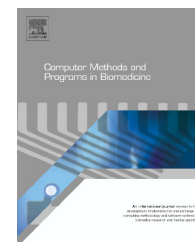




ELSEVIER

journal homepage: [www.intl.elsevierhealth.com/journals/cmpb](http://www.intl.elsevierhealth.com/journals/cmpb)

# Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification

Siuly Siuly<sup>a,\*</sup>, Yan Li<sup>b</sup>

<sup>a</sup> Centre for Applied Informatics, College of Engineering and Science, Victoria University, Melbourne, VIC, Australia

<sup>b</sup> Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, QLD, Australia

## ARTICLE INFO

### Article history:

Received 22 July 2014

Received in revised form

11 January 2015

Accepted 21 January 2015

### Keywords:

Electroencephalogram (EEG)

Optimum allocation

Principal component analysis

Least square support vector machine (LS-SVM)

Naive Bayes classifier (NB)

k-Nearest neighbor algorithm (KNN)

## ABSTRACT

The aim of this study is to design a robust feature extraction method for the classification of multiclass EEG signals to determine valuable features from original epileptic EEG data and to discover an efficient classifier for the features. An optimum allocation based principal component analysis method named as OA.PCA is developed for the feature extraction from epileptic EEG data. As EEG data from different channels are correlated and huge in number, the optimum allocation (OA) scheme is used to discover the most favorable representatives with minimal variability from a large number of EEG data. The principal component analysis (PCA) is applied to construct uncorrelated components and also to reduce the dimensionality of the OA samples for an enhanced recognition. In order to choose a suitable classifier for the OA.PCA feature set, four popular classifiers: least square support vector machine (LS-SVM), naive bayes classifier (NB), k-nearest neighbor algorithm (KNN), and linear discriminant analysis (LDA) are applied and tested. Furthermore, our approaches are also compared with some recent research work. The experimental results show that the LS-SVM.1v1 approach yields 100% of the overall classification accuracy (OCA), improving up to 7.10% over the existing algorithms for the epileptic EEG data. The major finding of this research is that the LS-SVM with the 1v1 system is the best technique for the OA.PCA features in the epileptic EEG signal classification that outperforms all the recent reported existing methods in the literature.

© 2015 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Epilepsy is one of the most common neurological disorders that hugely impair patient's daily lives. The disorder is associated with abnormal neural activities which may lead to seizures. According to a 2009 report by the World Health

Organization (WHO), worldwide over 50 million people suffer from epilepsy [1–4], while over the course of a life time, 100 million people will experience an epileptic episode [4]. This makes epilepsy the single most widespread serious neurological condition in the world today. In Australia, it is estimated that 2% of the population is affected by epilepsy [5]. Electroencephalography (EEG) is an important clinical tool that is used for the

\* Corresponding author. +61 3 991999511 (office).

E-mail addresses: [siuly.siuly@vu.edu.au](mailto:siuly.siuly@vu.edu.au) (S. Siuly), [liyan@usq.edu.au](mailto:liyan@usq.edu.au) (Y. Li).

<http://dx.doi.org/10.1016/j.cmpb.2015.01.002>

0169-2607/© 2015 Elsevier Ireland Ltd. All rights reserved.

purpose of epileptic detection as it is a condition related to the brain's electrical activity [6,7]. Epileptic activity can create clear abnormalities on a standard EEG and leaves its signature on it. Abnormal epileptic waveforms can be identified in EEG signal if there is abnormal neural activity during ictal state [8].

In order to find traces of epilepsy, visual marking of long EEG recordings by human experts was a very tedious, time-consuming and high-cost task [9]. On the other hand, due to visual analog analysis, different experts could give contradictory diagnosis for a same EEG segment. Therefore, it is a big challenge in the current biomedical research how to classify time-varying epileptic EEG signals as accurately as possible [10,11]. In order to perform classification upon signals, first the most important task is to extract distinguishing features or characteristics from the epileptic EEG data, which directly dictate the detection and classification accuracy. The features characterizing the original EEG are used as the input of a classifier to differentiate normal and epileptic EEGs. As optimal features play a very important role in the performance of a classifier, this study intends to find out a robust feature extraction method for the recognition of epileptic EEG signals.

Most recently, Shen et al. [12] introduced a method based on a cascade of wavelet-approximate entropy for feature extraction in the epileptic EEG signal classification. They tested three existing methods for classification: support vector machine (SVM), k-nearest neighbor (KNN), and radial basis function neural network (RBFNN), to determine which has the best performance in such a cascaded EEG analysis system. In 2012, Acharjee and Shahnaz [13] used twelve Cohen class kernel functions to transform EEG data in order to facilitate the time frequency analysis. The transformed data formulated a feature vector consisting of modular energy and modular entropy, and the feature vector was fed to an artificial neural network (ANN) classifier. Murugavel et al. [14] reported a method based on Lyapunov feature and a multi-class SVM for the classification of epileptic EEG signals. The wavelet transform (WT) and Lyapunov exponent were used to extract features from the EEG data and probabilistic neural network (PNN), radial basis function neural network (RBFNN) and multiclass SVM with 1vsA were applied to classify the five-category EEG signals. Ubeyli [15] developed an approach that integrated automatic diagnostic systems with spectral analysis techniques where the wavelet coefficients and power spectral density (PSD) values obtained by eigenvector methods were used as features. These features were fed to each of the seven classifiers: SVM, recurrent neural networks (RNN), PNN, mixture of experts (ME), modified mixture of experts (MME), combined neural networks (CNN), and multilayer perceptron neural network (MLPNN). By Ubeyli [16], a method based on eigenvector methods and multiclass SVMs with the ECOC was introduced for the classification of EEG signals. In the feature extraction stage, three eigenvector methods such as, the Pisarenko, multiple signal classification (MUSIC), and minimum-norm, were used to obtain the PSD values of the EEG signals and then these features were employed as the input of the multi-class SVMs. Guler and Ubeyli [17] developed another method based on the multiclass SVM with the ECOC where features were extracted from the epileptic EEG data by computing the wavelet coefficients and Lyapunov exponents. The PNN and MLPNN were also tested for their performances on the same

epileptic EEG data. In [18], Acharya et al. discussed various feature extraction methods and the results of the different automated epilepsy stage detection techniques in detail. In that paper, the authors briefly discussed some challenges to develop a computer aided diagnostic system to automatically identify the normal and abnormal activities using a minimum number of highly discriminating features for classification. Lee et al. [19] proposed several hybrid methods to classify normal and epileptic EEG signals using wavelet transform, phase-space reconstruction, and Euclidean distance based on a neural network with weighted fuzzy membership functions. The classification of abnormal activities of the brain functionality was achieved by understanding abnormal activities caused by changes in neuronal electrochemical activities through identifying the EEG signal features reported in [20] by Oweis and Abdulhay. Hilbert weighted frequency was used to help discriminate between healthy and seizure EEG patterns. In [21], Li et al. designed a method for feature extraction and pattern recognition of ictal EEGs, based upon empirical model decomposition (EMD) and SVM. First an EEG signal was decomposed into intrinsic mode functions (IMFs) using the EMD, and then the coefficients of the variation and the fluctuation index of the IMFs were extracted as features. Pachori et al. [22] presented a method for the classification of ictal and seizure-free EEG signals based on the EMD and the second-order difference plot (SODP). The EMD method decomposed an EEG signal into a set of symmetric and band-limited signals (the IMFs). The SODP of the IMFs provided an elliptical structure. The importance of the entropy based features was presented in [23] for recognizing the normal EEGs, and ictal as well as interictal epileptic seizures. Three non-linear features, such as wavelet entropy, sample entropy, and spectral entropy, were used to extract quantitative entropy features from the given EEGs, and these features were used as the input into two neural network models: recurrent Elman network and radial basis network for the classification. Siuly et al. [24] introduced a new approach based on simple random sampling (SRS) technique with least square support vector machine (LS-SVM) to classify EEG signals. In that paper, the SRS technique was applied in two stages to select random samples and sub-samples from each EEG channel data file and finally different features were evaluated from each sub-sample set to represent the original EEG signals and reduced the dimensionality of the data that were used as the inputs to the LS-SVM classifier.

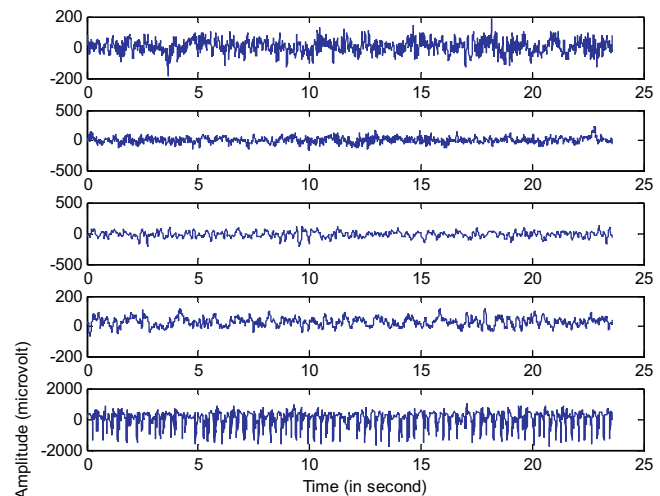
From the literature, it is observed that many of the feature extraction methods were not a perfect choice for non-stationary epileptic EEG data [25,26] and the reported methods were limited in their success rate and effectiveness. Some existing methods cannot work properly for large EEG data. None of the prior methods considered the variability of the observations within a window where the time variability is an important consideration for describing the characteristics of the original data. In most of the cases, the methods did not select their parameters through experimental evaluation, although the parameters significantly affect the classification performance.

This research intends to design a robust method for feature extraction based on the optimum allocation (OA) and principal component analysis (PCA), denoted as OA\_PCA. The reason

of using the OA method is to acquire representative sampling points from the huge amount of EEG data when the data is heterogeneous. It is worthy to mention that epileptic EEG data is heterogeneous in term of the time period. The idea to use the PCA is due to the fact that, in most of the EEG data, there is a large amount of redundant information unnecessary for diagnostic applications. The PCA method is practical when there are a large number of correlated variables in a dataset and it is believed that there are some redundancies in those variables. The PCA is an excellent method to make uncorrelated variables that can contribute to get an enhanced classification accuracy and it also can reduce the dimension of the data for easy handling. The core objective of this research is to develop a new method for feature extraction and also to discover a suitable classifier for the features that can improve overall classification accuracy (OCA) with low false alarm rate (FAR).

In order to acquire better representative features from the non-stationary EEG signals, it is required to make them stationary. That's why, this study partitions epileptic EEG data into a number of windows where the number of windows depends on the size of data and data structure (discussed in detail in Section 2.2.2). Then the OA scheme is effectively used to search representative sampling points from each window, which can efficiently describe the signals based on the variability of their observations within the window. After that, the PCA method is employed on the OA data to produce uncorrelated variables and also to reduce dimensionality. Because there is a possibility of much correlation among the brain electrical activities through the brain volume, and correlation may exist among different channels. The obtained principal components are treated as features in this study, called OA.PCA feature set. To find out an efficient classifier for the OA.PCA feature set, this study employs four prominent classifiers namely, least square support vector machine (LS-SVM), naive bayes classifier (NB),  $k$ -nearest neighbor algorithm (KNN), and linear discriminant analysis (LDA). In this research, the four different output coding approaches of the multi-class LS-SVM: error correcting output codes (ECOC), minimum output codes (MOC), One vs One (1vs1) and One vs All (1vsA), are also investigated to test which one is the best for the obtained features. The parameters of the proposed classification methods are selected by extensive experimental evaluations. The  $k$ -fold cross validation is employed to test the consistency of the proposed methods. The performance of each approach is evaluated by corrected percentage (CP) in each class, overall classification accuracy (OCA) and false alarm rate (FAR). In order to further evaluate the performances, we compare our proposed methods with some existing well-known algorithms. One of the major findings of this study is that LS-SVM with 1v1 (denoted as LS-SVM.1v1) is the best classifier for the OA.PCA features. To the best of our knowledge, the OA and PCA methods together have not been used on the epileptic EEG data for feature extraction so far.

The rest of the paper is organized as follows. In Section 2, the data used in this study are described, and the proposed methods are presented. Section 3 provides how the performances of the proposed methods are evaluated. Section 4 describes the experimental set-up with results and discussions. Comparisons among the proposed methods and also the



**Fig. 1 – Plot of different class of EEG signals (from top to bottom: Set Z, Set O, Set N, Set F and Set S).**

existing methods are discussed in Section 5. Finally, the conclusions are drawn in Section 6.

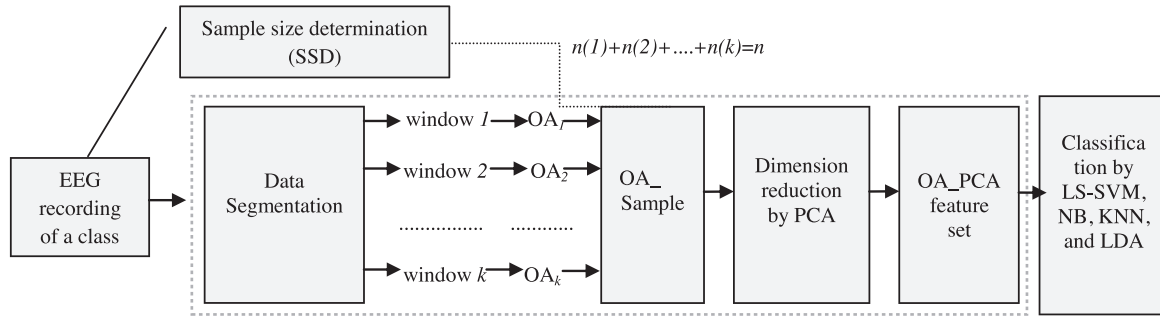
## 2. Data and methods

### 2.1. Data

In our experiments, we used an 'epileptic EEG dataset', which is publicly available in [27,28]. The whole database consists of five EEG data sets (denoted as Set Z, Set O, Set N, Set F, Set S), each containing 100 single-channel EEG signals of 23.6s from five separate classes. Fig. 1 illustrates a plot of different classes of EEG signals for each of the five sets (from top to bottom: Set Z, Set O, Set N, Set F and Set S). Each signal was chosen after visual inspection for artifacts, e.g. due to muscle activities or eye movements. All EEG recordings were made with the same 128-channel amplifier system, using an average common reference. The recorded data was digitised at 173.61 data points per second using 12-bit resolution. Band-pass filter settings were 0.53-40 Hz (12 dB/oct). Set Z and Set O were collected from surface EEG recordings of five healthy volunteers with eyes open and eyes closed, respectively. Set N, Set F and Set S were collected from the EEG records of the pre-surgical diagnosis of five epileptic patients. Signals in Set N and Set F were recorded in seizure-free intervals from five epileptic patients from the hippocampal formation of the opposite hemisphere of the brain and from within the epileptogenic zone, respectively. Set S contains the EEG records of five epileptic patients during seizure activity.

### 2.2. Methods

This study develops a new framework for classifying epileptic EEG signals presented in Fig. 2. The research work investigates and explores whether the OA-based PCA features, (named OA.PCA) are suitable in the epileptic EEG signal classification, and also evaluates which classifier is more suitable for the feature set. The four well-known classifiers, such as least square



**Fig. 2 – Architecture of the proposed system for the classification of epileptic EEG signals. Note:**  $OA_i$  = optimum allocation in the  $i$ th window;  $OA\_Sample$ : selected sample by the OA together from each window;  $OA\_PCA$ : feature vector set obtained by optimum allocation and principal component analysis;  $SSD = n$ ;  $n(1)$  = sample size by  $OA_1$  from window 1;  $n(2)$  = sample size by  $OA_2$  from window 2 and  $n(k)$  = sample size by  $OA_k$  from window  $k$ .

support vector machine (LS-SVM), naive bayes classifier (NB),  $k$ -nearest neighbor algorithm (KNN) and linear discriminant analysis (LDA), are tested on the  $OA\_PCA$  features set as the input. It can be seen from Fig. 2 that the proposed methodology is divided into six major parts: sample size determination (SSD), data segmentation, sample selection by OA (denoted as  $OA\_Sample$ ), dimension reduction by PCA, OA based PCA features (named as  $OA\_PCA$ ) and the classification part by the LS-SVM, NB, KNN and LDA. Brief explanations of these six parts are provided below.

#### 2.2.1. Sample size determination (SSD)

First step of the proposed method is to determine an appropriate sample size from an EEG channel data of a class. From statistical sense, sample is a set of observations from a population. In this study, the entire EEG data of a class (class means a specific category of EEG data. e.g. class 1 (Set Z): health persons with eyes open; class 2 (Set O): health persons with eyes closed etc.) is considered as a population where a sample is considered as a representative part of the population. An observation in a sample is called sample unit, and the sample size is the number of observations that are included in a sample. In this paper, the following formulas (Eqs. (1) and (2)) [29,30] is used to calculate the desired samples for each EEG class data.

$$n = \frac{Z^2 \times p \times (1 - p)}{e^2} \quad (1)$$

where  $n$  is the desired sample size;  $Z$  is the standard normal variate (the value for  $Z$  is found in statistical tables which contain the area under the normal curve) for the desired confidence level (1.96 for 95% confidence level and 2.58 for 99% confidence level) [31];  $p$  is the estimated proportion of an attribute, that is present in the population;  $e$  is the margin of error or the desired level of precision (e.g.  $e = 0.01$  for 99–100% confidence interval). If population is finite, the required sample size is given by

$$n_{\text{new}} = \frac{n}{1 + (n - 1)/N} \quad (2)$$

where  $N$  is the population size.

If the estimator  $p$  is not known, 0.50 (50%) is used as it produces the largest sample size. The larger sample size, the

more sure we can be that their answers truly reflect the population. In this research, we consider  $p = 0.50$  so that the sample size is maximum and  $Z = 2.58$  and  $e = 0.01$  for 99% confidence level in Eq. (1). In this study,  $N = 4097$  and thus we obtain from Eq. (2),  $n_{\text{new}} = 3288$  for each and every five classes (Set Z, Set O, Set N, Set F, Set S) of epileptic EEG data.

#### 2.2.2. Data segmentation

EEG signals are non-stationary and stochastic. But, during signal analysis, it is required to make an EEG signal stationary. The term ‘non-stationary’ means that the statistical properties of a signal change over time, such as the signal’s mean, variance, kurtosis, and skewness do not remain constant over the entire duration of the signal, but rather, change from one point in the signal to the next. ‘Stochastic’ refers to signals where the events in the signal occur in a random fashion and self-similar, at the simplest level means that if a portion of a signal is magnified, the magnified signal will look the same and have the same statistical properties as the original signal. Although an overall signal may not be stationary, usually smaller windows, or parts of those signals will exhibit stationarity. An EEG signal is stationary for a small amount of time. That’s why, we partition the recorded EEG time series signals of every category into several segments based on a specific time period to properly account for possible stationarities. In this methodology, we divide the EEG signals of a class into  $k$ -mutually exclusive groups, called ‘window’ considering a particular time period. For any experiments, the number ( $k$ ) of windows is determined empirically over time based on the data size. Suppose the sizes of the windows are  $N_1, N_2, \dots, N_k$ . Hence, the number of observations in window 1, window 2 and  $\dots$  window  $k$  are  $N_1, N_2, \dots, N_k$ , respectively.

In this research, the EEG signals of a class are segmented into four ( $k = 4$ ) windows with respect to a specific time period based on the data structure. Each window contains 5.9 s of EEG data as every channel consists of 4097 data points of 23.6 s. Thus, the sizes of the four windows (window 1, window 2, window 3 and window 4) are  $N_1 = 1024, N_2 = 1024, N_3 = 1024$  and  $N_4 = 1025$ , respectively.

#### 2.2.3. OA\_Sample

Optimum allocation (OA) refers to a method of sample allocation that is designed to provide the most precision. In

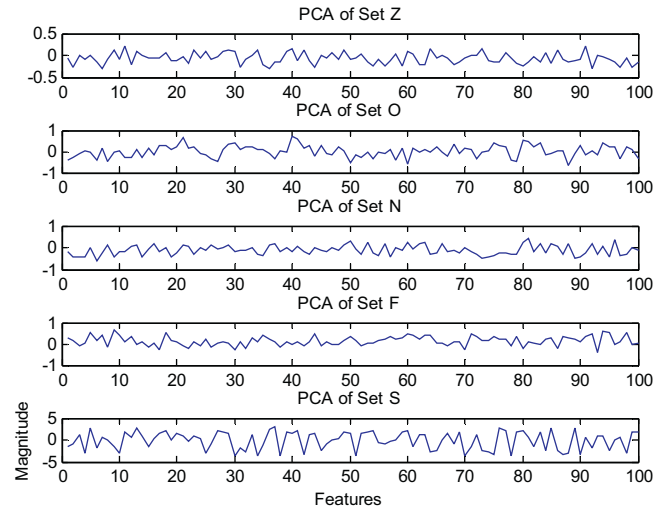


this study, the OA technique is used to determine the number of observations to be selected from different windows. To find out how a given total sample size,  $n$ , should be allocated among the  $k$  windows with the smallest possible variability, we calculate the best sample size for the  $i$ th window using the OA [29,32] technique by Eq. (3). All of the selected samples by the OA from all the windows together is denoted as OA.Sample in this study. The detailed discussion of the OA technique is available in references [29,32].

$$n(i) = \frac{N_i \sqrt{\sum_{j=1}^p s_{ij}^2}}{\sum_{i=1}^k \left( N_i \sqrt{\sum_{j=1}^p s_{ij}^2} \right)} \times n \quad i = 1, 2, \dots, k; \quad j = 1, 2, \dots, p \quad (3)$$

where  $n(i)$  is the required sample size of the  $i$ th window;  $N_i$  is the data size of the  $i$ th window;  $s_{ij}^2$  is the variance of the  $j$ th channel of the  $i$ th window; and  $n$  is the sample size of the EEG recording of a class obtained by Eq. (1). If  $n(1), n(2), \dots, n(k)$  are the sample sizes obtained by Eq. (3) from the window sizes,  $N_1, N_2, \dots, N_k$ , respectively, the sum of all obtained sample sizes from all the windows in a class will be equal to  $n$ , i.e.  $n(1) + n(2) + \dots + n(k) = n$  as shown in Fig. 1.

In this work,  $k$  (the number of windows in each class) = 4 and  $p$  (the number of channels in each class) = 100. For each of the five classes, the sizes of window 1, window 2, window 3 and window 4 are 1024, 1024, 1024 and 1025, respectively. The numbers of observations are determined to be selected from each window in a class using the OA by Eq. (3) as reported in Table 1. From Table 1, it is seen that in Set Z, the sample sizes: 797, 822, 837 and 832, are obtained by the OA from window 1, window 2, window 3 and window 4, respectively. The total of these four samples is 3288 that is equal to the total sample size,  $n = 3288$ . Note that the total sample size,  $n$  of a class (e.g. Set Z or Set O, Set N or Set F or Set S) is allocated among different windows with the smallest possible variability by the OA procedure. As shown in Table 1, the sample sizes are not equal in every window of a class due to different the variability of the observations in different windows. If the variability within a window is large, the size of the sample from that window is also large. On the other hand, if the variability of the observations within a window is small, the sample size will be small in that window. Finally, we create an OA.Sample for Set Z such as {797,822,837,832} which contains 3288 sample units. Similarly, we acquire the OA.Sample from Sets O, N, F, S, individually, which have 3288 sample units, for example {815, 840, 805, 828} for Set O; {839, 841, 780, 828} for Set N; {828, 833, 788, 839} for Set F; and {833, 844, 815, 796} for Set S. It is worthy to mention that each OA.Sample set has 100 dimensions because each class contains 100 channels of EEG data in the epileptic EEG database. Thus, the OA.Sample for each class consists of 3288 observations of 100 dimensions and all of the OA.Samples together from each of the five classes consists of 16,440 observations of 100 dimensions. The entire set ( $16,440 \times 100$ ) of the OA.Samples from the five classes is denoted as OA.Sample set in the experiments of this study.



**Fig. 3 – An example of PCA features for five classes EEG data sets.**

#### 2.2.4. Dimension reduction by PCA

At this step, the PCA is used to reduce the dimensionality of the OA.Sample set and also to generate fewer numbers of uncorrelated variables that is utilized as features for better classification of epileptic EEG signals. Generally the recorded multi-channel EEG signals are huge in number, containing a large amount of redundant information and highly correlated. Often different signals from different scalp sites do not provide the same amount of discriminative information. The PCA is a powerful technique to transform a number of correlated variables into a smaller number of uncorrelated variables called principal components, if there is a large amount of redundant information and a large number of correlated variables in a dataset. The principal components represent the most informative data carried by the original signals to provide the discriminative information about those signals. Thus, the PCA features work better in the EEG signal classification. A detailed explanation about the PCA is found in references [33–36]. Fig. 3 presents an example of the PCA features for the five classes of the EEG signals. In this figure, we consider 100 feature points for each of the five classes. The patterns of the PCA features for each class are similar as their original signal patterns (as shown in Fig. 1).

In this work, we apply the PCA method to the OA.Sample set of five classes that have 16,440 data points of 100 dimensions. Here eigenvectors with eigen values greater than one are chosen. Considering eigen values greater than one, we obtain a principal component set which contains 16,440 data points of 37 dimensions. Fig. 4 illustrates a box plot of whole feature vector set for five class datasets. The experimental results show that the EEG channels data are highly correlated with each other and the PCA components represent 68.216% of the variation that confirms an appropriate use of the PCA approach for this research. Thus the obtained principal components of the OA.Sample set are used as the features denoted as 'OA.PCA'.

#### 2.2.5. OA.PCA feature set

After reducing dimensions of the OA.Sample set by the PCA, the new generated feature vector set is denoted as 'OA.PCA'. In this

**Table 1 – Calculated sample size for each window by the OA scheme.**

Segment	Sample sizes obtained				
	Set Z	Set O	Set N	Set F	Set S
Window 1	797 (1024)	815 (1024)	839 (1024)	828 (1024)	833 (1024)
Window 2	822 (1024)	840 (1024)	841 (1024)	833 (1024)	844 (1024)
Window 3	837 (1024)	805 (1024)	780 (1024)	788 (1024)	815 (1024)
Window 4	832 (1025)	828 (1025)	828 (1025)	839 (1025)	796 (1025)
Total	3288	3288	3288	3288	3288

Note: In each cell, the number inside the parentheses is the window size (e.g. (1024) in the 1st cell) and the number outside of the parentheses is the calculated sample size (e.g. 797 in the 1st cell) obtained by the OA.

study, we obtain an OA\_PCA feature set that has 16,440 data points of 37 dimensions as discussed in the last section. This feature vector set is divided into a training set and a testing set using a 6-fold cross validation method, which is discussed in Section 3. In each of the six trials, the training set consists of 13,700, and the testing set consists of 2740 observations with 37 dimensions. Thus, the percentages of the training set and testing set are 83.33% and 16.67%, respectively. In this study, the training set is applied to train a classifier and the testing vectors are used to evaluate the accuracy and the effectiveness of the chosen classifiers in the classification. The testing set is used as the input to the classification algorithm. The obtained OA\_PCA feature set is fed to each of the four classifiers discussed in the following section.

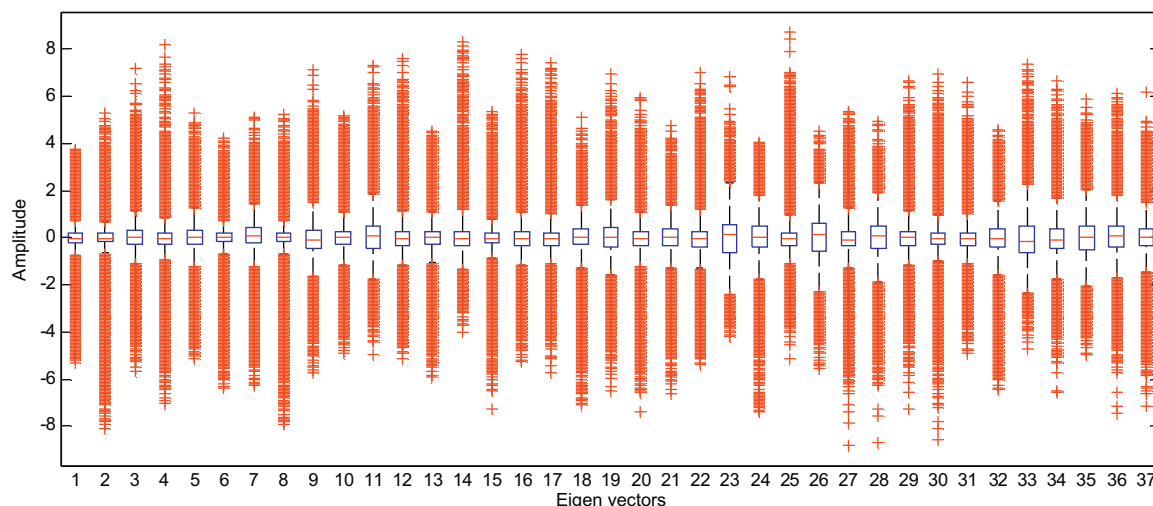
#### 2.2.6. Classification by the LS-SVM, NB, KNN and LDA

After the feature extraction, the feature vector set, OA\_PCA, is forwarded to each mentioned classifier. In order to choose the most appropriate classifier for the OA\_PCA feature set, this study employs four prominent classifiers: LS-SVM, NB, KNN, and LDA, for the classification of epileptic EEG signals. The brief explanation of those methods is provided below.

**Least square support vector machine (LS-SVM):** The LS-SVM proposed by Suykens and Vandewalle [37] becomes one of the most successful classifying models. The LS-SVMs use equality constraints instead of inequality constraints and solve linear equations instead of the quadratic programming that reduces the computational cost of the LS-SVM [38–40] compared to

the original SVM. A straightforward extension of LS-SVMs to multiclass problems has been proposed by Suykens and Vandewalle [38]. A common way of solving the multiclass categorization problem is to reformulate the problem into a set of binary classification problems using four output coding schemes: ECOC, MOC, 1vs1 and 1vsA. The detailed description of the multiclass LS-SVM algorithms can be found in literature [37–43]. In this research, the multiclass LS-SVM with four output coding schemes: ECOC, MOC, 1v1 and 1vA is denoted as LS-SVM\_ECOC, LS-SVM\_MOC, LS-SVM\_1v1 and LS-SVM\_1vA, respectively. Then they are all applied with the radial basis function (RBF) for the classification of multiclass EEG signals, using the OA\_PCA features as the input. Before the classification procedure, the hyper parameters of the LS-SVM ( $\gamma$  and  $\sigma^2$ ) with RBF kernel in each of the four output coding schemes are selected through the extensive experimental evaluations in this research, as discussed in Section 4.1.

**Naive Bayes classifier (NB):** A NB is a probabilistic classifier based on Bayes' theorem with strong (naive) independence assumptions. The NB assumes all the feature nodes are to be independent of each other for a given class, and typically the feature variables are assumed to have Gaussian distribution if they are continuous. The classification results are determined by the posterior probability  $P(y|x_1, x_2, x_3, \dots, x_n)$  described in detail in references [44,45], which can be transformed by using the chain rule and Bayes' Theorem. In this study, the class node,  $y$ , presents the five categories of the epileptic EEG signals and the feature nodes  $(x_1, x_2, x_3, \dots, x_n)$  represent the OA\_PCA

**Fig. 4 – Illustration of the entire feature vector set**

features. A detailed explanation of this method is available in references [44–46].

**k-Nearest neighbor algorithm (KNN):** The idea behind the KNN algorithm is quite straightforward in classifying objects based on the closest training observations presented in the feature space [34,47]. Here, an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its  $k$  nearest neighbors ( $k$  is a positive integer, typically small). If  $k = 1$ , then the object is simply assigned to the class of that single nearest neighbour [47,48]. The only adjustable parameter in the model is  $k$ , the number of the nearest neighbors to include in the estimate of a class membership. The value of  $P(y|x)$  is calculated simply as the ratio of members of class  $y$  among the  $k$  nearest neighbors of  $x$ . A detailed discussion of this method is available in references [34,47–49]. In this study, an appropriate value of  $k$  is selected after the experimental assessments, as described in Section 4.1.

**Linear discriminant analysis (LDA):** A LDA (also known as Fisher's LDA) builds a predictive model for a group membership. The LDA uses hyper-planes to separate data representing different classes [34]. The model is composed of a discriminant function based on linear combinations of predictor variables. The purpose of the LDA is to maximally separate the groups and to determine the most and parsimonious

$$FAR_i = \frac{\text{Number of of normal cases that are detected as not a normal cases in } i\text{th category}}{\text{Total number of normal cases}} \times 100 \quad (6)$$

way to separate groups and to discard variables which are little related to group distinctions for online systems. This technique has a very low computational requirement which makes it suitable. A detailed discussion of this method is available in references [34,50]. In this research, the OA.PCA features are used as the predictor variables in this model for epileptic EEG data classification.

### 3. Performance evaluation

In order to evaluate performances of the proposed framework, this study applies a  $k$ -fold cross-validation [51,52] method. In  $k$ -fold cross-validation procedure, a data set is partitioned into  $k$  mutually exclusive subsets (each subset is called 'fold') of equal size and the method is repeated  $k$  times. Each time, one of the subsets (folds) is used as a test set and the other  $k - 1$  subsets (folds) are put together to form a training set. Then the average accuracy across all  $k$  trials is computed.

In this study, we select  $k = 6$  as the 6-fold cross-validation is found adequate for this dataset. As mentioned in Section 2.2.5, we obtain a total of 16,440 feature vectors of 37 dimensions where each of the five classes holds 3288 feature vectors of the same dimensions. In this research, we divide the whole feature sets into six subsets where each subset is called fold. Each fold consists of 2740 feature vectors of the five classes taking 548 feature vectors from each class. Fig. 5 presents a design of how the extracted feature vectors of this study are partitioned into six mutually exclusive subsets (folds) according to the  $k$ -fold cross-validation system, and the procedure

is repeated 6 times. Each time, one subset (fold) is used as a testing set and the remaining five subsets (folds) are used as a training set. As shown in Fig. 5, in each iteration, the testing set is denoted as bold face and other five folds together is used as the training set. Thus, the training set consists of 13,700 feature vectors of 37 dimensions and the testing set contains 2740 feature vectors of the same dimension. The classification performances from each of six iterations on the testing set are obtained for all categories of the epileptic EEG signals and summarized in Table 6 in Section 4.2.

In this research, the performances of the proposed methods are assessed based on different statistical measures, such as corrected percentage (CP) in each class, overall classification accuracy (OCA) and false alarm rate (FAR). Their formulas are given below [14,51–57]:

- Corrected percentage (CP) in each class (also called sensitivity)

$$CP_i = \frac{\text{Number of correctly classified cases in } i\text{th class}}{\text{Actual number of cases in } i\text{th class}} \times 100 \quad (4)$$

- Overall classification accuracy (OCA)

$$OCA = \frac{\text{Number of total correct decisions for all classes}}{\text{Total number of cases for all classes}} \times 100 \quad (5)$$

- False alarm rate (FAR)

## 4. Experimental set-up

In this section, at first we select the parameters of the four mentioned classifiers: LS-SVM, NB, KNN and LDA, because the classification performances of a classifier depend on the values of the parameters as discussed in Section 4.1. Then we investigate the effectiveness of each classifier on the OA.PCA features with the optimally selected parameters discussed in Section 4.2. The classification by the multi-class LS-SVM is carried out in MATLAB (version 7.14, R2012a) using the LS-SVMlab toolbox (version 1.8) [58]. The classification of the NB method is also performed in MATLAB with the same version. On the other hand, the classification executions for the KNN and LDA methods are executed in SPSS package (version 21) in this study.

### 4.1. Parameter selection

As mentioned before, this study uses four classification methods, such as, LS-SVM, NB, KNN and LDA. The four output coding schemes of the multi-class LS-SVM: ECOC, MOC, 1vs1 and 1vsA, are explored in this research. As there are no specific guidelines to set the values of these parameters of the mentioned classifiers, this study presents a new way to select the parameters. In this paper, the RBF kernel function is employed for the multi-class LS-SVM as an optimal kernel function over different kernel functions that were tested. The multi-class LS-SVM with RBF has two important parameters,  $\gamma$  and  $\sigma^2$ , which should be appropriately chosen for achieving the desired performance as these parameters play a significant role in the classification performance. The NB classifier is a probabilistic

1 <sup>st</sup> iteration	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6
2 <sup>nd</sup> iteration	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6
3 <sup>rd</sup> iteration	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6
4 <sup>th</sup> iteration	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6
5 <sup>th</sup> iteration	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6
6 <sup>th</sup> iteration	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Fold-6

**Fig. 5 – Structure of performing 6-fold cross validation process.**

classifier based on Bayes' Theorem and there is no parameter in this method. The KNN model has only one parameter  $k$  which refers to the number of nearest neighbors. By varying  $k$ , the model can be made more or less flexible. In this study, the appropriate value of the model parameter,  $k$  is selected after experimental evaluations. In the LDA method, the parameters are obtained automatically through maximum likelihood estimation (MLE) method.

From the above discussion, it is clear that we only find out the optimal parameter values,  $\gamma$  and  $\sigma^2$ , for the LS-SVM classifier of each output coding system and an appropriate  $k$  value for the KNN classifier. As mentioned before, we denote the LS-SVM for the four output coding schemes: ECOC, MOC, 1v1 and 1vA, as LS-SVM\_ECOC, LS-SVM\_MOC, LS-SVM\_1v1 and LS-SVM\_1vA, respectively. In order to select appropriate combinations of  $(\gamma, \sigma^2)$  for all four LS-SVM approaches, we set up the ranges of  $\gamma$  and  $\sigma^2$  values as  $\gamma = 1, 10, 100, 1000$  and

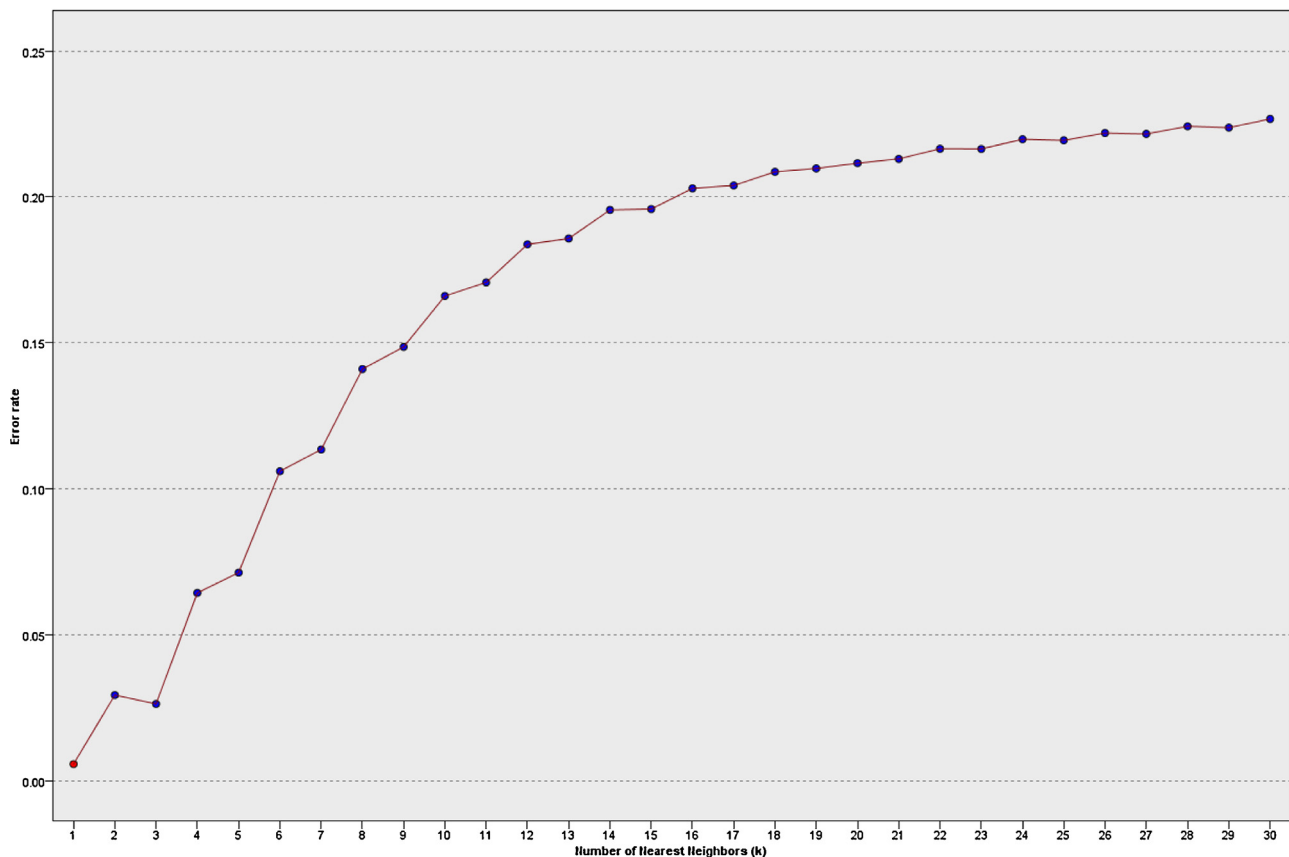
**Table 2 – Selection of hyper parameters  $(\gamma, \sigma^2)$  of the LS-SVM\_ECOC classifier.**

LS-SVM_ECOC		
Fold	Parameters $(\gamma, \sigma^2)$	OCA (%)
Fold-1	$\gamma = 100, \sigma^2 = 10$	99.96
Fold-2	$\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$ ; $\gamma = 10, \sigma^2 = 10$	100.00
Fold-3	$\gamma = 1, \sigma^2 = 10$ ; $\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$	100.00
Fold-4	$\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$	99.96
Fold-5	$\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$	100.00
Fold-6	$\gamma = 100, \sigma^2 = 10$	99.89

Note: Selected parameters for the LS-SVM\_ECOC approach:  $\gamma = 100$ ,  $\sigma^2 = 10$ .

$\sigma^2 = 1, 10, 100, 1000$  for the experimental evaluations. Thus, we obtain classification outcomes for all possible combinations of  $\gamma$  and  $\sigma^2$  and then take one combination as an optimal combination for each fold that reports the highest overall classification accuracy (OCA) rate for that fold. In each fold, the best combination of  $(\gamma, \sigma^2)$  for the LS-SVM\_ECOC, LS-SVM\_MOC, LS-SVM\_1v1 and LS-SVM\_1vA classifiers with their OCAs are presented in Tables 2–5, respectively. The selected combinations of  $(\gamma, \sigma^2)$  for each of the LS-SVMs is highlighted in bold font in every fold.

**k Selection Error log**



**Fig. 6 – k Selection error log in fold-1 of the testing set**



**Table 3 – Selection of hyper parameters ( $\gamma$ ,  $\sigma^2$ ) of the LS-SVM.MOC.**

LS-SVM.MOC		
Fold	Parameters ( $\gamma, \sigma^2$ )	OCA (%)
Fold-1	$\gamma = 100, \sigma^2 = 10$	99.96
Fold-2	$\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$	100.00
Fold-3	$\gamma = 1, \sigma^2 = 10$ ; $\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$	100.00
Fold-4	$\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$	99.96
Fold-5	$\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$	100.00
Fold-6	$\gamma = 100, \sigma^2 = 10$	99.89
Note: Selected parameters for the LS-SVM.MOC approach: $\gamma = 100, \sigma^2 = 10$ .		

**Table 4 – Selection of hyper parameters ( $\gamma$ ,  $\sigma^2$ ) of the LS-SVM.1v1.**

LS-SVM.1v1		
Fold	Parameters ( $\gamma, \sigma^2$ )	OCA (%)
Fold-1	$\gamma = 1, \sigma^2 = 10$ ; $\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 100$ ; $\gamma = 1000, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 100$	100.00
Fold-2	$\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 100$ ; $\gamma = 1000, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 100$	100.00
Fold-3	$\gamma = 1, \sigma^2 = 10$ ; $\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 100$ ; $\gamma = 1000, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 100$	100.00
Fold-4	$\gamma = 1, \sigma^2 = 10$ ; $\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$	100.00
Fold-5	$\gamma = 1, \sigma^2 = 10$ ; $\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 100$	100.00
Fold-6	$\gamma = 10, \sigma^2 = 10$	100.00
Note: Selected parameters for the LS-SVM.1v1 approach: $\gamma = 10$ , $\sigma^2 = 10$ .		

**Table 5 – Selection of hyper parameters ( $\gamma$ ,  $\sigma^2$ ) of the LS-SVM.1vA.**

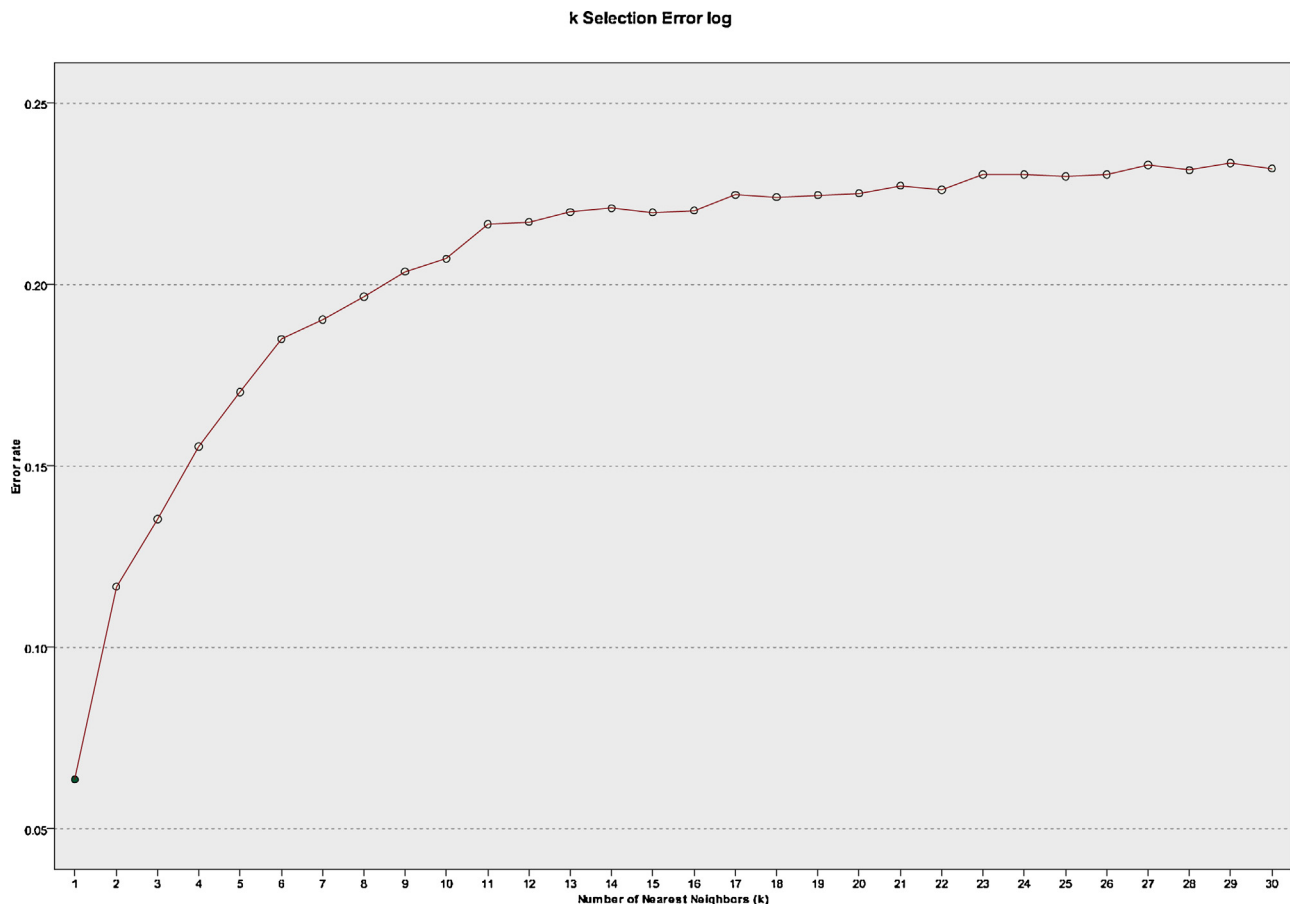
LS-SVM with 1vA		
Fold	Parameters ( $\gamma, \sigma^2$ )	OCA (%)
Fold-1	$\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$	99.96
Fold-2	$\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$	100.00
Fold-3	$\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$	100.00
Fold-4	$\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$	99.96
Fold-5	$\gamma = 10, \sigma^2 = 10$ ; $\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$	100.00
Fold-6	$\gamma = 100, \sigma^2 = 10$ ; $\gamma = 1000, \sigma^2 = 10$	99.85
Note: Selected parameters for the LS-SVM.1vA approach: $\gamma = 100$ , $\sigma^2 = 10$ .		

Table 2 reports that one combination of ( $\gamma, \sigma^2$ ) is obtained in fold-1, three combinations in fold-2, four combinations in fold-3, two combinations in fold-4, three combinations in fold-5 and one combination in fold-6. From this table, it is seen that  $\gamma = 100, \sigma^2 = 10$  is the common combination for the LS-SVM.ECOC approach in each fold. Thus we select  $\gamma = 100, \sigma^2 = 10$  as an optimal combination for the LS-SVM.ECOC method. Table 3 also shows that  $\gamma = 100, \sigma^2 = 10$  is the common combination that produces highest performance in every fold for the LS-SVM.MOC method. Hence  $\gamma = 100, \sigma^2 = 10$  is considered as the best combination for this method. In Table 4, it can be seen that the selected optimal parameters for the LS-SVM.1v1 classifier is  $\gamma = 10, \sigma^2 = 10$  that yields 100% classification accuracy in each fold. Table 5 reports that  $\gamma = 100, \sigma^2 = 10$  is the best pair for the hyper parameters of the LS-SVM.1vA method.

As mentioned before, there is only one model parameter in the KNN classifier that is  $k$ . The choice of  $k$  is essential in building up a model which can be regarded as one of the most important factors of the model that can strongly influence the quality of predictions.  $k$  should be set to a suitable value to minimize the probability of misclassification. The best choice of  $k$  depends upon the data. In this study, we select appropriate  $k$  value in automatic process following  $k$  selection error log as there is no a simple rule for selecting  $k$ . We consider the range of  $k$  value being in between 1 and 30. We pick an appropriate  $k$  value that results in the lowest error rate in each fold as the lowest error rate indicates having the best model. In the experimental results, we obtain the lowest error rate for  $k = 1$  in every fold. Figs. 6 and 7 show the two examples for fold-1 and fold-2 of the testing data set. In these two figures, it is seen that the number of nearest neighbors ( $k$ ) are plotted in X-axis and the error rate of the KNN model for each  $k$  value is presented in Y-axis. The error rate increases in conjunction with an increase in the number of nearest neighbors ( $k$ ). For example, in Fig. 6, the error rate is about 0.01 for  $k = 1$  while it is about 0.07 for  $k = 5$ . Again, in Fig. 7, the error rate is about 0.06 for  $k = 1$  when it is about 0.17 for  $k = 5$ . Hence the appropriate value of  $k$  is 1 for this epileptic dataset. Thus the lowest error rate is obtained with this value in each fold of the dataset.

#### 4.2. Results and discussions

Table 6 presents the classification results for the OA.PCA feature set with the classifiers: LS-SVM.ECOC, LS-SVM.MOC, LS-SVM.1v1, LS-SVM.1vA, NB, KNN and LDA on the epileptic EEG data. In this study, the average of the corrected percentage (CP) in a category (calculated by Eq. (3)) of the six folds is named as CP.average, and the variation among results from the six folds is denoted as standard deviation (SD). On the other hand, the average of the overall classification accuracy (OCA) (calculated by Eq. (5)) for the six folds is named as OCA.average, and the deviation among the six folds is denoted as standard deviation (SD). As shown in Table 6, the highest classification performance is achieved for the LS-SVM.1v1 classifier, which is 100% for each and every category. The LS-SVM.ECOC and LS-SVM.MOC classifiers get the second position and the LS-SVM.1vA in the third position according to the CP.average and OCA.average for the OA.PCA feature set. The LDA classifier



**Fig. 7 – k Selection error login fold-2 of the testing set**

yields the lowest performance among the seven classification methods. It can be seen from Table 6 that the LS-SVM classifier with each of the four output coding systems produces relatively better performances for the OA.PCA feature set compared to the other three classifiers: the NB, KNN and LDA. On the other hand, the NB produces a better performance than those of the KNN and LDA. Furthermore, Table 6 reports that the SD for every classifier is very low, which leads to the conclusion that the mentioned classifiers are consistent for the OA.PCA features. It is also observed that the SD of the classification performance for the LS-SVM.1v1 classifier is zero,

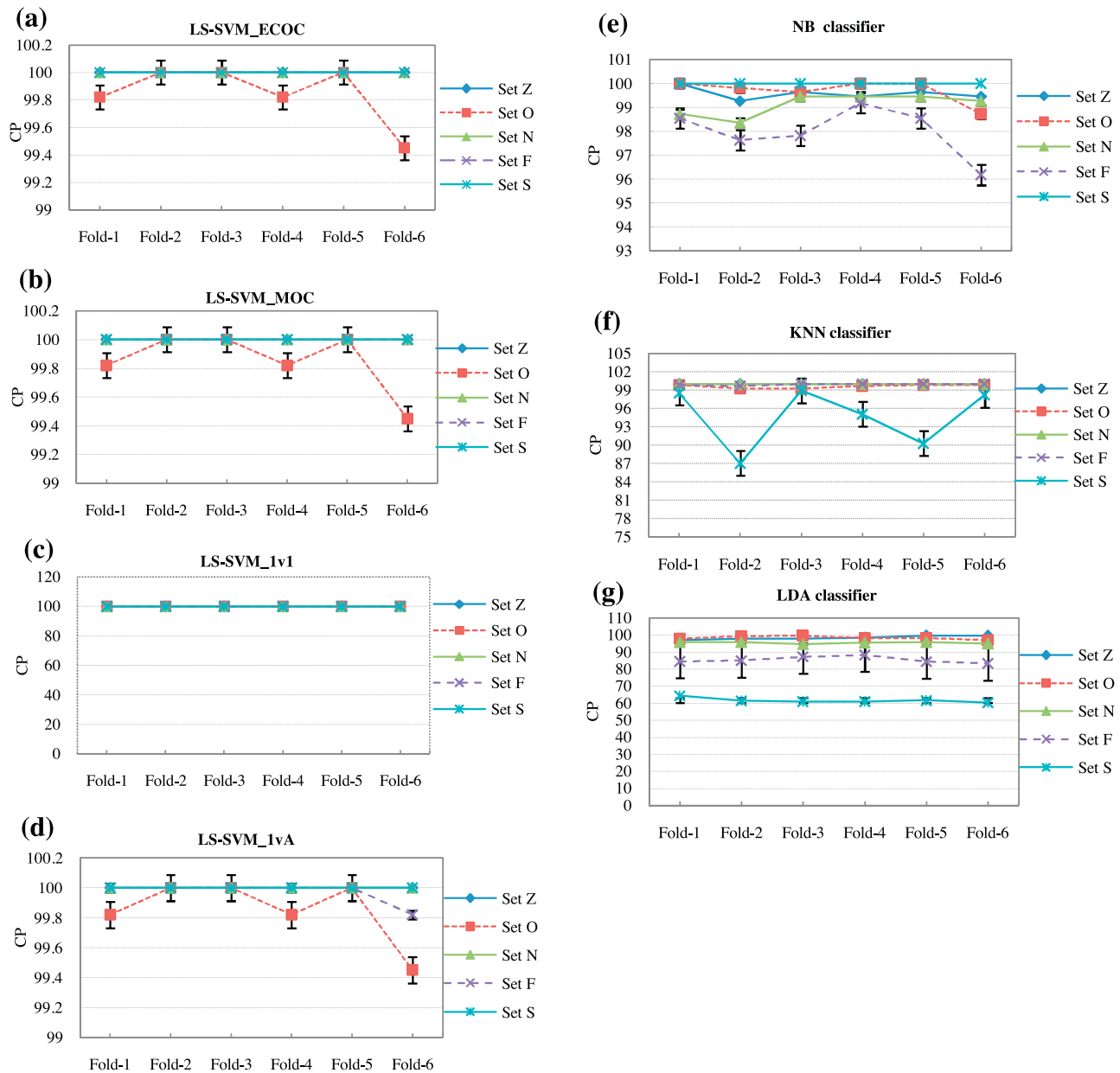
which indicates the robustness of the LS-SVM.1v1 for obtained features. The lower value of the SD represents the consistency of the proposed method. Hence the results demonstrate that the LS-SVM is a superior classifier for the epileptic EEG signal classification. The LS-SVM.1v1 is the best one.

In order to get a clear idea on, how the 6-fold cross-validation system produces the CP in each of the six folds in each class for every reported classifier, we provide Fig. 8(a)–(g). These figures present patterns of the corrected percentage (CP) in each class against each of the six folds. Fig. 8(c) shows that there are no fluctuations in the performance among the six

**Table 6 – Classification results on the epileptic EEG data.**

Different classifiers	CP.average $\pm$ SD					OCA.average $\pm$ SD
	Set Z	Set O	Set N	Set F	Set S	
LS-SVM.ECOC	100.0 $\pm$ 0.00	99.85 $\pm$ 0.21	100.0 $\pm$ 0.00	100.0 $\pm$ 0.00	100.0 $\pm$ 0.00	99.97 $\pm$ 0.04
LS-SVM.MOC	100.0 $\pm$ 0.00	99.85 $\pm$ 0.21	100.0 $\pm$ 0.00	100.0 $\pm$ 0.00	100.0 $\pm$ 0.00	99.97 $\pm$ 0.04
LS-SVM.1v1	100.0 $\pm$ 0.00	100.0 $\pm$ 0.00	100.0 $\pm$ 0.00	100.0 $\pm$ 0.00	100.0 $\pm$ 0.00	100.0 $\pm$ 0.00
LS-SVM.1vA	100.0 $\pm$ 0.00	99.85 $\pm$ 0.21	100.0 $\pm$ 0.00	99.96 $\pm$ 0.07	100.0 $\pm$ 0.00	99.96 $\pm$ 0.06
NB classifier	99.58 $\pm$ 0.25	99.70 $\pm$ 0.50	99.12 $\pm$ 0.47	97.98 $\pm$ 1.05	100 $\pm$ 0.00	99.24 $\pm$ 0.31
KNN classifier	100.0 $\pm$ 0.00	99.61 $\pm$ 0.27	100.0 $\pm$ 0.00	99.91 $\pm$ 0.15	94.67 $\pm$ 4.95	98.82 $\pm$ 1.01
LDA classifier	98.32 $\pm$ 1.03	98.25 $\pm$ 0.93	95.47 $\pm$ 0.46	85.40 $\pm$ 1.94	61.51 $\pm$ 1.41	87.79 $\pm$ 0.44

Set Z: normal persons with eyes open; Set O: normal persons with eyes closed; Set N: epileptic patients during seizure-free from the hippocampal formation of the opposite hemisphere of the brain; Set F: epileptic patients during seizure-free within the epileptogenic zone; Set S: epileptic patient during seizure activity.



**Fig. 8 – Individual classification performances of each of the six folds in each class for the proposed classifiers: (a) LS-SVM\_ECOC, (b) LS-SVM\_MOC, (c) LS-SVM\_1v1, (d) LS-SVM\_1vA, (e) NB, (f) KNN, and (g) LDA. Error bars represent the standard error.**

folds in each of the five classes for the LS-SVM\_1v1 classifier. The results indicate that the proposed method is fairly stable. It is also seen from Fig. 8(a), (b), (d) and (e)–(g) that the fluctuations in the performance among the different folds are very little in each class. From these figures, it is observed that the pattern of the CP in each fold for every class is very consistent, which indicates the robustness of the methods. The error bars in the graphs represent the standard errors indicating the fluctuations in the performances among the folds.

Table 7 illustrates the false alarm rate (FAR) (sometimes called ‘probability of false detection’) for the seven classifiers in each of the six folds for Set O, Set N, Set F and Set S. In this study, we consider Set Z as a normal case and compute the FAR using Eq. (5). As Shown in Table 7, the FARs are zero

in each of the folds in every class for the LS-SVM\_ECOC, LS-SVM\_MOC, LS-SVM\_1v1, LS-SVM\_1vA and KNN. That means, there are no incorrect responses in each fold in every class for those classifiers. For most of the cases, the FARs are also zero for the NB and LDA classifiers. For the value of the FAR, zero indicates that normal cases are detected as normal with 100% accuracy.

## 5. Comparisons

In order to choose the best classifier for the OA.PCA features, we compare our proposed methods in terms of their OCA against each fold and present the results in Fig. 9. It can be

**Table 7 – Obtained false alarm rate (FAR) for each of the proposed classifiers.**

Different classifiers	Data set	FAR			
		Set O	Set N	Set F	Set S
LS-SVM.ECOC	Fold-1	0.00	0.00	0.00	0.00
	Fold-2	0.00	0.00	0.00	0.00
	Fold-3	0.00	0.00	0.00	0.00
	Fold-4	0.00	0.00	0.00	0.00
	Fold-5	0.00	0.00	0.00	0.00
	Fold-6	0.00	0.00	0.00	0.00
LS-SVM.MOC	Fold-1	0.00	0.00	0.00	0.00
	Fold-2	0.00	0.00	0.00	0.00
	Fold-3	0.00	0.00	0.00	0.00
	Fold-4	0.00	0.00	0.00	0.00
	Fold-5	0.00	0.00	0.00	0.00
	Fold-6	0.00	0.00	0.00	0.00
LS-SVM.1v1	Fold-1	0.00	0.00	0.00	0.00
	Fold-2	0.00	0.00	0.00	0.00
	Fold-3	0.00	0.00	0.00	0.00
	Fold-4	0.00	0.00	0.00	0.00
	Fold-5	0.00	0.00	0.00	0.00
	Fold-6	0.00	0.00	0.00	0.00
LS-SVM.1vA	Fold-1	0.00	0.00	0.00	0.00
	Fold-2	0.00	0.00	0.00	0.00
	Fold-3	0.00	0.00	0.00	0.00
	Fold-4	0.00	0.00	0.00	0.00
	Fold-5	0.00	0.00	0.00	0.00
	Fold-6	0.00	0.00	0.00	0.00
NB classifier	Fold-1	0.00	0.00	0.00	0.00
	Fold-2	0.00	0.0036	0.0036	0.00
	Fold-3	0.00	0.0036	0.00	0.00
	Fold-4	0.00	0.0055	0.00	0.00
	Fold-5	0.00	0.0036	0.00	0.00
	Fold-6	0.00	0.0055	0.00	0.00
KNN classifier	Fold-1	0.00	0.00	0.00	0.00
	Fold-2	0.00	0.00	0.00	0.00
	Fold-3	0.00	0.00	0.00	0.00
	Fold-4	0.00	0.00	0.00	0.00
	Fold-5	0.00	0.00	0.00	0.00
	Fold-6	0.00	0.00	0.00	0.00
LDA classifier	Fold-1	0.00	0.011	0.020	0.00
	Fold-2	0.00	0.0055	0.015	0.0018
	Fold-3	0.0018	0.013	0.0073	0.00
	Fold-4	0.0018	0.0073	0.00	0.0073
	Fold-5	0.00	0.0036	0.0018	0.00
	Fold-6	0.00	0.00	0.0055	0.00

seen from the figure that the patterns of the OCA for the LS-SVMs in each of the four output coding systems are very close. That justifies that the four graph lines of the LS-SVM.ECOC, LS-SVM.MOC, LS-SVM.1v1, LS-SVM.1vA are laying in one line showing a higher performance compared to the other classifiers. This figure also demonstrates that the OCA of the LDA is a bit lower compared to the other six classifiers. From Table 6 and Fig. 9, it is clear that the LS-SVM results in a better performance with the OA.PCA feature set in the EEG signals classification than the other reported classifiers.

In order to further examine the efficiencies of the proposed classifiers with the OA.PCA features, we provide the comparisons of our approaches with some recently reported methods. In Table 8, we list a comparison of different methods in terms

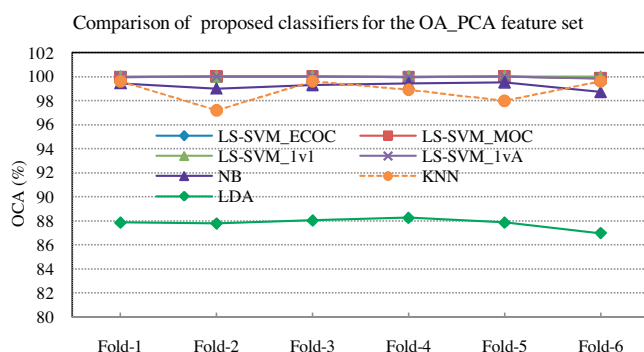
**Table 8 – Comparison with the existing methods on epileptic EEG database.**

Authors	Methods	OCA (%)
Siuly and Li (the proposed)	OA + PCA + LS-SVM.ECOC	99.97
	OA + PCA + LS-SVM.MOC	99.97
	<b>OA + PCA + LS-SVM.1v1</b>	<b>100.0</b>
	OA + PCA + LS-SVM.1vA	99.96
	OA + PCA + NB classifier	99.24
	OA + PCA + KNN classifier	98.82
Shen et al. [12]	OA + PCA + LDA classifier	87.79
	Wavelet + ApEn + SVM	99.97
	Wavelet + ApEn + KNN	99.10
Acharjee and Shahnaz [13]	Wavelet + ApEn + RBFNN	99.82
	Twelve Cohen class kernel functions + modular energy + modularentropy + ANN	98.00
Murugavel et al. [14]	WT and Lyapunov + SVM.1vA	96.00
	WT and Lyapunov + PNN	94.00
	WT and Lyapunov + RBFNN	93.00
Ubeyli [15]	Wavelet coefficients + power spectral density + SVM.ECOC	99.20
	Wavelet coefficients + power spectral density + MME classifier	98.68
	Wavelet coefficients + power spectral density + ME	95.00
	Wavelet coefficients + power spectral density + RNN	94.85
	Wavelet coefficients + power spectral density + PNN	95.30
	Wavelet coefficients + power spectral density + CNN	93.48
	Wavelet coefficients + power spectral density + MLPNN	90.48
Ubeyli [16]	Pisarenko, MUSIC, and minimum-norm + SVM.ECOC	99.30
	Pisarenko, MUSIC, and minimum-norm + MLPNN	92.90
Guler and Ubeyli [17]	Wavelet coefficients and lyapunov exponents + SVM.ECOC	99.28
	Wavelet coefficients and lyapunov exponents + PNN	98.05
	Wavelet coefficients and lyapunov exponents + MLPNN	93.63

WT = wavelet transform.

of the OCA for the epileptic EEG database. We present the results from our proposed methods and also the six reported research work [12–17]. The datasets used in these experiments are also the same. The highest OCA rate among the algorithms is highlighted in bold font.





**Fig. 9 – Comparisons of the proposed classifiers for the OA.PCA feature in terms of the OCA.**

From Table 8, it is obvious that the result obtained from the approach LS-SVM.1v1 is the best for this database among the recent reported approaches. The results indicate that the LS-SVM.1v1 approach improves the OCA by 0.03–7.10% over the existing algorithms for the epileptic EEG data.

## 6. Conclusions

This research presents an optimal feature extraction technique, named as OA.PCA, and also searches for a suitable classifier for the feature set to identify multi categories epileptic EEG signals. The proposed schemes demonstrate many advantages, such as high classification performance and very low FAR for all classifiers tested. The main conclusions of this study are summarized as follows:

- The OA.PCA system is very effective for feature extraction for the epileptic EEG data. The experimental results for the proposed classifiers, the LS-SVM.ECOC, LS-SVM.MOC, LS-SVM.1v1, LS-SVM.1vA, NB, KNN and LDA, confirm that the extracted features are very consistent to detect epileptic EEG signals.
- The parameters of the proposed methods are optimally selected through experimental evaluations indicating the reliability of the methods.
- The results show that the proposed LS-SVM.1v1 classifier with the OA.PCA features achieves the best performance compared to the other classifiers with the same features.
- The experimental results also indicate that our proposed approach, LS-SVM.1v1 outperforms the other six recently reported research results with the same epileptic EEG database. It demonstrates that our method is considered the best method for the epileptic EEG signal classification.

This study concludes that the LS-SVM.1v1 with the OA.PCA feature set is a promising technique for the epileptic EEG signals recognition. It can offer great potentials for the development of epilepsy analyses. The proposed technique can assist physicians or doctors to diagnose brain function abnormalities. One limitation for performing the proposed optimum allocation based algorithm is dealing with the problem of considering the window size based on

an empirical evaluation. In the future, we will extend the proposed approach to online and real-time applications in the multiclass classification problems.

## Conflict of interest

We hereby state that we don't have any financial and personal relationships with other people or organizations that could inappropriately influence (bias) this work.

## REFERENCES

- [2] WHO (World Health Organization) Report, (<http://www.who.int/mediacentre/factsheets/fs999/en/index.html>) (accessed February 2011).
- [3] ILAE IBE, WHO, Atlas Epilepsy Care in the World 2005: Global Campaign Against Epilepsy, World Health Organization, 2005.
- [4] E.H. Reynolds, The ILAE<sup>2</sup>/IBE<sup>1</sup>/WHO<sup>3</sup> global campaign against epilepsy: bringing epilepsy out of the shadows, *Epilepsy Behav.* 1 (2000) S3–S8.
- [5] R.G. Beran, L. Hall, J. Michelazzi, An accurate assessment of the prevalence ratio of epilepsy adequately adjusted by influencing factors, *Neuroepidemiology* 4 (2) (1985) 71–81.
- [6] N. Mc Grogan, Neural network detection of epileptic seizures in the electroencephalogram, [Online], Available: <http://www.new.ox.ac.uk/nmcgrogan/work/transfer>, (1999).
- [7] T. Sunil Kumar, V. Kanhangad, R.B. Pachori, Classification of seizure and seizure-free EEG signals using local binary patterns', *Biomed. Signal Process. Control* 15 (2015) 33–40.
- [8] N. Pradhan, P. Sadasivan, G. Arunodaya, Detection of seizure activity in EEG by an artificial neural network: a preliminary study, *Comput. Biomed. Res.* 29 (1996) 303–313.
- [9] R. Sharma, R.B. Pachori, Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions, *Expert Syst. Appl.* 2 (3) (2015) 1106–1117.
- [10] J.d. Bronzino, Principles of Electroencephalography. The Biomedical Engineering Handbook, second ed., CRC Press LLC, Boca Raton, FL, 2000.
- [11] S. Siuly, Y. Li, J. Wu, J. Yang, Developing a logistic regression model with cross-correlation for motor imagery signal recognition, in: IEEE/ICME International Conference on Complex Medical Engineering (ICME 2011), 22–25 May 2011, Harbin, China, 2011, pp. 502–507.
- [12] C.P. Shen, C.C. Chen, S.L. Hsieh, W.H. Chen, J.M. Chen, C.M. Chen, F. Lai, M.J. Chiu, High-performance seizure detection system using a wavelet-approximate entropy-fSVM cascade with clinical validation, *Clin. EEG Neurosci.* (2013), <http://dx.doi.org/10.1177/1550059413483451>.
- [13] P.P. Acharjee, C. Shahnaz, Multiclass epileptic seizure classification using time-frequency analysis of EEG signals, in: Seventh International Conference on Electrical and Computer Engineering, 20–22 December, 2012, Dhaka, Bangladesh, 2012, pp. 260–263.
- [14] A.S.M. Murugavel, S. Ramakrishnan, K. Balasamy, T. Gopalakrishnan, Lyapunov features based EEG signal classification by multi-class SVM, in: World Congress on Information and Communication Technologies, 2011, pp. 197–201.
- [15] E.D. Ubeyli, Decision support systems for time-varying biomedical signals: EEG signals classification, *Expert Syst. Appl.* 36 (2009) 2275–2284.
- [16] E.D. Ubeyli, Analysis of EEG signals by combining eigenvector methods and multiclass support vector machines, *Comput. Biol. Med.* 38 (2008) 14–22.

- [17] I. Guler, E.D. Ubeyli, Multiclass support vector machines for EEG-signal classification, *IEEE Trans. Inf. Technol. Biomed.* 11 (2) (2007) 117–126.
- [18] U.R. Acharya, S.V. Sree, G. Swapna, R.J. Martis, J.S. Suri, Automated EEG analysis of epilepsy: a review, *Knowl. Based Syst.* 45 (2013) 147–165.
- [19] S.H. Lee, J.S. Lim, J.K. Kim, J. Yang, Y. Lee, Classification of normal and epileptic seizure EEG signals using wavelet transform, phase-space reconstruction, and Euclidean distance, *Comput. Methods Programs Biomed.* 116 (1) (2014) 10–25.
- [20] R.J. Oweis, E.W. Abdulhay, Seizure classification in EEG signals utilizing Hilbert–Huang transform, *Biomed. Eng. Online* 10 (2011) 38–52.
- [21] S. Li, W. Zhou, Q. Yuan, S. Geng, D. Cai, Feature extraction and recognition of ictal EEG using EMD and SVM, *Comput. Biol. Med.* 43 (7) (2013) 807–816.
- [22] R.B. Pachori, S. Patidar, Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions, *Comput. Methods Programs Biomed.* 113 (2) (2014) 494–502.
- [23] S. Pravin Kumar, N. Sriraam, P.G. Benakop, B.C. Jinaga, Entropies based detection of epileptic seizures with artificial neural network classifiers, *Expert Syst. Appl.* 37 (2010) 3284–3291.
- [24] S. Siuly, Y. Li, P. Wen, EEG signal classification based on simple random sampling technique with least square support vector machines, *Int. J. Biomed. Eng. Technol.* 7 (4) (2011) 390–409.
- [25] S. Siuly, Y. Li, P. Wen, Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain computer interface, *Comput. Methods Programs Biomed.* 113 (3) (2014) 767–780.
- [26] S. Siuly, Y. Li, Discriminating the brain activities for brain–computer interface applications through the optimal allocation-based approach, *Neural Comput. Appl.* (2014), <http://dx.doi.org/10.1007/s00521-014-1753-3>.
- [27] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C.E. Elger, Indication of non linear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state, *Phys. Rev. E* 64 (2001) 061907.
- [28] EEG, EEG Time Series (Epileptic EEG Data), EEG, 2005, Nov, (<http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html>) [online].
- [29] M.N. Islam, An Introduction to Sampling Methods: Theory and Applications (revised ed.), Book World, Dhaka, 2007.
- [30] R.D. De Veaux, P.F. Velleman, D.E. Bock, Intro Stats, third ed., Pearson Addison Wesley, Boston, MA, 2008.
- [31] Computation of Z, ([http://ci.columbia.edu/ci/premba\\_test/c0331/s6/z.distribution.html](http://ci.columbia.edu/ci/premba_test/c0331/s6/z.distribution.html))
- [32] W.G. Cochran, Sampling Techniques, Wiley, New York, NY, 1977.
- [33] L.J. Cao, K.S. Chua, W.K. Chong, H.P. Lee, Q.M. Gu, A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine, *Neurocomputing* 55 (2003) 321–336.
- [34] R.O. Duda, P.E. Hart, D.G. Strok, Pattern Classification, second ed., John, Wiley & Sons, California, USA, 2001.
- [35] M. Chawla, H. Verma, K. Vinod, ECG modeling and QRS detection using principal component analysis, in: Proceedings of IET International Conference, Paper No. 04, MEDSIP06, Glasgow, UK, 2006.
- [36] M. Turk, A. Pentl, Eigenfaces for recognition, *J. Cogn. Neurosci.* 3 (1991) 71–86.
- [37] J.A.K. Suykens, and J. Vandewalle, Least Square Support Vector Machine classifier, *Neural Processing Letters* 9 (3) 293–300.
- [38] J.A.K. Suykens, T.V. Gestel, J.D. Brabanter, B.D. Moor, J. Vandewalle, Least Square Support Vector Machine, World Scientific, Singapore, 2002.
- [39] Y. Xing, X. Wu, Z. Xu, Multiclass least squares auto-correlation wavelet support vector machine, *ICIC Express Lett.* 2 (4) (2008) 345–350.
- [40] J.A.K. Suykens, J. Vandewalle, Multiclass least squares support vector machines, in: Proc. International Joint Conference on Neural Networks (IJCNN'99), Washington, DC, 1999.
- [41] F. Youping, C. Yungping, S. Wansheng, L. Yu, Multi-classification algorithm and its realization based on least square support vector machine algorithm, *J. Syst. Eng. Electron.* 16 (4) (2005) 901–907.
- [42] A. Sengpur, Multiclass least-square support vector machines for analog modulation classification, *Expert Syst. Appl.* 36 (2009) 6681–6685.
- [43] T.V. Gestel, J.A.K. Suykens, G. Lanckriet, A. Lambrechts, B. De Moor, J. Vandewalle, Multiclass LS-SVMs: moderated outputs and coding-decoding schemes, *Neural Process. Lett.* 15 (2002) 45–58.
- [44] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, *Mach. Learn.* 29 (1997) 131–163.
- [45] H. Zhang, The optimality of naive bayes, *The International Flairs Conference* 17 (2004).
- [46] R.M. Hope, Z. Wang, Z. Wang, Q. Ji, W.D. Gray, Workload classification across subjects using EEG, in: Proceedings of the human factors and ergonomics society 55th annual meeting -, 2011, pp. 202–206.
- [47] T. Cover, P. Hart, Nearest neighbor pattern classification, *IEEE Trans. Inf. Theor.* IT-13 (1) (1967) 21–27.
- [48] J. Han, M. Kamper, J. Pei, Data mining: Concepts and Techniques, Morgan Kaufmann, San Francisco, California, USA, 2005.
- [49] B. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, Cambridge, UK, 1996.
- [50] A.H. Fielding, Cluster and Classification Techniques for the Biosciences, Cambridge University Press, Cambridge, UK, 2007.
- [51] S. Siuly, Y. Li, Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain computer interface, *IEEE Trans. Neural Syst. Rehabil. Eng.* 20 (4) (2012) 526–538.
- [52] S. Siuly, Y. Li, P. Wen, Clustering technique-based least square support vector machine for EEG signal classification, *Comput. Methods Programs Biomed.* 104 (3) (2011) 358–372.
- [53] S. Siuly, Y. Li, P. Wen, Identification of motor imagery tasks through CC-LR algorithm in brain computer interface, *Int. J. Bioinf. Res. Appl.* 9 (2) (2013) 156–172.
- [54] S. Siuly, Y. Li, P. Wen, Analysis and classification of EEG signals using a hybrid clustering technique, in: The Proceedings of the 2010 IEEE/ICME International Conference on Complex Medical Engineering (CME2010), 2010, pp. 34–39.
- [55] S. Siuly, Y. Li, P. Wen, Classification of EEG signals using sampling techniques and least square support vector machines, in: The Proceedings of Fourth International Conference on Rough Sets and Knowledge Technology (RSKT 2009), LNCS 5589, 2009, pp. 375–382.
- [56] V. Bajaj, R.B. Pachori, Classification of seizure and nonseizure EEG signals using empirical mode decomposition, *IEEE Trans. Inf. Technol. Biomed.* 16 (6) (2012) 1135–1142.
- [57] V. Bajaj, R.B. Pachori, Epileptic seizure detection based on the instantaneous area of analytic intrinsic mode functions of EEG signals, *Open Biomed. Eng. Lett., Springer* 3 (1) (2013) 17–21.
- [58] LS-SVMlab Toolbox (version 1.8), (<http://www.esat.kuleuven.ac.be/sista/lssvmlab/>)