Expert Systems with Applications 37 (2010) 8659–8666

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

EEG signal classification using PCA, ICA, LDA and support vector machines

Abdulhamit Subasi [a,*], M. Ismail Gursoy [b]

[a] International Burch University, Faculty of Engineering and Information Technologies, Sarajevo, Bosnia and Herzegovina
[b] Kahta Vocational School of Higher Education, Adiyaman University, Adiyaman, Turkey

ARTICLE INFO

Keywords:
Electroencephalogram (EEG)
Epileptic seizure
Discrete wavelet transform (DWT)
Independent component analysis (ICA)
Principal component analysis (PCA)
Linear discriminant analysis (LDA)
Support vector machines (SVM)

ABSTRACT

In this work, we proposed a versatile signal processing and analysis framework for Electroencephalogram (EEG). Within this framework the signals were decomposed into the frequency sub-bands using DWT and a set of statistical features was extracted from the sub-bands to represent the distribution of wavelet coefficients. Principal components analysis (PCA), independent components analysis (ICA) and linear discriminant analysis (LDA) is used to reduce the dimension of data. Then these features were used as an input to a support vector machine (SVM) with two discrete outputs: epileptic seizure or not. The performance of classification process due to different methods is presented and compared to show the excellent of classification process. These findings are presented as an example of a method for training, and testing a seizure prediction method on data from individual petit mal epileptic patients. Given the heterogeneity of epilepsy, it is likely that methods of this type will be required to configure intelligent devices for treating epilepsy to each individual's neurophysiology prior to clinical operation.

1. Introduction

Electroencephalograms (EEGs) are recordings of the electrical potentials produced by the brain. Analysis of EEG activity has been achieved principally in clinical settings to identify pathologies and epilepsies since Hans Berger's recording of rhythmic electrical activity from the human scalp. In the past, interpretation of the EEG was limited to visual inspection by a neurophysiologist, an individual trained to qualitatively make a distinction between normal EEG activity and abnormalities contained within EEG records. The advance in computers and the technologies related to them has made it potential to successfully apply a host of methods to quantify EEG changes (Bronzino, 2000).

Compared with other biomedical signals, the EEG is extremely difficult for an untrained observer to understand, partially as a consequence of the spatial mapping of functions onto different regions of the brain and electrode placement. Besides, data processing can be determination of reduced features set including only data needed for quantification, as in evoked response recordings, or feature extraction and subsequent pattern recognition, as in automated spike detection during monitoring for epileptic seizure activity. In early attempts to show a relationship between the EEG and behavior, analog frequency analyzers were used to examine the EEG data. This approach is based on earlier interpretation that the EEG spectrum contains some characteristic waveforms that fall primarily within four frequency bands – δ (<4 Hz), θ (4–8 Hz), α (8–13 Hz), and β (13–30 Hz). Even though unsatisfactory, these initial efforts did bring in the use of frequency analysis to the study of brain wave activity. Although power spectral analysis provides a quantitative measure of the frequency distribution of the EEG at the expense of other details in the EEG such as the amplitude distribution and information relating to the presence of particular EEG patterns. Hence time–frequency signal-processing algorithms such as discrete wavelet transform (DWT) analysis are necessary to address different behavior of the EEG in order to describe it in the time and frequency domain. It should also be emphasized that the DWT is suitable for analysis of non-stationary signals, and this represents a major advantage over spectral analysis. Hence the DWT is well suited to locating transient events. Such transient events as spikes can occur during epileptic seizures (Adeli, Zhou, & Dadmehr, 2003; Bronzino, 2000; D'Alessandro et al., 2003; Subasi, 2007).

An exciting application of seizure prediction technology is its potential for use in therapeutic epilepsy devices to trigger intervention to prevent seizures before they begin. Seizure prediction has been investigated by type to include prediction by studying preictal features, prediction by fast detection, prediction by classification, and prediction by probability estimation. Studies in seizure prediction vary widely in their theoretical approaches to the problem, validation of results, and the amount of data analyzed. Some relative weaknesses in this literature are the lack of extensive testing on baseline data free from seizures, the lack of technically rigorous validation and quantification of algorithm

* Corresponding author. Address: International Burch University, Faculty of Engineering and Information Technologies, Francuske Revolucije bb. Ilidza, Sarajevo, 71210, Bosnia and Herzegovina. Tel.: +387 33 782 100; fax: +387 33 782 131.
E-mail address: asubasi@ibu.edu.ba (A. Subasi).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.06.065

performance in many studies (Adeli et al., 2003; D'Alessandro et al., 2003; Subasi, 2006, 2007).

Principal component analysis (PCA), independent component analysis (ICA) and linear discriminant analysis (LDA) are well-known methods for feature extraction (Cao, Chua, Chong, Lee, & Gu, 2003; Wang & Paliwal, 2003; Widodo & Yang, 2007). Feature extraction means transforming the existing features into a lower-dimensional space which is useful for feature reduction to avoid the redundancy due to high-dimensional data. In this work, DWT has been applied for the time–frequency analysis of EEG signals for the classification using wavelet coefficients. EEG signals were decomposed into frequency sub-bands using DWT. Then a set of statistical features was extracted from these sub-bands to represent the distribution of wavelet coefficients. PCA, ICA and LDA are used to reduce the dimension of data. Then these features were used as an input to a support vector machine (SVM) with two discrete outputs: epileptic or non-epileptic seizure. The accuracy of the various classifiers will be assessed and cross-compared, and advantages and limitations of each technique will be discussed. The simulation shows that SVM by feature extraction using PCA, ICA or LDA can always perform better than that without feature extraction. Furthermore, among the three methods, the best performance is achieved in LDA feature extraction.

2. Materials and methods

2.1. Subjects and data recording

We used the publicly available data described in Andrzejak et al. (2001). The complete data set[1] consists of five sets (denoted A–E) each containing 100 single-channel EEG segments. Sets A and B consisted of segments taken from surface EEG recordings that were carried out on five healthy volunteers using a standardized electrode placement scheme. Volunteers were relaxed in an awake state with eyes open (A) and eyes closed (B), respectively. Sets C, D, and E originated from EEG archive of presurgical diagnosis. EEGs from five patients were selected, all of whom had achieved complete seizure control after resection of one of the hippocampal formations, which was therefore correctly diagnosed to be the epileptogenic zone. Segments in set D were recorded from within the epileptogenic zone, and those in set C from the hippocampal formation of the opposite hemisphere of the brain. While sets C and D contained only activity measured during seizure free intervals, set E only contained seizure activity. All EEG signals were recorded with the same 128-channel amplifier system, using an average common reference. The data were digitized at 173.61 samples per second using 12 bit resolution. Band-pass filter settings were 0.53–40 Hz (12 dB/oct). In this study, we used two dataset (A and E) of the complete dataset as in Subasi (2007). Typical EEGs are given in Fig. 1.

2.2. Analysis using discrete wavelet transform

A signal is said to be stationary if it does not change much over time. Fourier transform can be applied to the stationary signals. However, like EEG, plenty of signals may contain non-stationary or transitory characteristics. Thus it is not ideal to directly apply Fourier transform to such signals. In such a situation time–frequency methods such as wavelet transform must be used. In wavelet analysis, a variety of different probing functions may be used. This concept leads to the defining equation for the continuous wavelet transform (CWT):

$$W(a,b) = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)dt \tag{1}$$

where b acts to translate the function across $x(t)$, and the variable a acts to vary the time scale of the probing function, ψ. If a is greater than one, the wavelet function, ψ, is stretched along the time axis, and if it is less than one (but still positive) it contacts the function. While the probing function ψ could be any of a number of different functions, it always takes on an oscillatory form, hence the term "wavelet." The * indicates the operation of complex conjugation, and the normalizing factor $\frac{1}{\sqrt{|a|}}$ ensures that the energy is the same for all values of a. In applications that require bilateral transformations, it would be preferred a transform that produces the minimum number of coefficients required to recover accurately the original signal. The *discrete wavelet transform* (DWT) achieves this parsimony by restricting the variation in translation and scale, usually to powers of 2. For most signal and image processing applications, DWT-based analysis is best described in terms of filter banks. The use of a group of filters to divide up a signal into various spectral components is termed *sub-band coding*. This procedure is known as multi-resolution decomposition of a signal $x[n]$. Each stage of this scheme consists of two digital filters and two down-samplers by 2. The first filter, $h[\cdot]$ is the discrete mother wavelet, high-pass in nature, and the second, $g[\cdot]$ is its mirror version, low-pass in nature. The down-sampled outputs of first high-pass and low-pass filters provide the detail, D1 and the approximation, A1, respectively (Adeli et al., 2003; Marchant, 2003; Semmlow, 2004).

Selection of appropriate wavelet and the number of levels of decomposition is very important in analysis of signals using DWT. The number of levels of decomposition is chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the signal that correlate well with the frequencies required for classification of the signal are retained in the wavelet coefficients. Since the EEG signals do not have any useful frequency components above 30 Hz, the number of levels was chosen to be 5. Thus the signal is decomposed into the details D1–D5 and one final approximation, A5. The ranges of various frequency bands are shown in Table 1. Figs. 2 and 3 show five different levels of approximation and details of an EEG signal taken from an unhealthy and healthy subject, respectively. These approximation and detail records are reconstructed from the Daubechies 4 (DB4) wavelet filter (Adeli et al., 2003).

The extracted wavelet coefficients provide a compact representation that shows the energy distribution of the EEG signal in time and frequency. Table 1 presents frequencies corresponding to different levels of decomposition for Daubechies order 4 wavelet with a sampling frequency of 173.6 Hz. In order to further decrease the dimensionality of the extracted feature vectors, statistics over the set of the wavelet coefficients was used (Kandaswamy, Kumar, Ramanathan, Jayaraman, & Malmurugan, 2004). The following statistical features were used to represent the time–frequency distribution of the EEG signals:

(1) Mean of the absolute values of the coefficients in each sub-band.
(2) Average power of the wavelet coefficients in each sub-band.
(3) Standard deviation of the coefficients in each sub-band.
(4) Ratio of the absolute mean values of adjacent sub-bands.

Features 1 and 2 represent the frequency distribution of the signal and the features 3 and 4 the amount of changes in frequency distribution. These feature vectors, calculated for the frequency bands A5 and D3–D5, were used for classification of the EEG signals (Kandaswamy et al., 2004).

[1] EEG time series are available under (http://www.meb.unibonn.de/epileptologie/science/physik/eegdata.html).

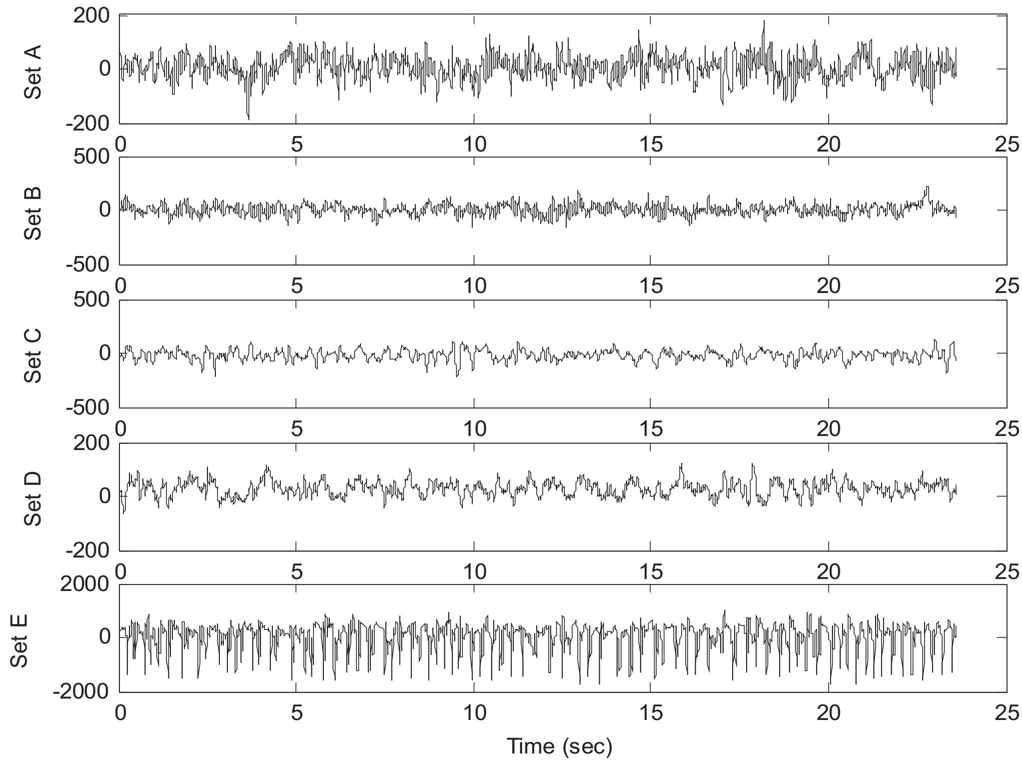

Fig. 1. Examples of five different sets of EEG signals taken from different subjects.

Table 1
Frequencies corresponding to different levels of decomposition for Daubechies 4 filter wavelet with a sampling frequency of 173.6 Hz.

Decomposed signal	Frequency range (Hz)
D_1	43.4–86.8
D_2	21.7–43.4
D_3	10.8–21.7
D_4	5.4–10.8
D_5	2.7–5.4
A_5	0–2.7

2.3. Feature extraction methods

2.3.1. Principal component analysis (PCA)

Principal component analysis (PCA) is a well-established method for feature extraction and dimensionality reduction. In PCA, we seek to represent the d-dimensional data in a lower-dimensional space. This will reduce the degrees of freedom; reduce the space and time complexities. The objective is to represent data in a space that best expresses the variation in a sum-squared error sense. This technique is mostly useful for segmenting signals from multiple sources. It facilitates significantly if we know how many independent components exist ahead of time, as with standard clustering methods. The basic approach in principal components is theoretically rather simple. First, the d-dimensional mean vector μ and $d \times d$ covariance matrix Σ are computed for the full data set. Next, the eigenvectors and eigenvalues are computed, and sorted according to decreasing eigenvalue. Call these eigenvectors e_1 with eigenvalue λ_1, e_2 with eigenvalue λ_2, and so on. Sub-sequently, the largest k such eigenvectors are chosen. In practice, this is done by looking at a spectrum of eigenvectors. Often there will be dimension implying an inherent dimensionality of the subspace governing the "signal." The other dimensions are noise. Form a $k \times k$ matrix A whose columns consist of the k eigenvectors. Preprocess data according to:

$$x' = A^t(x - \mu) \tag{2}$$

It can be shown that this representation minimizes a squared error criterion. Details are given in Cao et al. (2003), Duda, Hart, and Strok (2001).

2.3.2. Independent component analysis (ICA)

ICA is a feature extraction method that transform multivariate random signal into a signal having components that are mutually independent. Independent components can be extracted from the mixed signals by using this method. In this manner, independence denotes the information carried by one component cannot be inferred from the others. Statistically this means that joint probability of independent quantities is obtained as the product of the probability of each of them. Suppose there are c independent scalar source signals $x_i(t)$ for $i = 1, \ldots, c$ where we can consider t to be a time index $1 \leqslant t \leqslant T$. For notational convenience we group the c values at an instant into a vector $x(t)$ and assume, further, that the vector has zero mean. Because of our independence assumption, and an assumption of no noise, we the multivariate density function can be written as

$$p(x(t)) = \prod_{i=1}^{c} p(x_i(t)) \tag{3}$$

Suppose that a d-dimensional data vector is observed at each moment,

$$y(t) = Ax(t) \tag{4}$$

where A is a $c \times d$ scalar matrix, and below we shall require $d \geqslant c$. The task of independent component analysis is to recover the

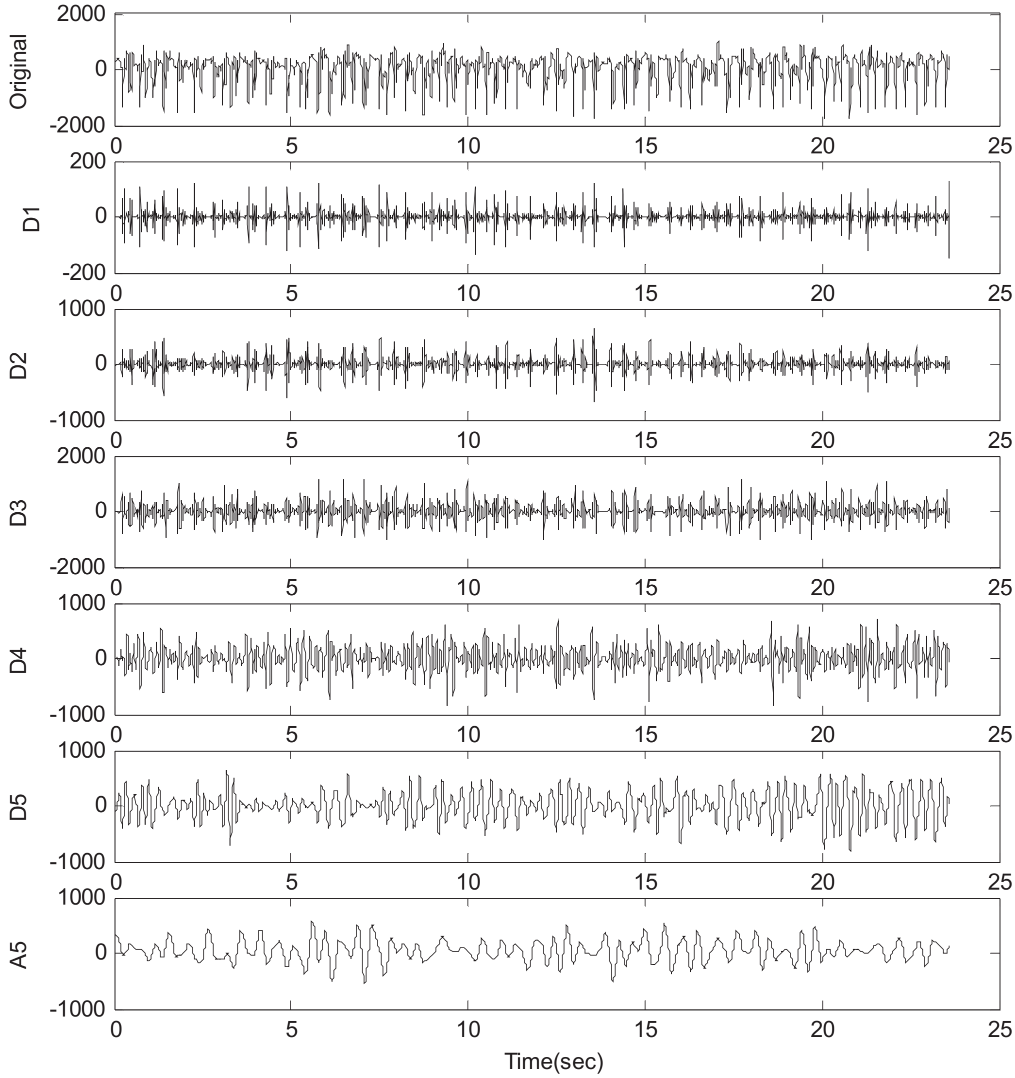

Fig. 2. Approximate and detailed coefficients of EEG signal taken from unhealthy subject (epileptic patient).

source signals from the sensed signals. More specifically, we seek a real matrix W such that

$$z(t) = Wy(t) = WAx(t) \tag{5}$$

where z is an estimate of the sources $x(t)$. Of course we seek $W = A^{-1}$, but neither A nor its inverse are known.

We approach the determination of A by maximum-likelihood techniques. We use an estimate of the density, parameterized by a $\hat{p}(y;a)$ and seek the parameter vector a that minimizes the difference between the source distribution and the estimate. That is, a is the basis vectors of A and thus $\hat{p}(y;a)$ is an estimate of the $p(y)$. Details are given in Cao et al. (2003), Duda et al. (2001), and Widodo and Yang (2007).

2.3.3. Linear discriminant analysis (LDA)

The aim of LDA is to create a new variable that is a combination of the original predictors. This is accomplished by maximizing the differences between the predefined groups, with respect to the new variable. The goal is to combine the predictor scores in such a way that, a single new composite variable, the discriminant score, is formed. This can be viewed as an excessive data dimension reduction technique that compresses the p-dimensional predictors into a one-dimensional line. At the end of the process it is hoped that each class will have a normal distribution of discriminant scores but with the largest possible difference in mean scores for the classes. In reality, the degree of overlap between the discriminant score distributions can be used as a measure of the success of the technique. Discriminant scores are calculated by a discriminant function which has the form:

$$D = w_1Z_1 + w_2Z_2 + w_3Z_3 + \cdots + w_pZ_p \tag{6}$$

As a result a discriminant score is a weighted linear combination of the predictors. The weights are estimated to maximize the differences between class mean discriminant scores. Generally, those predictors which have large dissimilarities between class means will have larger weights, at the same time weights will be small when class means are similar (Fielding, 2007).

2.4. Support vector machines (SVMs)

Support vector machines (SVMs) are build on developments in computational learning theory. Because of their accuracy and ability to deal with a large number of predictors, they have more attention in biomedical applications. The majority of the previous classifiers separate classes using hyperplanes that split the classes, using a flat plane, within the predictor space. SVMs broaden the

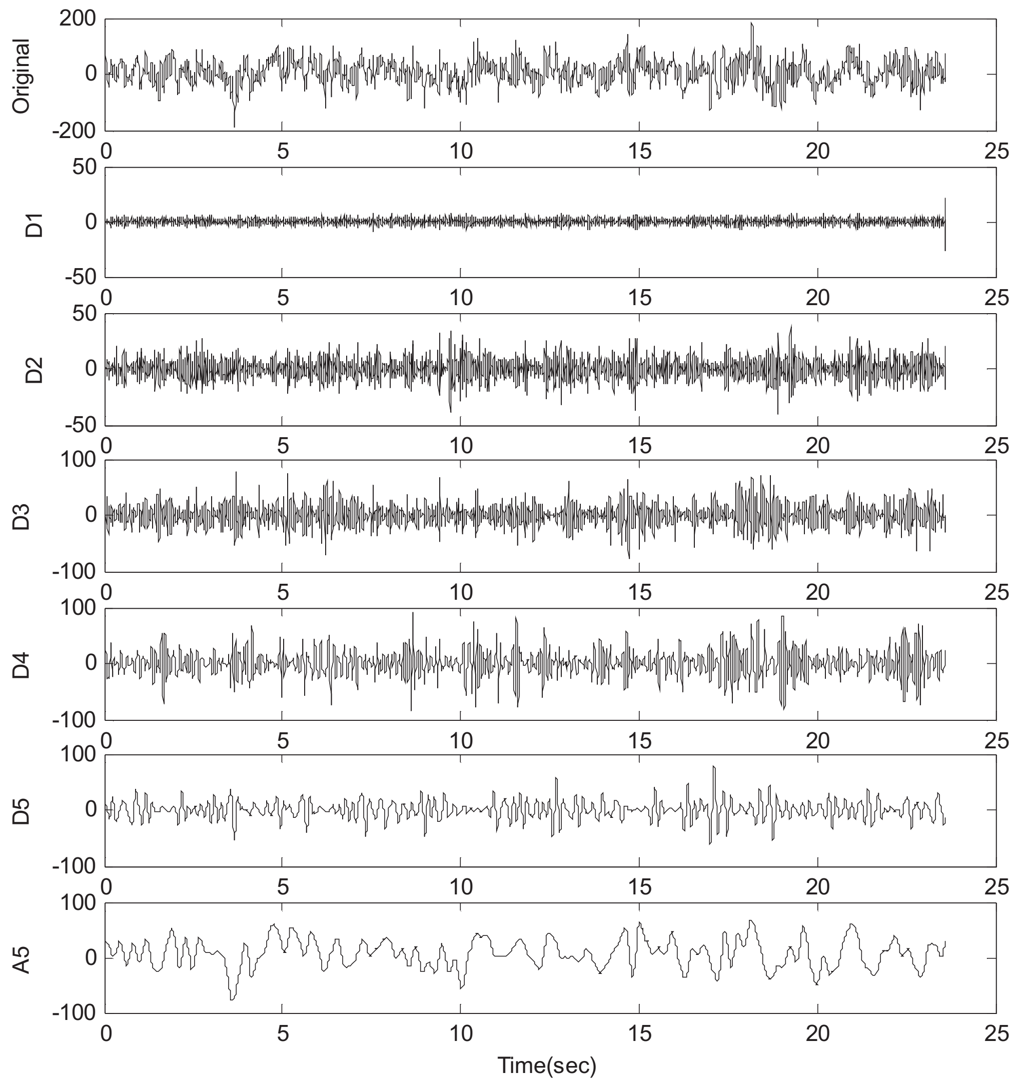

Fig. 3. Approximate and detailed coefficients of EEG signal taken from a healthy subject.

concept of hyperplane separation to data that cannot be separated linearly, by mapping the predictors onto a new, higher-dimensional space in which they can be separated linearly.

The method's name derives from the support vectors, which are lists of the predictor values taken from cases that lie closest to the decision boundary separating the classes. It is practical to assume that these cases have the greatest impact on the location of the decision boundary. In fact, if they were removed they could have large effects on its location. Computationally, finding the best location for the decision plane is an optimization problem that makes uses of a kernel function to build linear boundaries through nonlinear transformations, or mappings, of the predictors. The intelligent component of the algorithm is that it locates a hyperplane in the predictor space which is stated in terms of the input vectors and dot products in the feature space. The dot product can then be used to find the distances between the vectors in this higher-dimensional space. A SVM locates the hyperplane that divides the support vectors without ever representing the space explicitly. As an alternative a kernel function is used that plays the role of the dot product in the feature space. The two classes can only be separated absolutely by a complex curve in the original space of the predictor. The best linear separator cannot totally separate the two classes. On the other hand, if the original predictor values can be projected into a more suitable feature space, it is possible to separate completely the classes with a linear decision boundary. As a result, the problem becomes one of finding the suitable transformation. The kernel function, which is central to the SVM approach, is also one of the main problems, especially with respect to the selection of its parameter values. It is also crucial to select the magnitude of the penalty for violating the soft margin between the classes. This means that successful construction of a SVM necessitates some decisions that should be informed by the data to be classified (Abe, 2005; Burbidge, Trotter, Buxton, & Holden, 1998; Burges, 1998; Duda et al., 2001; Fielding, 2007).

The basic support vector classifier is very similar to the perceptron. Both are linear classifiers, assuming separable data. In perceptron learning, the iterative procedure is stopped when all samples in the training set are classified correctly. For linearly separable data, this means that the found perceptron is one solution arbitrarily selected from an (in principle) infinite set of solutions. In contrast, the support vector classifier chooses one particular solution: the classifier which separates the classes with maximal margin. The margin is defined as the width of the largest 'tube' not containing samples that can be drawn around the decision boundary (see Fig. 4). It can be proven that this particular solution has the highest generalization ability.

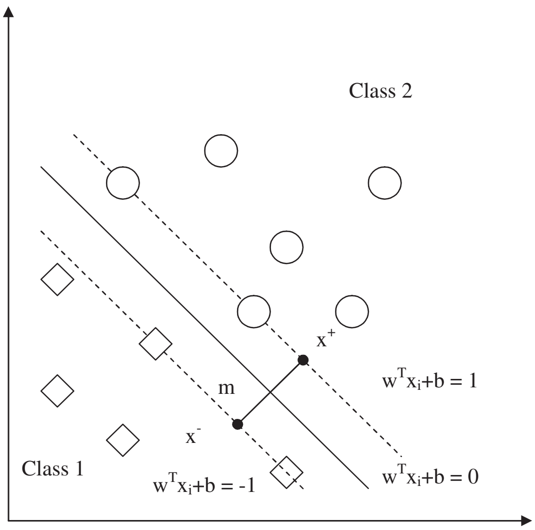

Fig. 4. The linear support vector classifier.

The support vector classifier has many advantages. A unique global optimum for its parameters can be found using standard optimization software. Nonlinear boundaries can be used without much extra computational effort. Moreover, its performance is very competitive with other methods. A drawback is that the problem complexity is not of the order of the dimension of the samples, but of the order of the number of samples. For large sample sizes $N_S > 1000$ general quadratic programming software will often fail and special-purpose optimizers using problem-specific speedups have to be used to solve the optimization. Details are given (Abe, 2005; Burbidge et al., 1998; Burges, 1998; Cortes & Vapnik, 1995; Duda et al., 2001; Fielding, 2007; Van der Heijden, Duin, de Ridder, & Tax, 2004; Vapnik, 1995).

3. Results and discussion

In this study, we used EEG signals of normal and epileptic patients in order to perform a comparison between the PCA, ICA and LDA by using SVM. EEG recordings were divided into sub-band frequencies such as α, β, δ and θ by using DWT. Then a set of statistical features was extracted from the wavelet sub-band frequencies δ (1–4 Hz), θ (4–8 Hz), α (8–13 Hz) and β (13–30 Hz). After normalization, the EEG signals were decomposed using wavelet transform and the statistical features were extracted from the sub-bands. Then dimension of these features are reduced by using ICA, PCA and LDA. A classification system based on SVM was implemented using these data as inputs.

The objective of the modelling phase in this application was to develop classifiers that are able to identify any input combination as belonging to either one of the two classes: normal or epileptic. For developing neural network classifiers, 800 examples were randomly taken from the 1600 examples and used for training the neural networks, and the remaining 800 examples were kept aside and used for testing the developed models. The class distribution of the samples in the training and test data set is summarized in Table 2.

Additionally, because the problem involves classification into two classes, sensitivity and specificity were used as a performance measure. In order to analyze the output data obtained from the application, sensitivity (true positive ratio) and specificity (true negative ratio) are calculated by using confusion matrix. The sensitivity value (true positive, same positive result as the diagnosis of expert neurologists) was calculated by dividing the total of diagnosis numbers to total diagnosis numbers that are stated by the expert neurologists. Sensitivity, also called the true positive ratio, is calculated by the formula:

Table 2
Class distribution of the samples in the training and test data sets.

Class	Training set	Test set	Total
Epileptic	400	400	800
Normal	400	400	800
Total	800	800	1600

$$\text{Sensitivity} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \tag{7}$$

On the other hand, specificity value (true negative, same diagnosis as the expert neurologists) is calculated by dividing the total of diagnosis numbers to total diagnosis numbers that are stated by the expert neurologists. Specificity, also called the true negative ratio, is calculated by the formula:

$$\text{Specificity} = \text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\% \tag{8}$$

3.1. Experimental results

Epileptic seizure detection in EEG can be thought as a sort of pattern recognition concept. It consists of data acquisition, signal processing, feature extraction, feature reduction and seizure detection. A novel EEG signal classification method is proposed, which is based on DWT, the dimension reduction (based on ICA, PCA and LDA) and SVM classification. The procedure of the proposed system can be summarized as follows:

Step 1: The features calculated with statistical features parameter from time–frequency domain using DWT.
Step 2: We extract the features using ICA, PCA and LDA algorithm to reduce the dimensionality. This step is performed to remove the irrelevant features which are redundant and even degrade the performance of the classifier.
Step 3: The classification process for epileptic seizure detection is carried out using SVM-based classification.

The procedure was repeated on EEG recordings of all subjects (healthy and epileptic patients). In this work, the radial basis function (RBF) kernel is used as the kernel function of SVMs. There are two parameters related with this kernel: σ and γ. The upper bound σ for penalty term and kernel parameter γ plays a critical role in performance of SVMs. Hence, inappropriate selection of parameters σ and γ, may cause over-fitting or under-fitting problem. Therefore, we should find optimal σ and γ so that the classifier can accurately classify the data input. In this work, we use 10-fold cross-validation to investigate the appropriate kernel parameter σ, and γ. Principally, all the pairs of (σ, γ) for RBF kernel are tried and the one with the best cross-validation accuracy is selected. After the selection of optimal kernel parameters σ, and γ, the whole training data was trained once more to construct the final classifier.

In this work, the training process carried out using RBF kernel to PCA + SVM, ICA + SVM, and LDA + SVM. After training, we used three different feature extraction methods and get the test results which are shown in Table 3. By using PCA, ICA and LDA features are extracted from original feature sets. In addition, the number of support vectors (SVs) decreased due to feature extraction. As seen in Table 3, the classification rate with LDA feature extraction is highest (100%) and ICA came second (99.5%). The PCA had lowest correct classification percentage (98.75%) compared to LDA and

Table 3
The values of statistical parameters of the ICA, PCA and LDA models for EEG signal classification.

Feature extraction method	Accuracy	Specificity	Sensitivity
PCA (%)	98.75	98.5	99.00
ICA (%)	99.5	99	100
LDA (%)	100	100	100

ICA counterparts. Also the simulation shows that SVM by feature extraction using PCA, ICA, or LDA can always perform better than that without feature extraction (98%). The excellent of LDA is also shown by the number of SVs which is reduced and smaller than PCA and ICA. In these circumstances, classification process using ICA feature extraction needs fewer numbers of SVs than PCA feature extraction. This fact can be explained that ICA finds the components not only uncorrelated but independent. Independent components are more valuable for classification rather than uncorrelated components. However, according to training time, the classification process using LDA feature extraction and SVMs is relatively longer than PCA and ICA feature extraction. Furthermore, it is obvious that kernel parameter selection is crucial to get good performance. Besides, the use of appropriate kernel parameter will overcome the problems of under-fitting and over-fitting so the best classification process is yielded.

3.2. Discussion

Although the previous works have shown good performance on the EEG signal classification, there still remain some problems to be solved. First, the number of available EEG patterns for the classifier training is not much more, which shows us that the generalization ability of a classifier dominates the accuracy of online EEG classification. On the other hand, the classifiers used in the previous works, for instance, the ANNs did not minimize the generalization error bound for unseen EEG patterns. In this work, SVM is implemented to overcome this limitation. Second, the systems in previous works sent all the extracted features into the classifiers directly. But, due to a great deviation in EEG pattern distribution there exist mixed distribution between classes in general. As a result, if a feature transformation mechanism that can minimize the within-class scatter and maximize the between-class scatter is set into the system, it can be anticipated that the size of between-class overlap region can be significantly reduced and the classification performance can be significantly improved. In order to achieve this, the PCA, ICA, and LDA algorithms are used in proposed structure.

Based on the results of the present study and experience in the EEG signal classification problem, we would like to emphasize the following:

1. The high classification accuracy of the SVM classifier gives insights into the features used for defining the EEG signals. The conclusion drawn in the applications demonstrated that the DWT coefficients are the features, which well represent the EEG signals, and by the usage of these features a good distinction between classes can be obtained.
2. Support vector machines (SVMs) are based on preprocessing the data to represent patterns in a high dimension—typically much higher than the original feature space. With an appropriate nonlinear mapping to a sufficiently high dimension, data from two categories can always be separated by a hyperplane. As a result, while the original features bring sufficient information for good classification, mapping to a higher-dimensional feature space make available better discriminatory evidence that are absent in the original feature space. The problem of training an SVM is to select the nonlinear functions that map the input to a higher-dimensional space. Often this choice will be informed by the designer's knowledge of the problem domain. In the absence of such information, we might choose to use polynomials, Gaussians or other basis functions. The dimensionality of the mapped space can be arbitrarily high (though in practice it may be limited by computational resources). For training the SVMs we chose Radial Basis Function (RBF) and tried to find an appropriate kernel parameters σ, and γ. The optimal σ, and γ values can only be ascertained after trying out different values. In addition, the choice of γ parameter in the SVM is crucial in order to have a suitably trained SVM. The SVM has to be trained for different kernel parameters until to get the best result (Cortes & Vapnik, 1995; Ubeyli, 2008; Vapnik, 1995).
3. Subasi (2007) evaluated the diagnostic accuracy of the Mixture of Expert (ME) model and ANN on the same EEG data sets (A and E) (Andrzejak et al., 2001) and the total classification accuracy of the ME model was 94.5% and ANN was 93.2%. Thus, the accuracy rates of the SVM with the ICA, PCA and LDA for this application were found to be significantly higher than that of the ANN and ME model presented in the previous study (Subasi, 2007).
4. Nigam and Graupe (2004) used the same EEG data sets (A and E) by using different feature extraction with ANN and the total classification accuracy of their model was 97.2%. The SVM used for this application indicated higher performance than that of the ANN model presented by Nigam and Graupe (2004) also.
5. The classification results and the values of statistical parameters indicated that the SVM with the ICA, PCA and LDA had considerable success in the EEG signals classification by comparing with the ANN. The proposed combined PCA, ICA and LDA methods with SVM approach can be evaluated in classification of the non-stationary biomedical signals.
6. The testing performance of the SVM-based diagnostic system is found to be satisfactory and we think that this system can be used in clinical studies after it is developed. This application brings objectivity to the evaluation of EEG signals and its automated nature makes it easy to be used in clinical practice. Besides the feasibility of a real-time implementation of the expert diagnosis system, diagnosis may be made more accurately by increasing the variety and the number of parameters.

4. Conclusion

Diagnosing epilepsy is a difficult task requiring observation of the patient, an EEG, and gathering of additional clinical information. SVMs that classifies subjects as having or not having an epileptic seizure provides a valuable diagnostic decision support tool for physicians treating potential epilepsy, since differing etiologies of seizures result in different treatments. Conventional classification methods of EEG signals using mutually exclusive time and frequency domain representations does not give efficient results. In this work, EEG signals were decomposed into time–frequency representations using DWT and statistical features were calculated to represent their distribution. Using statistical features extracted from the DWT sub-bands of EEG signals, three feature extraction method; namely PCA, ICA, and LDA, were used with SVM and cross-compared in terms of their accuracy relative to the observed epileptic/normal patterns. The comparisons were based on two scalar performance measures derived from the confusion matrices; namely specificity and sensitivity. The result of EEG signal classification using SVMs shows that nonlinear feature extraction can improve the performance of classifier with respect to reduce the number of support vector. According to this result, the application of nonlinear feature extraction and SVMs can serve as a promising alternative for intelligent diagnosis system in the future. Also it is demonstrated that dimension reduction by PCA, ICA and LDA can improve the generalization performance of SVM.

References

Abe, S. (2005). *Support vector machines for pattern classification*. London: Springer.
Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods, 123*, 69–87.
Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E, 64*, 061907.
Bronzino, J. D. (2000). Principles of electroencephalography (2nd ed.). In J. D. Bronzino (Ed.). *The biomedical engineering handbook*. Boca Raton: CRC Press LLC.
Burbidge, R., Trotter, M., Buxton, B., & Holden, S. (1998). Drug design by machine learning: Support vector machines for pharmaceutical data analysis. *Computers and Chemistry, 26*, 5–14.
Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2*(2), 1–47.
Cao, L. J., Chua, K. S., Chong, W. K., Lee, H. P., & Gu, Q. M. (2003). A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing, 55*, 321–336.
Cortes, C., & Vapnik, V. (1995). Support vector networks. *Machine Learning, 20*(3), 273–297.
D'Alessandro, M., Esteller, R., Vachtsevanos, G., Hinson, A., Echauz, A., & Litt, B. (2003). Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: A report of four patients. *IEEE Transactions on Biomedical Engineering, 50*(5), 603–615.
Duda, R. O., Hart, P. E., & Strok, D. G. (2001). *Pattern classification* (2nd ed.). John Wiley & Sons.
Fielding, A. H. (2007). *Cluster and classification techniques for the biosciences*. Cambridge, UK: Cambridge University Pres.
Kandaswamy, A., Kumar, C. S., Ramanathan, R. P., Jayaraman, S., & Malmurugan, N. (2004). Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine, 34*(6), 523–537.
Marchant, B. P. (2003). Time–frequency analysis for biosystem engineering. *Biosystems Engineering, 85*(3), 261–281.
Nigam, V. P., & Graupe, D. (2004). A neural-network-based detection of epilepsy. *Neurological Research, 26*(1), 55–60.
Semmlow, J. L. (2004). *Biosignal and biomedical image processing: MATLAB-based applications*. New York: Marcel Dekker, Inc..
Subasi, A. (2006). Automatic detection of epileptic seizure using dynamic fuzzy neural networks. *Expert Systems with Applications, 31*, 320–328.
Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications, 32*, 1084–1093.
Ubeyli, E. D. (2008). Analysis of EEG signals by combining eigenvector methods and multiclass support vector machines. *Computers in Biology and Medicine, 38*, 14–22.
Van der Heijden, F., Duin, R. P. W., de Ridder, D., & Tax, D. M. J. (2004). *Classification parameter estimation and state estimation: An engineering approach using MATLAB*. England: John Wiley & Sons Ltd..
Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer.
Wang, X., & Paliwal, K. K. (2003). Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern Recognition, 36*, 2429–2439.
Widodo, A., & Yang, B. (2007). Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors. *Expert Systems with Applications, 33*, 241–250.