$. On the other hand, the probability of y, when y belongs to class 0, can be calculated as $1 - \pi = 1 - \mathrm{P}(y = 1|x_1, x_2, \ldots, x_n) = \mathrm{P}(y = 0|x_1, x_2, \ldots, x_n)$. In (3), β_0 is an intercept and $\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficient related to the independent variables $x_1, x_2, \ldots, x_n$. These parameters are estimated by maximum likelihood estimation (MLE) [32].

In this work, we consider the mentioned six features of a feature vector set (training/testing) as the six input variables ($x_1 = mean$ values, $x_2 = maximum$ values, $x_3 = minimum$ values, $x_4 = standard\ deviation$ values, $x_5 = median$ values and $x_6 = mode$ values.) in (3) for the both datasets. We treat the dependent variable y as $\mathrm{right\ hand} = 0$ and $\mathrm{right\ foot} = 1$ for dataset IVa; and $\mathrm{right\ foot} = 0$ and $\mathrm{left\ hand} = 1$ for dataset IVb. Finally we obtain the prediction results that directly provide the class label 0 or 1 with the samples.

We also compare the performance of the LS-SVM with a kernel logistic regression classifier. Kernel logistic regression is a nonlinear form of logistic regression. It can be achieved via the so-called "kernel trick" which has the ability to classify data with nonlinear boundaries and also can accommodate data with very high dimensions. A detailed description of the kernel logistic regression is available in [40] and [41]. A final solution of the kernel logistic regression could be achieved using the following equation:

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x) + b \tag{4}$$

where x_i represents ith input feature vector of d dimensions, n is the number of feature vectors and b is the model parameter. The vector α_i contains the parameters which define decision boundaries in the kernel space and $K(x_i, x)$ is a kernel function. The most commonly used kernel in practical applications is the radial basis function (RBF) kernel defined as $K(x_i, x) = \exp(-(\|x_i - x\|)^2/2\sigma^2)$ which is also used in this study. Here σ is a kernel parameter controlling the sensitivity of the kernel. The parameters of this method are automatically estimated by iteratively re-weighted least square procedure [41].

In the kernel logistic regression, we utilize the feature vectors and their class labels as the same process of the logistic regression for the inputs. Finally, we acquire the output of the kernel logistic regression as an estimate of a posterior probability of the class membership. In Section IV, the classification results of these three classifiers are presented for datasets IVa and IVb. The following section discusses how the performance of the proposed algorithm is evaluated through a 10-fold cross-validation procedure in this paper.

C. Performance Evaluation

The classification accuracy has been one of the main pitfalls in the developed BCI systems. It directly affects the decision made in a BCI output. The classification accuracy from an experiment is calculated by dividing the number of correctly classified samples by the total number of samples [33]. In this study, a k-fold cross-validation [34], [35] is used to calculate classification accuracy for assessing the performance of the proposed method. In k-fold cross-validation procedure, a data set is partitioned into k mutually exclusive subsets of approximately equal size and the method is repeated k times (folds). Each time, one of the subsets is used as a test set and the other $k-1$ subsets are put together to form a training set. Then the average accuracy across all k trials is computed.

In this study, we select $k = 10$ as it is a common choice for the k-fold cross-validation. As mentioned in Section III-B3, we obtain a total of 235 feature vectors of six dimensions from the two-class MI EEG signals of a subject in each of the two datasets, IVa and IVb. Fig. 6 presents a design of how the extracted feature vectors of this study are partitioned into 10 mutually exclusive subsets according to the k-fold cross-validation

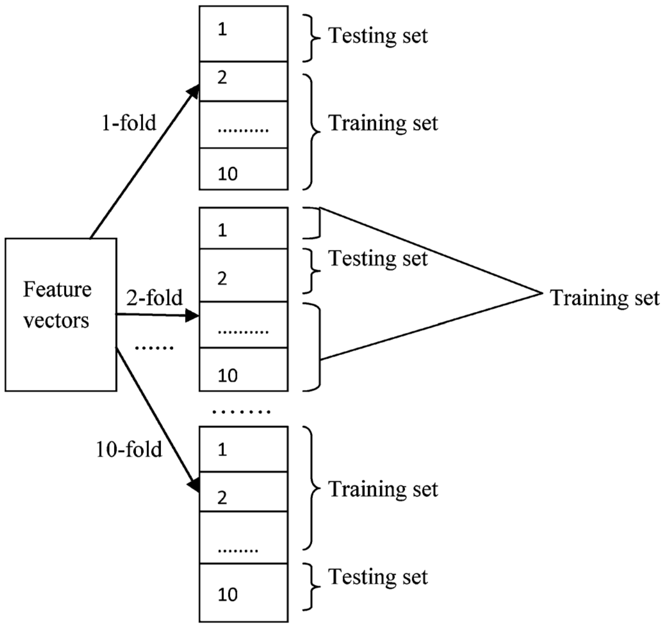

Fig. 6. Partitioning design of the obtained feature vectors for the 10-fold cross-validation method.

system. As shown in Fig. 6, the feature vector set of each subject is divided into 10 subsets and the procedure is repeated 10 times (the folds). Each time, one subset is used as a testing set and the remaining nine subsets are used as a training set, which is illustrated in Fig. 6. The obtained classification accuracy of each of 10 times on the testing set is averaged called "10-fold cross-validation accuracy" in this paper.

IV. Experiments and Results

Before classification, the hyper parameters of the LS-SVM classifier is tuned by a two-step grid search algorithm discussed in Section IV-A as the classification performance of the LS-SVM depends on the parameters and the values chosen of the parameters significantly affect the classification accuracy. Section IV-B discusses how the variables are set up in the logistic regression classifier and the kernel logistic regression classifier. The results obtained by the LS-SVM, the logistic regression and the kernel logistic regression for the cross-correlation based features are compared to each other for datasets IVa and IVb in Section IV-C. Section IV-D presents a comparative study for our proposed method with eight existing methods in the literature. In this research, the classification by the LS-SVM is carried out in MATLAB (ver. 7.7, R2008b) using the LS-SVMlab toolbox (ver. 1.5) [36] and the classification of the logistic regression and kernel logistic regression are executed through MATLABArsenal [MATLAB Classification Wrapper 1.00 (Debug version)] package [42].

In this study, the training set is applied to train the classifier and the testing vectors are used to verify the accuracy and the effectiveness of the classifiers for the classification of the two-class MI data. Our proposed algorithm is employed on each subject of both datasets separately, as the MI EEG signals are naturally highly subject-specific depending on physical and mental

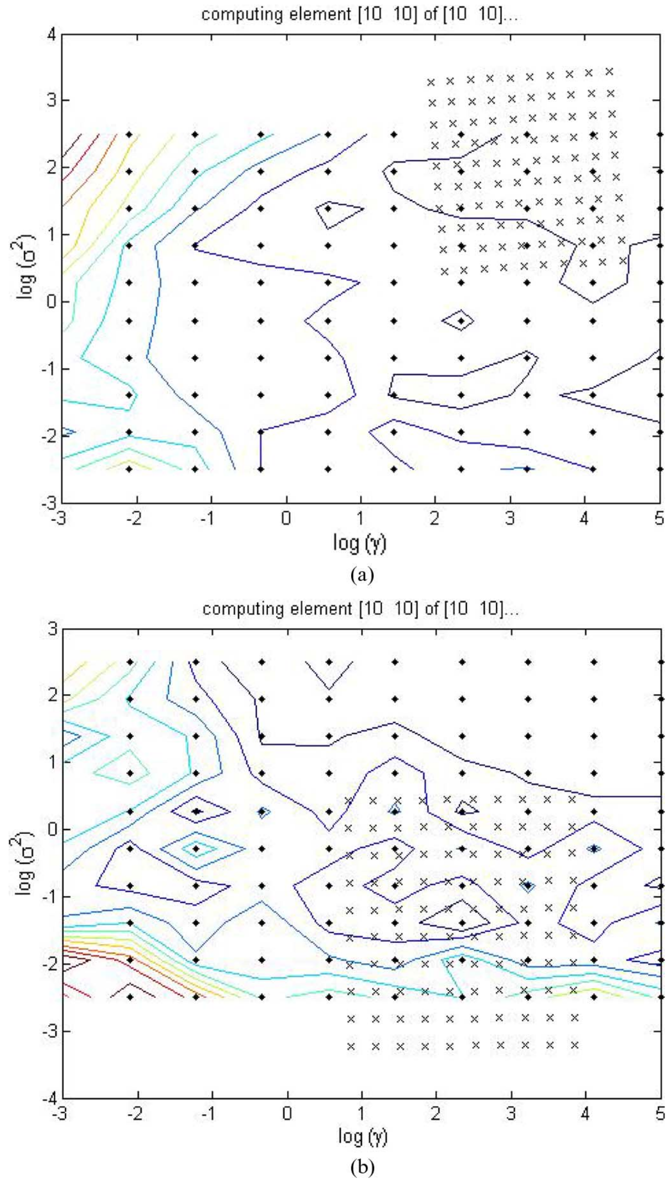

Fig. 7. (a) Process of the two-step grid search for optimizing the parameters γ (gamma) and σ^2 (sig2) of the LS-SVM classifier in the one-fold of subject aa of dataset IVa. (b) Process of the two-step grid search for optimizing the parameters γ (gamma) and σ^2 (sig2) of the LS-SVM classifier in the one-fold of dataset IVb.

tasks. All experimental results are presented based on the testing set in this paper.

A. Tunning the Hyper Parameters of the LS-SVM Classifier

To improve the generalization performance of the LS-SVM classifier, it is needed to select its two parameters (γ, σ^2) through an appropriate procedure. These parameters play an important role in the classification performance. The regularization parameter γ (gamma) determines tradeoff between minimizing the training error and minimizing the model complexity. The parameter σ^2 (sig2) is the bandwidth and implicitly defines the nonlinear mapping from the input space to a high dimensional feature space. Large values of γ and σ^2 may lead to an over-fitting problem for the training data [21], [37], so the values must be chosen carefully. This study applies a

TABLE I
OPTIMAL VALUES OF THE PARAMETERS γ AND σ^2 OF THE LS-SVM FOR DATASET IVa

	Obtained optimal parameter values of γ and σ^2 of the LS-SVM									
Subject	aa		al		av		aw		ay	
Parameters	γ	σ^2	γ	σ^2	γ	σ^2	γ	σ^2	γ	σ^2
1-fold	82.13	8.84	65.89	5.52	22.16	5.06	260.30	11.74	58.12	4.69
2-fold	31.59	7.04	60.05	8.32	220.64	7.88	181.25	11.19	72.15	1.53
3-fold	202.90	16.09	245.97	2.24	7.33	6.46	401.84	14.81	45.85	1.88
4-fold	128.22	18.62	80.27	5.66	921.84	1.78	343.21	12.92	92.86	1.05
5-fold	30.79	9.48	315.18	14.67	4.94	8.66	200.97	0.79	42.59	0.65
6-fold	58.76	20.79	632.36	13.27	10.11	3.92	141.36	12.81	78.74	1.39
7-fold	46.71	7.73	261.91	11.72	719.69	5.42	193.22	8.44	26.18	0.73
8-fold	29.75	10.85	64.29	2.21	9.23	3.99	179.95	13.72	149.09	0.84
9-fold	208.08	32.27	49.22	1.65	10.25	4.04	75.48	7.41	246.47	1.52
10-fold	29.49	10.57	79.59	1.99	16.63	5.25	605.63	13.71	49.67	1.35

Note: up to two digits decimal considered.

two-step grid search technique in order to obtain the optimum values of the hyper parameters for the LS-SVM. This section describes the process on how the parameters are selected through a two-step grid search algorithm. A grid search is a two dimensional minimization procedure based on an exhaustive search in a limited range [38]. It tries values of each parameter across a specified search range using geometric steps. In each iteration, one leaves a point, and fits a model on the other data points. The performance of the model is estimated based on the one-point-left-out. This procedure is repeated for each data point. Finally, all the different estimates of the performance are combined. The two-step grid search procedure is provided in the free LS-SVM toolbox [36] to develop the LS-SVM model [39].

In this research, the two-step grid search method is applied in each of the 10 folds of a subject of the both datasets for selecting the optimal parameter values of the LS-SVM. The obtained values of the parameters for each fold are used in the LS-SVM algorithm to obtain the reliable performance of the proposed method. Fig. 7(a) and (b) shows the process of the two-step grid search for optimizing the parameters γ (gamma) and σ^2 (sig2) of the LS-SVM classifier for dataset IVa (for the one-fold of subject $\mathbf{aa}$) and dataset IVb (for the one-fold), respectively.

Firstly, the optimal range of the parameters is determined in the first step of a grid search. The grids denoted as "♦" in the first step is 10×10, and the searching step is for a crude search with a large step size. The optimal search area is determined by the error contour line. The grids denoted as "$\times$" in the second step is also 10×10, and the searching step is the specified search with a small step size. The contour lines indicate the value levels of the cost function in the grid search. Using this method, the obtained optimal combinations of γ and σ^2 for the LS-SVM is presented in Table I for dataset IVa and in Table II for dataset IVb.

As shown in Tables I and II, the optimal values of hyper parameters for the LS-SVM are obtained in each of 10-folds for each subject of the two datasets through the two-step grid search algorithm. In this study, the classification results of each fold are achieved using the optimal parameter values in each subject of the both datasets.

TABLE II
OPTIMAL VALUES OF THE PARAMETERS γ AND σ^2 OF THE LS-SVM FOR DATASET IVb

	Obtained optimal parameter values γ and σ^2 of the LS-SVM	
Parameters	γ	σ^2
1-fold	7.0451	5.3714
2-fold	6.9404	1.5538
3-fold	47.7992	3.7401
4-fold	107.1772	1.7566
5-fold	10.0366	1.8417
6-fold	320.4905	3.3605
7-fold	57.1816	2.7994
8-fold	820.5462	1.3092
9-fold	569.3277	2.1852
10-fold	31.9349	1.8465

B. Variable Selections in the Logistic Regression and Kernel Logistic Regression Classifiers

Although the parameters of the logistic regression are obtained automatically through maximum likelihood estimation (MLE) method, the variable selections are an important task for the logistic regression model. In this research, the logistic regression presented in (3) is used to estimate the probability of the dependent variable using independent variables as the input. For each of the two datasets, we consider the MI tasks as a dependent variable, termed y, and the six statistical features (discussed in Section III-B3) are treated as six independent variables. The six independent variables used in (3) are $x_1 = mean$ values, $x_2 = maximum$ values, $x_3 = minimum$ values, $x_4 = standard\ deviation$ values, $x_5 = median$ values and $x_6 = mode$ values. It is known that the dependent variable y has two values, 0 and 1, in the logistic regression. For dataset IVa, the right-hand MI class is treated as 0 and the right-foot MI class as 1. For dataset IVb, we denote the right-foot MI class as 0 and the left-hand MI class as 1. In the kernel logistic regression, the model parameters in (4) are automatically anticipated by the iteratively re-weighted least square procedure [41]. The feature vectors and class labels of the kernel logistic regression in (4) are considered as the same as those in the logistic regression.

C. Performances on Both Datasets

Table III presents the classification results for the LS-SVM, the logistic regression and the kernel logistic regression clas-

TABLE III
CLASSIFICATION RESULTS BY THE 10-FOLD CROSS-VALIDATION METHOD ON TESTING SET OF DATASET IVa

Subject	10-fold cross-validation accuracy (%) (mean ± standard deviation)		
	LS-SVM	Logistic regression	Kernel logistic regression
aa	97.88±4.56	95.31±7.17	97.03±5.62
al	99.17±1.76	87.26±10.07	96.20±4.99
av	98.75±2.81	94.89±7.77	95.74±7.32
aw	93.43±6.87	94.93±5.14	94.51±4.88
ay	89.36±5.74	75.33±12.92	83.42±10.97
Average	95.72±4.35	89.54±8.61	93.38±6.76

sifiers for the five subjects of dataset IVa. In Table III, the results of each subject are reported in terms of mean $\pm$ standard deviation of the accuracy over a 10-fold cross-validation method on the testing set. It is observed from Table III that the proposed LS-SVM classifier for the cross-correlation features produces an accuracy of 97.88% for subject $\mathbf{aa}$, 99.17% for subject $\mathbf{al}$, 98.75% for subject $\mathbf{av}$, 93.43% for subject $\mathbf{aw}$ and 89.36% for subject $\mathbf{ay}$; while these values are 95.31%, 87.26%, 94.89%, 94.93%, 75.33%, respectively, for the logistic regression classifier; and 97.03%, 96.20%, 95.74%, 94.51%, 83.42%, respectively, for the kernel logistic regression with the same features. Based on the experimental results, the classification success rates of the proposed LS-SVM classifier are higher than those of the logistic regression and the kernel logistic regression in four out of five subjects.

Table III also reports that the standard deviations for the proposed approach are much lower compared to those of the logistic regression and the kernel logistic regression in those four subjects. The lower values of standard deviation indicate the consistency of the proposed method. As seen from Table III, the proposed LS-SVM provides the best results with an average classification accuracy of 95.72% whereas this value is 89.54% for the logistic regression and 93.38% for the kernel logistic regression classifier. The average classification accuracy for the proposed method increases by 6.18% in comparison to the logistic regression model and 2.34% to the kernel logistic regression.

In what follows, we provide the details on how the 10-fold cross-validation system produces the classification accuracy in each of the 10-folds for one subject applying the LS-SVM, the logistic regression and the kernel logistic regression classifiers. Figs. 8(a)–(e) plot the comparative results of each of the 10-folds for the five subjects for dataset IVa. The figures show the individual classification accuracies against each of the 10-folds for the logistic regression, kernel logistic regression and the proposed LS-SVM on the testing sets for subjects $\mathbf{aa}$, $\mathbf{al}$, $\mathbf{av}$, $\mathbf{aw}$, $\mathbf{ay}$, respectively. From these figures, it is observed that in most of the cases, the proposed LS-SVM classifier yields a better performance for each of the 10-folds compared to the logistic regression and the kernel logistic regression. An increasing tendency of prediction accuracy in every fold of all subjects for the LS-SVM is shown in these figures. From Figs. 8(a)–(e), the fluctuations of the performance of the proposed method are smaller among the 10-folds for each subject compared to the logistic regression model and the kernel logistic regression model, indicating that the proposed method is fairly stable.

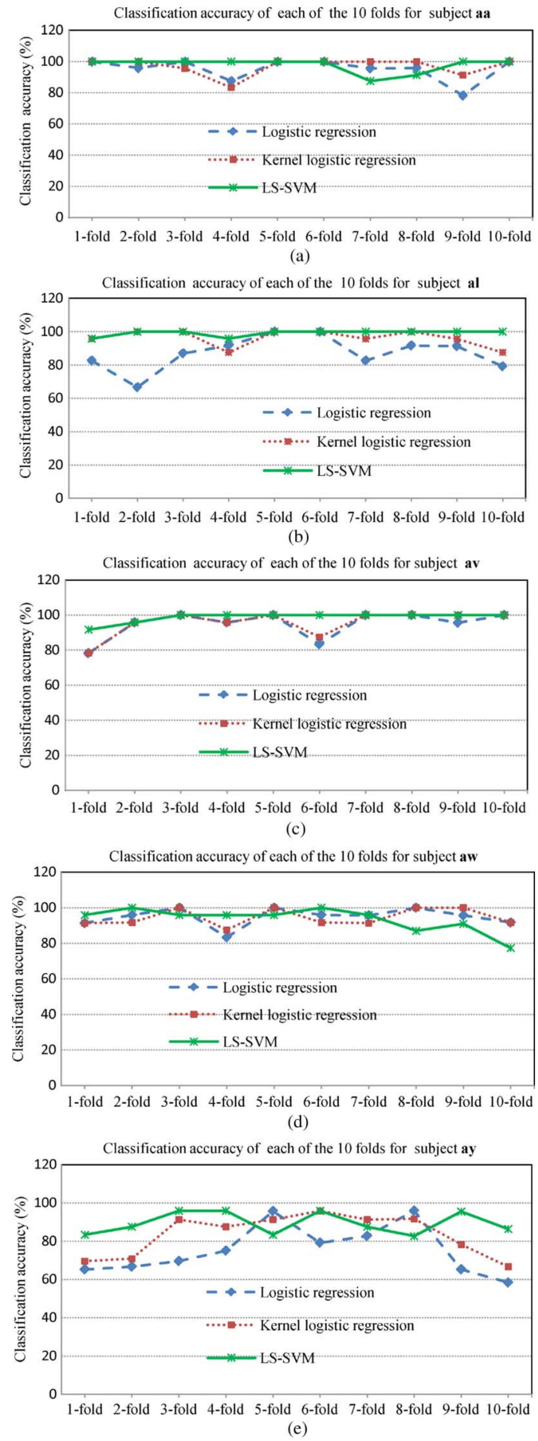

Fig. 8. Comparisons of the individual classification accuracies among the logistic regression, kernel logistic regression and the LS-SVM for each of the 10-folds: (a) subject $\mathbf{aa}$ (b) subject $\mathbf{al}$ (c) subject $\mathbf{av}$ (d) subject $\mathbf{aw}$ (e) subject $\mathbf{ay}$ in dataset IVa.

In Table IV, we provide the classification accuracy for the proposed LS-SVM, the logistic regression and the kernel logistic regression models using the 10-fold cross validation procedure for dataset IVb. As shown in Table IV, the classification accuracy is 97.89% for the LS-SVM while this value is 95.31% for the logistic regression and 94.87% for the kernel logistic regression. The results show a 2.58% improvement in the proposed LS-SVM compared to the logistic regression and 3.02%

TABLE IV
CLASSIFICATION RESULTS BY THE 10-FOLD CROSS-VALIDATION METHOD ON TESTING SET OF DATASET IVb

Methods	10-fold cross-validation accuracy (%) (mean ± standard deviation)
LS-SVM	97.89±2.96
Logistic regression	95.31±5.88
Kernel logistic regression	94.87±6.98

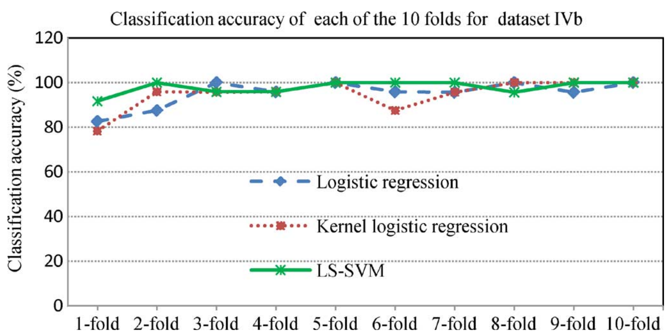

Fig. 9. Comparisons of the individual classification accuracies among the logistic regression, kernel logistic regression and the LS-SVM for each of the 10-folds in dataset IVb.

over the kernel logistic regression for the same inputs. The standard deviation value is also smaller in the LS-SVM compared to the logistic regression and the kernel logistic regression, which reflects the consistency of the LS-SVM. The results in terms of the 10-fold cross-validation accuracy on both datasets, displayed in Table III and Table IV, demonstrate that the proposed LS-SVM classifier is superior compared to the logistic regression and the kernel logistic regression methods for the same features.

From Fig. 9, it is observed that in most of the 10-folds, the proposed LS-SVM generates higher accuracies and also the variations of the performance among the 10-folds are smaller compared to those of the logistic regression and the kernel logistic regression. These indicate that the proposed method is more reliable for the MI signal classification. From Fig. 8(a)–(e) and Fig. 9, it is clear that the proposed algorithm achieves a better classification performance both individually and overall, compared to the logistic regression and the kernel logistic regression.

In order to report the performance with a different channel data as a reference signal, we use the electrode position C3 (according to the 10/20 system) as a reference signal instead of Fp1 in the present algorithm for each subject of the both datasets. Like Fp1, C3 is also selected from the right-hand class from dataset IVa, while it is from the right-foot class for dataset IVb.

Using the 10-fold cross-validation procedure, the proposed LS-SVM classifier yields the classification accuracy of 99.58%, 94.94%, 98.64%, 93.26%, and 91.06% for subjects **aa**, **al**, **av**, **aw**, **ay**, respectively, for the reference signal of channel C3; whereas these values are 97.88%, 99.17%, 98.75%, 93.43%, and 89.36% for the reference signal of Fp1 in dataset IVa as shown in Table V. The overall accuracy for the LS-SVM is 95.49% for channel C3 and 95.72% for channel Fp1. For the same dataset, the logistic regression generates the 10-fold cross-validation accuracy of 97.88% for subject **aa**, 78.32% for subject **al**, 97.83% for subject **av**, 96.18% for subject **aw** and 68.91% for subject **ay** using channel C3 as the reference signal; whereas those values are 95.31%, 87.26%, 94.89%, 94.93%, 75.33% for the reference signal of channel Fp1. The average classification accuracy of the logistic regression reaches at 87.82% for the reference signal C3 and 89.54% for the reference signal Fp1. For the same dataset, the kernel logistic regression produces classification accuracy of 96.16%, 89.78%, 96.58%, 93.15%, 85.09% for subjects **aa**, **al**, **av**, **aw**, **ay**, respectively, for the reference channel C3; whereas these values are 97.03%, 96.20%, 95.74%, 94.51%, 83.42% for the reference channel Fp1as provided in Table V. Thus, the kernel logistic regression achieves the overall accuracy for five subjects as 92.15% for the reference signal C3 and 93.38% for the Fp1. For dataset IVb, we obtain an accuracy of 97.88% for the proposed LS-SVM algorithm with the reference signal C3; while this value is 97.89% for the reference signal Fp1as shown in Table VI. On the other hand, the logistic regression classifier produces a classification accuracy of 86.41% for the channel C3 and 95.31% for the channel Fp1 for the same dataset. For reference channel C3, the kernel logistic regression is able to generate the classification performance of 95.36% where this value is 94.87% for the reference channel Fp1. From the results of both the reference signals C3 and Fp1, it is observed that the performance of the proposed algorithm does not differ significantly from changing the reference signal. It proves the robustness of the method.

In addition, to investigate the performance of the proposed six features, we add other three features, *inter quartile range* (IQR), *1/4 percentile* (P_{25}) (*first quartile*, $Q_1 = P_{25}$) and *3/4 percentile* (P_{75}) (*third quartile*, $Q_3 = P_{75}$), into our existing feature set. As mentioned before, our existing feature set consists of six features which are {*mean*, *median*, *mode*, *standard deviation*, *maximum* and *minimum*}. Adding the three features $\{IQR$, P_{25} and $P_{75}\}$ into the existing feature set, we get a feature set of nine features, which is {*mean*, *median*, *mode*, *standard deviation*, *maximum* and *minimum*, IQR, P_{25} and P_{75}}. Table V and Table VI present the classification results of three classifiers, the LS-SVM, the logistic regression and the kernel logistic regression for the nine features set in comparing with the results of the existing six features set for two reference signals, Fp1 and C3, for datasets IVa and IVb, respectively. As shown in Table V, the proposed LS-SVM based algorithm with the nine features achieves the classification accuracy of 96.51%, 97.05%, 97.48%, 95.21%, and 90.63% for subjects **aa**, **al**, **av**, **aw**, **ay**, respectively, in dataset IVa, for the reference signal Fp1. These values are 97.88%, 99.17%, 98.75%, 93.43%, and 89.36% for the existing feature set with the same reference signal. For the reference signal C3, the proposed method with nine features is able to provide accuracy of 98.29% for subject **aa**, 95.78% for subject **al**, 98.22% for subject **av**, 94.91% for subject **aw**, and 91.02% for subject **ay**; while these values are 99.58%, 94.94%, 98.64%, 93.26%, and 91.06%, respectively, for the six features set. From Table V, it is also seen that the average classification rates of the proposed method are 95.38% for the nine features and 95.72% for the six features with the reference signal Fp1; while these values are 95.65% and 95.49%, respectively, with the reference signal C3.

TABLE V
CLASSIFICATION RESULTS (%) OF THE THREE CLASSIFIERS FOR THE NINE FEATURES AND THE SIX FEATURES FOR THE REFERENCE SIGNALS, FP1 AND C3, IN DATASET IVa

Sub	Fp1 ref signal		C3 ref signal		Fp1 ref signal		C3 ref signal		Fp1 ref signal		C3 ref signal	
	SVM_9	SVM_6	SVM_9	SVM_6	LR_9	LR_6	LR_9	LR_6	KLR_9	KLR_6	KLR_9	KLR_6
aa	96.51	97.88	98.29	99.58	97.90	95.31	97.05	97.88	98.30	97.03	94.87	96.16
al	97.05	99.17	95.78	94.94	85.60	87.26	88.93	78.32	96.20	96.20	90.18	89.78
av	97.48	98.75	98.22	98.64	95.31	94.89	97.83	97.83	95.74	95.74	96.16	96.58
aw	95.21	93.43	94.91	93.26	94.89	94.93	92.00	96.18	95.74	94.51	94.86	93.15
ay	90.63	89.36	91.02	91.06	82.14	75.33	66.36	68.91	89.35	83.42	86.36	85.09
avg	95.38	95.72	95.65	95.49	91.17	89.54	88.43	87.82	95.07	93.38	92.49	92.15

Note: Sub = subject; avg = average; ref = reference; SVM_9 = LS − SVM with the nine features; SVM_6 = LS − SVM with the six features; LR_9 = logistic regression with the nine features; LR_6 = logistic regression with the six features; KLR_9 = kernel logistic regression with the nine features; KLR_6 = kernel logistic regression with the six features.

TABLE VI
CLASSIFICATION RESULTS (%) OF THE THREE CLASSIFIERS FOR THE NINE FEATURES AND THE SIX FEATURES FOR THE REFERENCE SIGNALS, FP1 AND C3, IN DATASET IVb

Methods	Nine features		Six features	
	Fp1 ref signal	C3 ref signal	Fp1 ref signal	C3 ref signal
LS-SVM	97.48	97.88	97.89	97.88
Logistic regression	94.47	87.70	95.31	86.41
Kernel logistic regression	95.31	96.65	94.87	95.36

For the same dataset, Table V reports that the logistic regression with the reference signal Fp1 obtains 97.90%, 85.60%, 95.31%, 94.89%, and 82.14% classification accuracy for subjects $\mathbf{aa}$, $\mathbf{al}$, $\mathbf{av}$, $\mathbf{aw}$, $\mathbf{ay}$, respectively, for the nine features; whilst those values are 95.31%, 87.26%, 94.89%, 94.93%, and 75.33%, respectively, for the six features. With the reference signal C3, the logistic regression produces 97.05% for subject $\mathbf{aa}$, 88.93% for subject $\mathbf{al}$, 97.83% for subject $\mathbf{av}$, 92.00% for subject $\mathbf{aw}$ and 66.36% for subject $\mathbf{ay}$ for the nine features; while these are 97.88%, 78.32%, 97.83%, 96.18%, and 68.91% for the six features. As shown in Table V, the overall performance of the logistic regression model is 91.17% for the nine features and 89.54% for the six features with the reference signal Fp1, and 88.43%, and 87.82%, respectively, for the reference signal C3.

On the other hand, it can be seen from Table V that the kernel logistic regression with the reference signal Fp1 yields the classification accuracy of 98.30%, 96.2%, 95.74%, 95.74%, and 89.35% for subjects $\mathbf{aa}$, $\mathbf{al}$, $\mathbf{av}$, $\mathbf{aw}$, $\mathbf{ay}$, respectively, for the nine features; whereas these values are 97.03%, 96.20%, 95.74%, 94.51%, and 83.42%, respectively, for the six features. With the reference signal C3, this algorithm achieves 94.87% for subject $\mathbf{aa}$, 90.18% for subject $\mathbf{al}$, 96.16% for subject $\mathbf{av}$, 94.86% for subject $\mathbf{aw}$and 86.36% for subject $\mathbf{ay}$, respectively, for the nine features; while those values are 96.16%, 89.78%, 96.58%, 93.15%, and 85.09%, respectively, for the six features. The average accuracies of this algorithm are 95.07% for the nine features and 93.38% for the six features with the reference signal Fp1; where the values are 92.49% and 92.15%, respectively, for the reference signal C3.

In dataset IVb, the LS-SVM classifier with the three added features $\{IQR, P_{25}$ and $P_{75}\}$ generates 97.48% accuracy for the reference signal Fp1 and 97.88% for the reference signal C3; where these values are 97.89% and 97.88%, respectively, for the six features as shown in Table VI. For the reference signal Fp1, the classification accuracies of the logistic regression are obtained as 94.47% for the nine features and 95.31% for the six features, while these values are 87.70% and 86.41% for the reference signal C3. On the other hand, the kernel logistic regression with the nine features provides the classification performance of 95.31% for the reference signal Fp1 and 96.65% for the reference signal C3; while these values are 94.87% and 95.36% for the six features.

From the discussions, we can see that there is no significant difference of performance between the nine features and the six features. If there are outliers (an *outlier* is an observation that lies an abnormal distance from other values in a set of data) in the data, the IQR *i*s more representative than the standard deviation as an estimate of the spread of the body of the data. The IQR is less efficient than the *standard deviation* as an estimate of the spread when the data is approximately normally distributed. For the same type of distribution (normal distribution), P_{25} and P_{75} are not good measures to represent a distribution. As the datasets used in this study are almost symmetric and there are no obvious outliers, that is why, we do not get significantly better performance when the three features $\{IQR, P_{25}$ *and* $P_{75}\}$ are added into the six features.

D. Performance Comparisons With the Existing Techniques

In order to further examine the efficiency of the proposed algorithm, this section provides the comparisons of our approach with other eight recently reported techniques. Those eight existing algorithms for dataset IVa are discussed in Section II. Table VII reports the comparison results of the classification accuracy rates for the proposed method and the eight algorithms for dataset IVa. This table shows the classification performance for the five subjects as well as the overall mean accuracy values. The highest classification accuracy rate among the nine algorithms is highlighted in bold font for each subject and their averages.

From Table VII, it is noted that the proposed cross-correlation based LS-SVM algorithm provides better classification accuracies than the other eight algorithms in all of the five subjects. The highest classification rates are 97.88% for subject $\mathbf{aa}$, 99.17% for subject $\mathbf{al}$, 98.75% for subject $\mathbf{av}$, 93.43% for subject $\mathbf{aw}$ and 89.36% for subject $\mathbf{ay}$ produced by our proposed approach.

TABLE VII
PERFORMANCE COMPARISONS FOR DATASET IVa

Authors	Method	Classification accuracy rate (%)					
		aa	**al**	**av**	**aw**	**ay**	Average
Proposed method	CC-based LS-SVM	**97.88**	**99.17**	**98.75**	**93.43**	**89.36**	**95.72**
Siuly et al. [13]	CT-based LS-SVM	92.63	84.99	90.77	86.50	86.73	88.32
Lu et al. [17]	R-CSP with aggregation	76.8	98.2	74.5	92.9	77.0	83.9
Lotte et al. [16]	SSRCSP	70.54	96.43	53.57	71.88	75.39	73.56
Lotte et al. [16]	TRCSP	71.43	96.43	63.27	71.88	86.9	77.98
Lotte et al. [16]	WTRCSP	69.64	98.21	54.59	71.88	85.32	75.93
Lotte et al. [16]	SRCSP	72.32	96.43	60.2	77.68	86.51	78.63
Lu et al. [15]	R-CSP with generic learning	69.6	83.9	64.3	70.5	82.5	74.20
Yong et al. [14]	Sparse spatial filter optimization	57.5	86.9	54.4	84.4	84.3	73.50

Note: $\mathrm{CC} = \mathrm{cross-correlation\ technique}$; $\mathrm{CT} = \mathrm{Clustering\ technique}$; $\mathrm{R-CSP} = \mathrm{Regulized\ common\ spatial\ pattern}$; $\mathrm{CSP} = \mathrm{Common\ spatial\ pattern}$

Further looking at the performance comparisons in Table VII, it is noted that the proposed algorithm is ranked first in terms of the average accuracy (95.72%), while the CT based LS-SVM algorithm [13] comes second (88.32%), R-CSP with aggregation [17] is third (83.9%), and so on. The sparse spatial filter optimization [14] is the last (73.50%). The results indicate that the proposed method achieves by 7.40%–22.22% improvements over all the eight existing algorithms for BCI competition III, dataset IVa.

V. CONCLUSION

The translation of brain activities into control signals in BCI systems requires a robust and accurate classification of the various types of information. In this paper, we present a cross-correlation based LS-SVM algorithm for improving the classification accuracy of the MI-based EEG signals in BCI systems. The proposed scheme utilizes a cross-correlogram based feature extraction procedure for the MI signals, and develops a LS-SVM classifier for the classification of the extracted MI features. We apply the same features as the inputs to a logistic regression and the kernel logistic regression models for comparing the performance of the proposed LS-SVM classifier. In addition, we compare our proposed approach with eight other recently reported methods. As the parameters of the LS-SVM can significantly affect the classification performance, we use a two-step grid search algorithm for selecting optimal combinations of parameters of the LS-SVM classifier. The methods are tested on datasets IVa and IVb of BCI Competition III. All experiments on both datasets are evaluated through a 10-fold cross-validation process, which indicates the reliability of the obtained results. The main conclusions of this study are summarized as follows.

1) The proposed method is promising for a two-class MI EEG signal classification. The feasibility of the approach has been verified with BCI competition III, datasets IVa and IVb.
2) The cross-correlation feature extraction procedure is effective for the classification performance even when the data size is very large. The experimental results from the three classifiers, the LS-SVM, the logistic regression and the kernel logistic regression, confirm that the extracted features are reliable for capturing the valuable information from the original MI signal patterns.
3) To further investigate the reliability of the obtained features, we add other three features $\{IQR,\ P_{25}\ and\ P_{75}\}$, into the current six features and then employ the LS-SVM, the logistic regression and the kernel logistic regression algorithms as the inputs, separately. The results show that the performance of the nine features is not much improved in comparison with those of the six features for each of the three algorithms.
4) The experimental results using the proposed algorithm are consistent because the parameter values of the LS-SVM classifier are optimally selected through the two-step grid search algorithm rather than by the manual selection.
5) The results show that the proposed LS-SVM classifier achieves a better performance compared to the logistic regression and the kernel logistic regression classifiers for the same feature vectors in both datasets.
6) The experimental results also indicate that the proposed approach outperforms the other eight recently reported methods in BCI Competition III, dataset IVa, by at least 7.40%. It demonstrates that our method performs the best for the MI signal classification in BCI applications.

This study concludes that the cross-correlation based LS-SVM algorithm is a promising technique for MI signal recognition and it offers great potentials for the development of MI-based BCI analyses which assist clinical diagnoses and rehabilitation tasks. In the future, we will extend the proposed cross-correlation based LS-SVM algorithm to multiclass classification problems.

REFERENCES

[1] B. Y. Wang, X. Gao, B. Hong, C. Jia, and S. Gao, "Brain-computer interfaces based on visual evoked potentials," *IEEE Eng. Med. Biol. Mag.*, vol. 27, no. 5, pp. 64–71, Sep.–Oct. 2008.

[2] D. J. McFarland and J. R. Wolpaw, "Brain-computer interfaces for communication and control," *Commun. ACM*, vol. 54, no. 5, pp. 60–66, 2011.

[3] G. Pfurtscheller, C. Brunner, A. Schlogl, and F. Lopes da Silva, "Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks," *Neuroimage*, vol. 31, no. 1, pp. 153–159, 2006.

[4] T. Kayikcioglu and O. Aydemir, "A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data," *Pattern Recognit. Lett.*, vol. 31, pp. 1207–1215, 2010.

[5] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanable, and K. R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[6] M. Grosse-Wentrup, C. Liefhold, K. Gramann, and M. Buss, "Beamforming in noninvasive brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1209–1219, Apr. 2009.

[7] 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. Biomed. Eng.*, vol. 55, no. 6, pp. 1733–1743, Jun. 2008.

[8] J. Long, Y. Li, and Z. Yu, "A semi-supervised support vector machine approach for parameter setting in motor imagery-based brain computer interfaces," *Cognitive Neurodynamics*, vol. 4, pp. 207–216, 2010.

[9] J. Meng, G. Liu, G. Huang, and X. Zhu, "Automated selecting subset of channels based on CSP in motor imagery brain-computer system," in *Proc. 2009 IEEE Int. Conf. Robot. Bioinformat.*, Guilin, China, Dec. 19–23, 2009, pp. 2290–2294.

[10] G. M. Hieftje, R. I. Bystroff, and R. Lim, "Application of correlation analysis for signal-to-noise enhancement in flame spectrometry: Use of correlation in determination of rhodium by atomic fluorescence," *Analytical Chem.*, vol. 45, no. 2, pp. 253–258, 1973.

[11] S. Dutta, A. Chatterjee, and S. Munshi, "An automated hierarchical gait pattern identification tool employing cross-correlation-based feature extraction and recurrent neural network based classification," *Expert Syst.*, vol. 26, no. 2, pp. 202–217, 2009.

[12] H. Esen, F. Ozgen, M. Esen, and A. Sengur, "Modelling of a new solar air heater through least-squares support vector machines," *Expert Syst. Appl.*, vol. 36, pp. 10673–10682, 2009.

[13] Siuly, Y. Li, and P. Wen, "Clustering technique-based least square support vector machine for EEG signal classification," *Comput. Methods Programs Biomed.*, vol. 104, pp. 358–372, 2011.

[14] X. Yong, R. K. Ward, and G. E. Birch, "Sparse spatial filter optimization for EEG channel reduction in brain-computer interface," in *Proc. ICASSP 2008*, pp. 417–420.

[15] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial patterns with generic learning for EEG signal classification," in *Proc. IEEE 31st Annu. Int. Conf. EMBS*, Sep. 2–6, 2009, pp. 6599–6602.

[16] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.

[17] H. Lu, H. L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial patterns with aggregation for EEG classification in small-sample setting," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2936–2945, Dec. 2010.

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

[19] B. Blankertz, K. R. Muller, D. J. Krusierski, G. Schalk, J. R. Wolpaw, A. Schlgl, G. Pfurtscheller, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.

[20] S. Dutta, A. Chatterjee, and S. Munshi, "Correlation techniques and least square support vector machine combine for frequency domain based ECG beat classification," *Med. Eng. Phys.*, vol. 32, no. 10, pp. 1161–1169, Dec. 2010.

[21] S. Chandaka, A. Chatterjee, and S. Munshi, "Cross-correlation aided support vector machine classifier for classification of EEG signals," *Expert Syst. Appl.*, vol. 36, pp. 1329–1336, 2009.

[22] T. A. L. Wren, K. P. Do, S. A. Rethlefsen, and B. Healy, "Cross-correlation as a method for comparing dynamic electromyography signals during gait," *J. Biomechan.*, vol. 39, pp. 2714–2718, 2006.

[23] M. N. Islam, *An Introduction to Statistics and Probability*, 3rd ed. Dhaka, Bangladesh: Mullick, pp. 160–161.

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

[25] U. Thissen, B. Ustun, W. J. Melssen, and L. M. C. Buydens, "Multivariate calibration with least-square support vector machines," *Analytical Chem.*, vol. 76, pp. 3099–3105, 2004.

[26] S. Siuly, Y. Li, and P. Wen, "Classification of EEG signals using sampling techniques and least square support vector machines," in *RSKT 2009*. Berlin, Germany: Springer, vol. 5589, Lecture Notes Computer Science, pp. 375–382.

[27] J. A. K. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle, *Least Square Support Vector Machine*. Singapore: World Scientific, 2002.

[28] Siuly, Y. Li, and P. Wen, "Analysis and classification of EEG signals using a hybrid clustering technique," in *Proc. 2010 IEEE/ICME Int. Conf. Complex Med. Eng. (CME2010)*, pp. 34–39.

[29] W. Caesarendra, A. Widodo, and B. S. Yang, "Application of relevance vector machine and logistic regression for machine degradation assessment," *Mechan. Syst. Signal Process.*, vol. 24, pp. 1161–1171, 2010.

[30] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*. New York: Wiley, 1989.

[31] A. Subasi and E. Ercelebi, "Classification of EEG signals using neural network and logistic regression," *Comput. Methods Programs Biomed.*, vol. 78, pp. 87–99, 2005.

[32] A. Subasi, A. Alkan, E. Koklukaya, and M. K. Kiymik, "Wavelet neural network classification of EEG signals by using AR model with MLE preprocessing," *Neural Netw.*, vol. 18, pp. 985–997, 2005.

[33] Siuly, Y. Li, and P. Wen, "EEG signal classification based on simple random sampling technique with least square support vector machines," *Int. J. Biomed. Eng. Technol.*, 2011, to be published.

[34] S. Abdulkadir, "Multiclass least-square support vector machines for analog modulation classification," *Expert Syst. Appl.*, vol. 36, pp. 6681–6685, 2009.

[35] S. Ryali, K. Supekar, D. A. Abrams, and V. Menon, "Sparse logistic regression for whole-brain classification of fMRI data," *NeuroImage*, vol. 51, pp. 752–764, 2010.

[36] LS-SVMlab Toolbox (Version 1.5) [Online]. Available: http://www.esat.kuleuven.ac.be/sista/lssvmlab/

[37] J. A. K. Suykens and J. Vandewalle, "Least square support vector machine classifiers," *Neural Process. Lett.*, vol. 9, pp. 293–300, 1999.

[38] L. Xie, Y. Ying, and T. Ying, "Classification of tomatoes with different genotypes by visible and short-wave near-infrared spectroscopy with least-square support vector machines and other chemometrics," *J. Food Eng.*, vol. 94, pp. 34–39, 2009.

[39] X. L. Li, Y. He, and C. Q. Wu, "Least square support vector machine analysis for the classification of paddy seeds by harvest year," *Proc. ASABE*, vol. 51, no. 5, pp. 1793–1799, 2008.

[40] G. C. Cawley and N. L. C. Talbot, "Efficient approximate leave-one-out cross-validation for kernel logistic regression," *Mach. Learn.*, vol. 71, pp. 243–264, 2008.

[41] S. P. Rahayu, S. W. Purnami, A. Embong, and J. M. Zain, "Kernel logistic regression-linear for leukemia classification using high dimensional data," *JUTI*, vol. 7, no. 3, pp. 145–150, 2009.

[42] MATLABArsenal [Online]. Available: http://www.informedia.cs.cmu.edu/yanrong/MATLABArsenal/MATLABArsenal.zip

Siuly Siuly received the B.Sc. (Hons) and M.Sc. degrees in statistics from the University of Dhaka, Bangladesh. She is currently working toward the Ph.D. degree in biomedical engineering in the Department of Mathematics and Computing, University of Southern Queensland (USQ), Toowoomba, Australia.

Her research interests are EEG signal classification, signal processing, machine learning, biomedical informatics, and pattern recognition.

Yan Li received the B.Sc. and M.Sc. degrees from Huazhong University of Science and Technology, Wuhan, China, and the Ph.D. degree from the Flinders University of South Australia, Adelaide, Australia.

She is an Associate Professor of Department of Mathematics and Computing at University of Southern Queensland (USQ), Toowoomba, Australia. Her research interests include a wide range area of bio-medical engineering, artificial intelligence, blind signal processing, and independent component analysis.