
Click to activate...
LANZAROTE
Caliente

Computer Methods and Programs in Biomedicine (2005) **78**, 87—99

ELSEVIER

Computer Methods and Programs in Biomedicine

www.intl.elsevierhealth.com/journals/cmpb

Classification of EEG signals using neural network and logistic regression

Abdulhamit Subasi [a,*], **Ergun Erçelebi** [b]

[a] *Department of Electrical and Electronics Engineering, Kahramanmaras Sutcu Imam University, 46601 Kahramanmaraş, Turkey*
[b] *Department of Electrical and Electronics Engineering, University of Gaziantep, 27310 Gaziantep, Turkey*

Received 26 May 2004; received in revised form 12 October 2004; accepted 26 October 2004

KEYWORDS
EEG;
Epileptic seizure;
Lifting-based discrete wavelet transform (LBDWT);
Logistic regression (LR);
Multilayer perceptron neural network (MLPNN)

Summary Epileptic seizures are manifestations of epilepsy. Careful analyses of the electroencephalograph (EEG) records can provide valuable insight and improved understanding of the mechanisms causing epileptic disorders. The detection of epileptiform discharges in the EEG is an important component in the diagnosis of epilepsy. As EEG signals are non-stationary, the conventional method of frequency analysis is not highly successful in diagnostic classification. This paper deals with a novel method of analysis of EEG signals using wavelet transform and classification using artificial neural network (ANN) and logistic regression (LR). Wavelet transform is particularly effective for representing various aspects of non-stationary signals such as trends, discontinuities and repeated patterns where other signal processing approaches fail or are not as effective. Through wavelet decomposition of the EEG records, transient features are accurately captured and localized in both time and frequency context. In epileptic seizure classification we used lifting-based discrete wavelet transform (LBDWT) as a preprocessing method to increase the computational speed. The proposed algorithm reduces the computational load of those algorithms that were based on classical wavelet transform (CWT). In this study, we introduce two fundamentally different approaches for designing classification models (classifiers) the traditional statistical method based on logistic regression and the emerging computationally powerful techniques based on ANN. Logistic regression as well as multilayer perceptron neural network (MLPNN) based classifiers were developed and compared in relation to their accuracy in classification of EEG signals. In these methods we used LBDWT coefficients of EEG signals as an input to classification system with two discrete outputs: epileptic seizure or non-epileptic seizure. By identifying features in the signal we want to provide an automatic system that will support a physician in the diagnosing process. By applying LBDWT in connection with MLPNN, we obtained novel and reliable classifier architecture. The comparisons between the developed classifiers were primarily based on analysis of the receiver operating characteristic (ROC) curves as well as a number of scalar performance

* Corresponding author.
E-mail addresses: asubasi@ksu.edu.tr (A. Subasi), ercelebi@gantep.edu.tr (E. Erçelebi).

doi:10.1016/j.cmpb.2004.10.009

measures pertaining to the classification. The MLPNN based classifier outperformed the LR based counterpart. Within the same group, the MLPNN based classifier was more accurate than the LR based classifier.

1. Introduction

The human brain is obviously a complex system and exhibits rich spatiotemporal dynamics. Among the noninvasive techniques for probing human brain dynamics, electroencephalography (EEG) provides a direct measure of cortical activity with millisecond temporal resolution. EEG is a record of the electrical potentials generated by the cerebral cortex nerve cells. There are two different types of EEG depending on where the signal is taken in the head: scalp or intracranial. For scalp EEG, the focus of this research, small metal discs, also known as electrodes, are placed on the scalp with good mechanical and electrical contact. Intracranial EEG is obtained by special electrodes implanted in the brain during a surgery. In order to provide an accurate detection of the voltage of the brain neuron current, the electrodes are of low impedance (<5 kΩ). The changes in the voltage difference between electrodes are sensed and amplified before being transmitted to a computer program to display the tracing of the voltage potential recordings. The recorded EEG provides a continuous graphic exhibition of the spatial distribution of the changing voltage fields over time.

Epileptic seizure is an abnormality in EEG recordings and is characterized by brief and episodic neuronal synchronous discharges with dramatically increased amplitude. This anomalous synchrony may occur in the brain locally (partial seizures), which is seen only in a few channels of the EEG signal, or involving the whole brain (generalized seizures), which is seen in every channel of the EEG signal.

EEG signals involve a great deal of information about the function of the brain. But classification and evaluation of these signals are limited. Since there is no definite criterion evaluated by the experts, visual analysis of EEG signals in time domain may be insufficient. Routine clinical diagnosis needs to analysis of EEG signals. Therefore, some automation and computer techniques have been used for this aim. Since the early days of automatic EEG processing, representations based on a Fourier transform have been most commonly applied. This approach is based on earlier observations that the EEG spectrum contains some characteristic waveforms that fall primarily within four frequency bands—delta (<4 Hz), theta (4—8 Hz), alpha (8—13 Hz) and beta (13—30 Hz). Such methods have proved beneficial for various EEG characterizations, but fast Fourier transform (FFT), suffer from large noise sensitivity. Parametric power spectrum estimation methods such as autoregressive (AR), reduces the spectral loss problems and gives better frequency resolution. But, since the EEG signals are non-stationary, the parametric methods are not suitable for frequency decomposition of these signals [1,2].

A powerful method was proposed in the late 1980s to perform time-scale analysis of signals: the wavelet transforms (WT). This method provides a unified framework for different techniques that have been developed for various applications [2—18]. Since the WT is appropriate for analysis of non-stationary signals and this represents a major advantage over spectral analysis, it is well suited to locating transient events, which may occur during epileptic seizures.

Wavelet's feature extraction and representation properties can be used to analyze various transient events in biological signals. Adeli et al. [2] gave an overview of the discrete wavelet transform (DWT) developed for recognizing and quantifying spikes, sharp waves and spike-waves. They used wavelet transform to analyze and characterize epileptiform discharges in the form of 3-Hz spike and wave complex in patients with absence seizure. Through wavelet decomposition of the EEG records, transient features are accurately captured and localized in both time and frequency context. The capability of this mathematical microscope to analyze different scales of neural rhythms is shown to be a powerful tool for investigating small-scale oscillations of the brain signals. A better understanding of the dynamics of the human brain through EEG analysis can be obtained through further analysis of such EEG records.

Numerous other techniques from the theory of signal analysis have been used to obtain representations and extract the features of interest for classification purposes. Neural networks and statistical pattern recognition methods have been applied to EEG analysis. Neural network detection systems have been proposed by a number of researchers [19—29]. Pradhan et al. [19] used the raw EEG as an input to a neural network while Weng and Khorasani [20] used the features proposed by Gotman [21] with an adaptive structure neural network, but his results show a poor false detection rate. Petrosian et al. [22] showed that the ability of specifically designed and trained recurrent neural networks (RNN), combined with wavelet preprocessing, to predict the onset of epileptic seizures both on scalp and intracranial recordings only one-channel of electroencephalogram.

In order to provide faster and efficient algorithm, Folkers et al. [11] proposed a versatile signal processing and analysis framework for bioelectrical data and in particular for neural recordings and 128-channel EEG. Within this framework the signal is decomposed into subbands using fast wavelet transform algorithms, executed in real-time on a current digital signal processor hardware platform.

This paper aims to compare the traditional method of logistic regression to the more advanced neural network techniques, as mathematical tools for developing classifiers for the detection of epileptic seizure in multi-channel EEG. In the neural network techniques, the multilayer perceptron neural network (MLPNN) will be used with backpropagation and Levenberg—Marquardt training algorithm. The choice of this network was based on the fact that it is the most popular type of artificial neural networks (ANNs). In these methods we used lifting-based discrete wavelet transform (LBDWT) coefficients of EEG signals as an input to classification system with two discrete outputs: epileptic seizure or non-epileptic seizure. We provide faster wavelet decomposition in multi-channel EEG without any special hardware, by using LBDWT in a multi-channel EEG. The accuracy of the classifiers will be assessed and cross-compared, and advantages and limitations of each technique will be discussed.

2. Materials and method

2.1. Subjects and data recording

The EEG data used in our study were downloaded from 24-h EEG recorded from both epileptic patients and normal subjects. The following bipolar EEG channels were selected for analysis: F7-C3, F8-C4, T5-O1 and T6-O2. In order to assess the performance of the classifier, we selected 500 EEG segments containing spike and wave complex, artifacts and background normal EEG. Twenty absence seizures (petit mal) from five epileptic patients admitted for video-EEG monitoring were analyzed. The total recording time was 452.8 h with an average duration of 22.8 ± 2.4 h. The subjects consisted of three males and two females, age 28.87 ± 15.27 (mean $\pm$ SD; range 6—43) with a diagnosis of epilepsy and no other accompanying disorders. Recordings were done under video control to have an accurate determination of the different stage of the seizure. The different stages of EEG signals were determined by two physicians. EEG data were acquired with Ag/AgCl disc electrodes placed using the 10—20 international electrode placement system. The recordings band-pass filtered (1—70 Hz) EEG. The filtered EEG signals were segmented to 5-s (1000 sample) durations. Four-channel recordings containing epileptiform events (spikes, spike and waves) were digitized at 200 samples per second using 12-bit resolution. All EEG were taken during restful wakefulness stage but some portions of the EEG contained EMG artifacts. Digitized data were stored on an optical disc for further processing.

2.2. Visual inspection and validation

Two neurologists with experience in the clinical analysis of EEG signals separately inspected every recording included in this study to score epileptic and normal signals. Each event was filed on the computer memory and linked to the tracing with its start and duration. These were then revised by the two experts jointly to solve disagreements and set up the training set for the program, consenting to the choice of threshold for the epileptic seizure detection. The agreement between the two experts was evaluated—for the testing set—as the rate between the numbers of epileptic seizures detected by both experts. A further step was then performed with the aim of checking the disagreements and setting up a "gold standard" reference set. When revising this unified event set, the human experts, by mutual consent, marked each state as epileptic or normal. They also reviewed each recording entirely for epileptic seizures that had been overlooked by all during the first pass and marked them as definite or possible. This validated set provided the reference evaluation to estimate the sensitivity and specificity of computer scorings. Nevertheless, a preliminary analysis was carried out solely on events in the training set, as each stage in these sets had a definite start and duration.

2.3. Wavelet transform analysis

The discrete wavelet transform is a versatile signal-processing tool that finds many engineering and scientific applications. One area in which the DWT has been particularly successful is the epileptic seizure

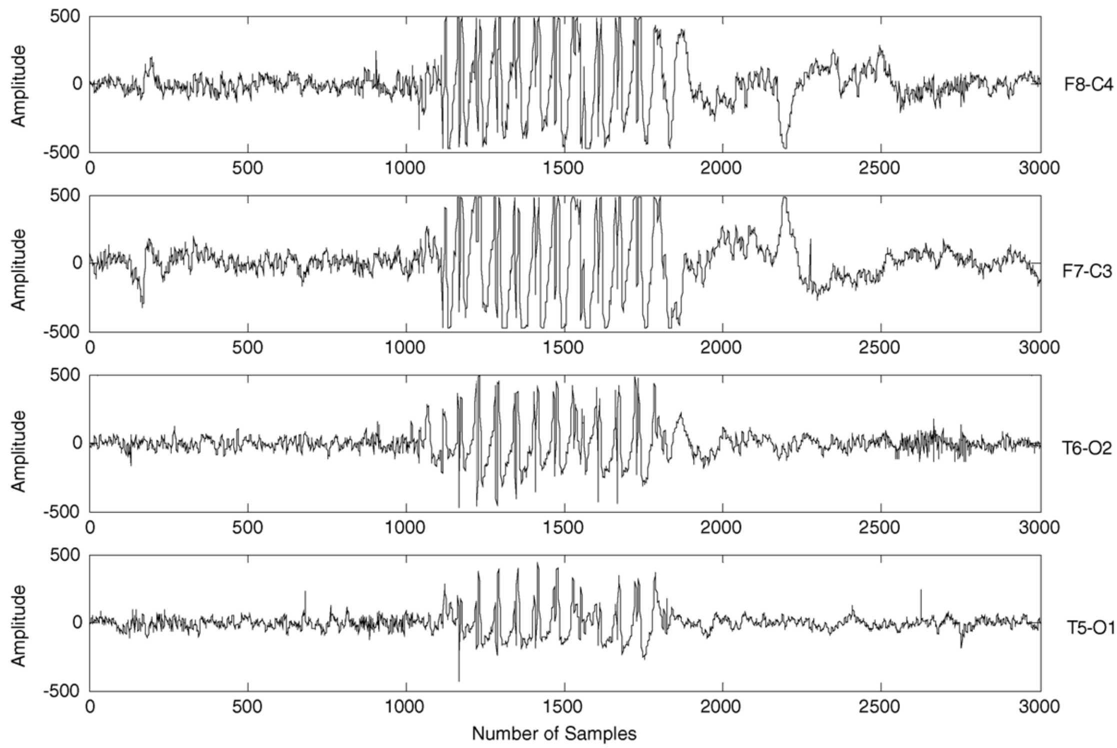

Fig. 1 Epileptic EEG signal.

detection because it captures transient features and localizes them in both time and frequency content accurately. However, the conventional convolution based implementation of the DWT has high computational and memory requirements. Recently, lifting based implementation of the DWT has been proposed to overcome these drawbacks [30,31]. The lifting scheme is a new method for constructing biorthogonal wavelets. The basic idea behind the lifting scheme is a relationship among all biorthogonal wavelets that share the same scaling function such that one can construct the desired wavelet form a simple one. Any wavelet with FIR filters can be factorized into a finite number of alternating lifting and dual lifting steps starting from the lazy wavelet, by a finite number of lifting or dual lifting. The main difference with such classical constructions is that it entirely relies on the spatial domain. Therefore, it is ideally suited for constructing wavelets that lack translation and dilation, and thus the Fourier transform is no longer available. This scheme is called second-generation wavelets. Obviously, it can be used to construct first-generation wavelets and leads to a faster, fully in-place implementation of the wavelet transform. The lifting based wavelet transform implementation not only helps in reducing the number of computations but also achieves lossy to lossless performance with finite precision. The computational efficiency of the lifting implementation can be up to 100% higher than the traditional direct convolution based implementation [30,31]. Detailed derivations related to LBDWT are given in Appendix A.

The proposed method was applied on a wide variety of EEG data for both epileptic and normal signals. Four channels of EEG (F7-C3, F8-C4, T5-O1 and T6-O2) recorded from a patient with absence seizure epileptic discharges are shown in Fig. 1 and normal EEG signal shown in Fig. 2. Fig. 3 shows six different levels of approximation (identified by A1—A5 and displayed in the left column) and details (identified by D1—D5 and displayed in the right column) of an epileptic EEG signal. Fig. 4 shows six different levels of approximation (identified by A1—A5 and displayed in the left column) and details (identified by D1—D5 and displayed in the right column) of a normal EEG signal. These approximation and detail records are reconstructed from the DB4 wavelet filter. Approximation A4 is obtained by superimposing details D5 on approximation A5. Approximation A3 is obtained by superimposing details D4 on approximation A4 and so on. Finally, the original signal is obtained by superimposing details D1 on approximation A1. LBDWT acts like a mathematical microscope, zooming into small scales to reveal compactly spaced events in time and zooming out into large scales to exhibit the global waveform patterns.

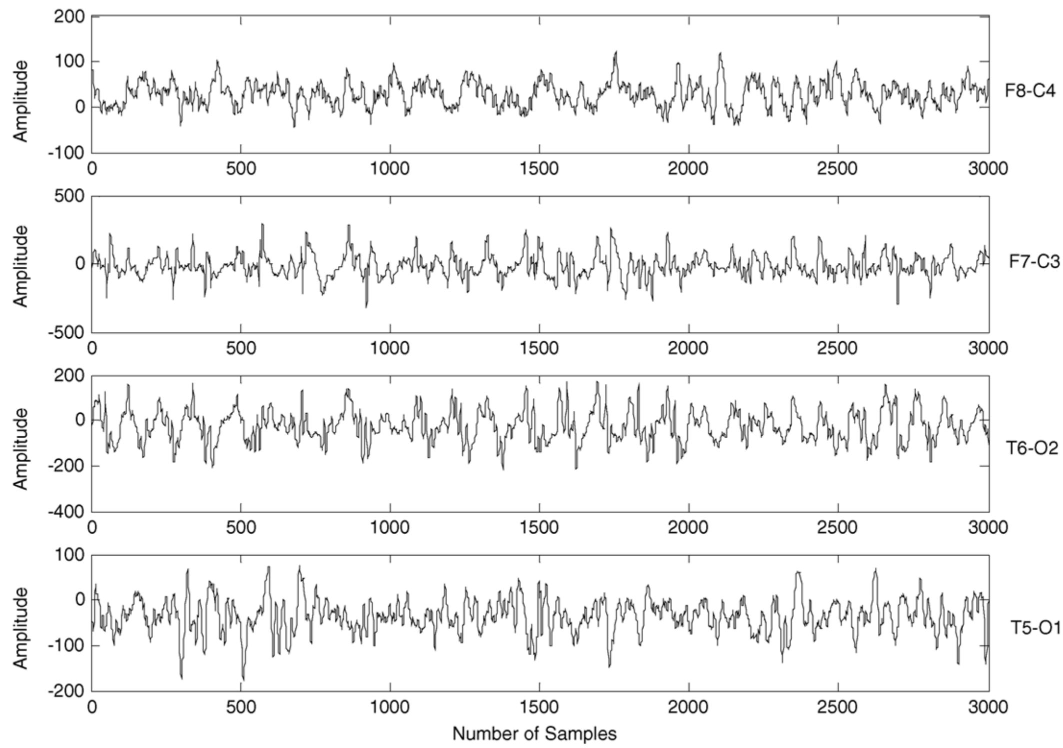

Fig. 2 Normal EEG signal.

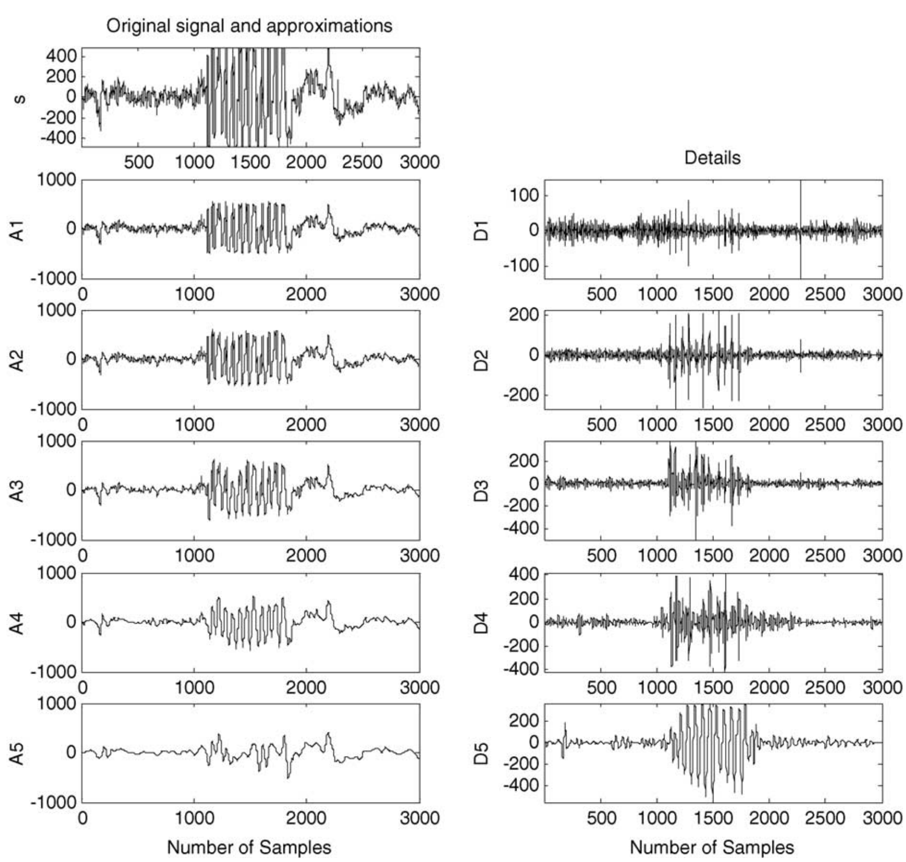

Fig. 3 Approximate and detailed coefficients of epileptic EEG signal.

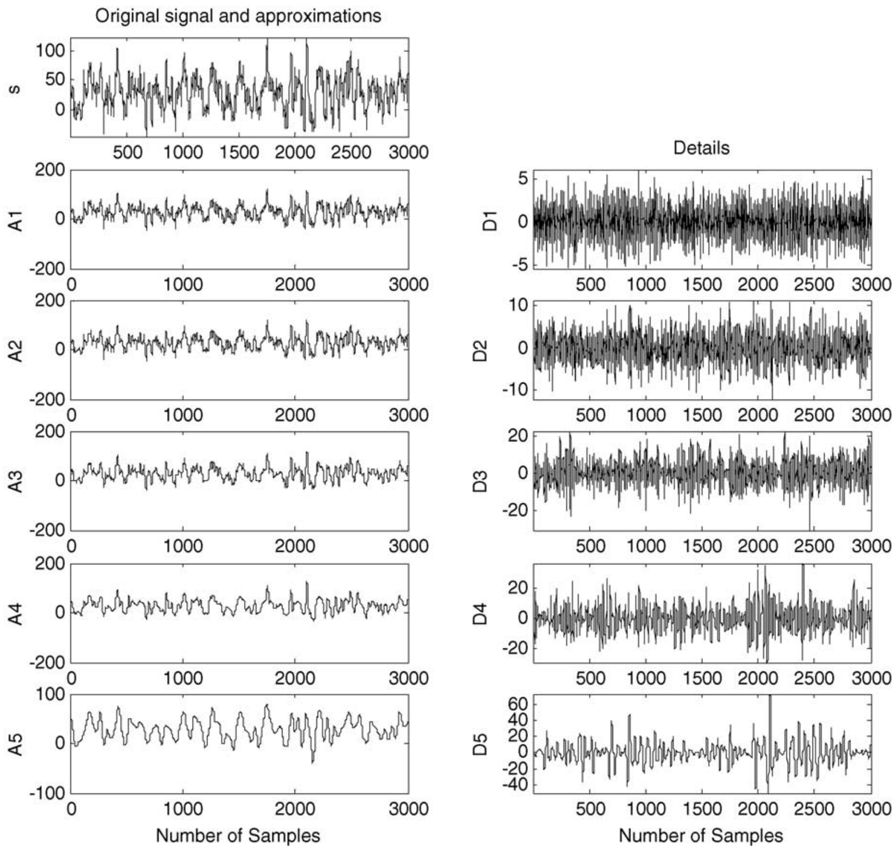

Fig. 4 Approximate and detailed coefficients of normal EEG signal.

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 four wavelet with a sampling frequency of 200 Hz. It can be seen from Table 1 that the components A5 decomposition is within the delta range (1—4 Hz), D5 decomposition is within the theta range (4—8 Hz), D4 decomposition is within the alpha range (8—13 Hz) and D3 decomposition is within the beta range (13—30 Hz). Lower level decompositions corresponding to higher frequencies have negligible magnitudes in a normal EEG.

Table 1 Frequencies corresponding to different levels of decomposition for Daubechies four filter wavelet with a sampling frequency of 200 Hz

Decomposed signal	Frequency range (Hz)
D1	50—100
D2	25—50
D3	12.5—25
D4	6.25—12.5
D5	3.125—6.25
A5	0—3.125

2.4. Logistic regression

Logistic regression [32—35] is a widely used statistical modeling technique in which the probability, P_1, of dichotomous outcome event is related to a set of explanatory variables in the form

$$\mathrm{logit}(P_1) = \ln\left(\frac{P_1}{1-P_1}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \quad (1)$$

In Eq. (1), β_0 is the intercept and $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients associated with the explanatory variable $x_1, x_2, \ldots, x_n$. These input variables are the average of the wavelet coefficients (D3—D5 and A5) of four-channel EEG signals. A dichotomous variable is restricted to two values such as yes/no, on/off, survive/die or 1/0, usually representing the occurrence or non-occurrence of some event (for example, epileptic seizure/not). The explanatory (independent) variables may be continuous, dichotomous, discrete or combination. The use of ordinary linear regression (OLR) based on least squares method with dichotomous outcome would lead to meaningless results. As in Eq. (1), the response (dependent) variable is the natural logarithm of the odds ratio representing the ratio between the probability that an event will occur to the probability that it will not occur (e.g., probability of being epileptic or not). In general, logistic regression imposes less stringent requirements than OLR, in that it does not assume linearity of the relationship between the explanatory variables and the response variable and does not require Gaussian distributed independent variables. Logistic regression calculates the changes in the logarithm of odds of the response variable, rather than the changes in the response variable itself, as OLR does. Because the logarithm of odds is linearly related to the explanatory variables, the regressed relationship between the response and explanatory variables is not linear. The probability of occurrence of an event as function of the explanatory variables is nonlinear as derived from Eq. (1) as

$$P_1(x) = \frac{1}{1+e^{-\mathrm{logit}\,(P_1(x))}} = \frac{1}{1+e^{-(\beta_0+\sum_{i=1}^{n}\beta_i x_i)}} \quad (2)$$

Unlike OLR, logistic regression will force the probability values (P_1) to lie between 0 and 1 ($P_1 \to 0$ as the right-hand side of Eq. (2) approaches $-\infty$, and $P_1 \to 1$ as it approaches $+\infty$). Commonly, the maximum likelihood estimation (MLE) method is used to estimate the coefficients β_0, β_1, ..., β_n in the logistic regression equation [32—35]. This method is different from that based on ordinary least squares (OLS) for estimating the coefficients in linear regression. The OLS method seeks to minimize the sum of squared distances of all the data points from the regression line. On the other hand, the MLE method seeks to maximize the log likelihood, which reflects how likely it is (the odds) that the observed values of the dependent variable may be predicted from the observed values of the independent variables. Unlike OLS method, the MLE method is an iterative algorithm, which starts with an initial arbitrary estimate of the regression equation coefficients and proceeds to determine the direction and magnitude of change in the coefficients that will increase the likelihood function. After this initial function is determined, residuals are tested and a new estimate is computed with an improved function. This process is repeated until some convergence criterion (e.g., Wald test, log likelihood-ratio test, classification tables, etc.) is reached. In the current study, the coefficients were obtained by minimizing (using Newton's method) the log likelihood function defined as the sum of the logarithms of the predicted probabilities of occurrence for those cases where the event occurred and the logarithms of the predicted probabilities of non-occurrence for those cases where the event did not occur [35,36].

2.5. Artificial neural networks

Artificial neural networks are computing systems made up of large number of simple, highly interconnected processing elements (called nodes or artificial neurons) that abstractly emulate the structure and operation of the biological nervous system. Learning in ANNs is accomplished through special training algorithms developed based on learning rules presumed to mimic the learning mechanisms of biological systems. There are many different types and architectures of neural networks varying fundamentally in the way they learn, the details of which are well documented in the literature [36—40]. In this paper, neural network relevant to the application being considered (i.e., classification of EEG data) will be employed for designing classifiers, namely the MLPNN.

The architecture of MLPNN may contain two or more layers. A simple two-layer ANN consists only of an input layer containing the input variables to the problem and output layer containing the solution of the problem. This type of networks is a satisfactory approximator for linear problems. However, for approximating nonlinear systems, additional intermediate (hidden) processing layers are employed to handle the problem's nonlinearity and complexity. Although it depends on complexity of the function or the process being modeled, one hidden layer may be sufficient to map an arbitrary function to any degree of accuracy. Hence, three-layer architecture ANNs were adopted for the present study. Fig. 5 shows the typical

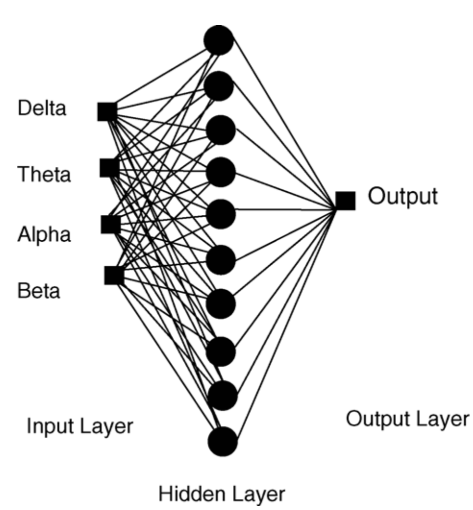

Fig. 5 Artificial neural network architecture.

structure of a fully connected three-layer network.

The determination of appropriate number of hidden layers is one of the most critical tasks in neural network design. Unlike the input and output layers, one starts with no prior knowledge as to the number of hidden layers. A network with too few hidden nodes would be incapable of differentiating between complex patterns leading to only a linear estimate of the actual trend. In contrast, if the network has too many hidden nodes it will follow the noise in the data due to over-parameterization leading to poor generalization for untrained data. With increasing number of hidden layers, training becomes excessively time-consuming. The most popular approach to finding the optimal number of hidden layers is by trial and error [36—40]. In the present study, MLPNN consisted of one input layer, one hidden layer with 21 nodes and one output layer.

Training algorithms are an integral part of ANN model development. An appropriate topology may still fail to give a better model, unless trained by a suitable training algorithm. A good training algorithm will shorten the training time, while achieving a better accuracy. Therefore, training process is an important characteristic of the ANNs, whereby representative examples of the knowledge are iteratively presented to the network, so that it can integrate this knowledge within its structure. There are a number of training algorithms used to train a MLPNN and a frequently used one is called the backpropagation training algorithm [36—40]. The backpropagation algorithm, which is based on searching an error surface using gradient descent for points with minimum error, is relatively easy to implement. However, backpropagation has some problems for many applications. The algorithm is not guaranteed to find the global minimum of the error function since gradient descent may get stuck in local minima, where it may remain indefinitely. In addition to this, long training sessions are often required in order to find an acceptable weight solution because of the well-known difficulties inherent in gradient descent optimization. Therefore, a lot of variations to improve the convergence of the backpropagation were proposed. Optimization methods such as second-order methods (conjugate gradient, quasi-Newton, Levenberg—Marquardt (L—M)) have also been used for ANN training in recent years. The Levenberg—Marquardt algorithm combines the best features of the Gauss—Newton technique and the steepest-descent algorithm, but avoids many of their limitations. In particular, it generally does not suffer from the problem of slow convergence [41].

Table 2 Class distributions of the samples in the training and the validation data sets

Class	Training set	Validation set	Total
Epileptic	102	88	190
Normal	198	112	310
Total	300	200	500

2.6. Development of logistic regression model and ANNs

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 the logistic regression and neural network classifiers, 300 examples were randomly taken from the 500 examples and used for deriving the logistic regression models or for training the neural networks. The remaining 200 examples were kept aside and used for testing the validity of the developed models. The class distribution of the samples in the training, validation and test data set is summarized in Table 2.

We divided four-channel EEG recordings into subbands frequencies by using LBDWT as in Figs. 3 and 4. Since four-frequency band, which are alpha (D4), beta (D3), theta (D5) and delta (A5) is sufficient for the EEG signal processing, these wavelet subband frequencies (delta (1—4 Hz), theta (4—8 Hz), alpha (8—13 Hz), beta (13—30 Hz)) are applied to LR and MLPNN input (as in Fig. 5). Then we take the average of the four channels and give these wavelet coefficients (D3—D5 and A5) of EEG signals as an input to ANN and LR.

The MLPNN was designed with LBDWT coefficients (D3—D5 and A5) of EEG signal in the input layer; and the output layer consisted of one node representing whether epileptic seizure detected or not. A value of "0" was used when the experimental investigation indicated a normal EEG pattern and "1" for epileptic seizure. The preliminary architecture of the network was examined using one and two hidden layers with a variable number of hidden nodes in each. It was found that one hidden layer is adequate for the problem at hand. Thus, the sought network will contain three layers of nodes. The training procedure started with one hidden node in the hidden layer, followed by training on the training data (300 data sets), and then by testing on the validation data (200 data sets) to examine the network's prediction performance on cases never used in its development. Then, the same procedure was run repeatedly each time the network was expanded by adding one more node to the hidden layer, until the best architecture and set of connection weights were obtained. Using the backpropagation (L—M) algorithm for training, a training rate of 0.01 (0.005) and momentum coefficient of 0.95 (0.9) were found optimum for training the network with various topologies. The selection of the optimal network was based on monitoring the variation of error and some accuracy parameters as the network was expanded in the hidden layer size and for each training cycle. The sum of squares of error representing the sum of square of deviations of ANN solution (output) from the true (target) values for both the training and test sets was used for selecting the optimal network. The optimum number of nodes in hidden layer is found as 21.

Additionally, because the problem involves classification into two classes, accuracy, sensitivity and specifity were used as a performance measure. These parameters were obtained separately for both the training and validation sets each time a new network topology was examined. Computer programs that we have written for the training algorithm based on backpropagation of error and L—M were used to develop the MLPNNs.

2.7. Evaluation of performance

The coherence of the diagnosis of the expert neurologists and diagnosis information was calculated at the output of the classifier. Prediction success of the classifier may be evaluated by examining the confusion matrix. 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:

$$\text{sensitivity} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \qquad (3)$$

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:

$$\text{specifity} = \text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\% \qquad (4)$$

Neural network and logistic regression analysis were also compared to each other by receiver operating characteristic (ROC) analysis. ROC analysis is an appropriate means to display sensitivity and specificity relationships when a predictive output for two possibilities is continuous. In its tabular form, the ROC analysis displays true and false positive and negative totals and sensitivity and specificity for each listed cutoff value between 0 and 1.

In order to perform the performance measure of the output classification graphically, the ROC curve was calculated by analyzing the output data obtained from the test. Furthermore, the performance of the model may be measured by calculating the region under the ROC curve. The ROC curve is a plot of the true positive rate (sensitivity) against the false positive rate (1 − specificity) for each possible cutoff. A cutoff value is selected that may classify the degree of epileptic seizure detection correctly by determining the input parameters optimally according to the used model.

3. Results and discussion

Logistic regression model and MLPNN classifier were developed using the 300 training examples, while the remaining 200 examples were used for validation of the model. Note that although logistic regression does not involve training, we will use "training examples" to refer to that portion of database used to derive the regression equations. In order to perform fair comparison between the neural network and logistic regression-based model, only the 300 data sets were used in developing the model and the remaining data sets were kept aside for model validation. The developed logistic model was run on the 300 for training and 200 for validation examples.

Table 3 shows a summary of the performance measures. It is obvious from Table 3 that the MLPNN trained with L—M algorithm is ranked first in terms of its classification accuracy of the EEG signals epileptic/normal data (93%), while the MLPNN trained with backpropogation came second (92%). The logistic regression-based classifier had lower accuracy (89%) compared to the neural network-based counterparts. The MLPNN trained with L—M algorithm was able to accurately predict (detect) epileptic cases, 92.8% of sensitivity compared to 91.6% using the the MLPNN trained with backpropogation, while the logistic regression-based classifiers indicated a detection accuracy of only 89.2%.

Also, the area under ROC curves for the three classifiers (logistic regression, MLPNN trained with

Table 3 Comparison of logistic regression and neural network models for EEG signals

Classifier type	Correctly classified	Specifity	Sensitivity	Area under ROC curve
Logistic regression	89	90.3	89.2	0.853
MLPNN with backprop	92	91.4	91.6	0.889
MLPNN with L—M	93	92.3	92.8	0.902

backpropogation and L—M) is given in Table 3. When the area under the ROC curve in Table 3 is examined, the MLPNN trained with L—M has achieved an acceptable classification success with the value 0.902. However, the area under the curve has been found to be 0.889 in MLPNN trained with backpropogation and 0.853 in the logistic regression analysis. Thus, it can be seen clearly that the performance of the MLPNN trained with L—M is better than MLPNN trained with backpropogation and the logistic regression model.

In this study, EEG recordings were divided into subbands frequencies as alpha, beta, theta and delta by using LBDWT (Figs. 3 and 4). Then, wavelet subband frequencies (delta (1—4 Hz), theta (4—8 Hz), alpha (8—13 Hz), beta (13—30 Hz)) are applied to LR and MLPNN. For solving pattern classification problem MLPNN employing backpropagation and L—M training algorithms were used. Effective training algorithm and better-understood system behavior are the advantages of this type of neural network. Selection of network input parameters and performance of classifier are important in epileptic seizure detection. The efficiency of this technique can be explained by using the result of experiments. This paper clearly demonstrates that our method is applicable for detecting epileptic seizure. The qualities of the method are that it is simple to apply, and it does not require high computation power. The method can be used as a standalone tool, but it can be implemented as a building block of a brain—computer interface for computer-assisted EEG diagnostics.

The classification efficiency, which is defined as the percentage ratio of the number of EEG signals correctly classified to the total number of EEG signals considered for classification, also depends on the type of wavelet chosen for the application. In order to investigate the effect of other wavelets on classifications efficiency, tests were carried out using other wavelets. Apart from db4, Haar, Symmlet of order 10 (sym10), Coiflet of order 4 (coif4), Daubechies of order 2 (db2) and Daubechies of order 8 (db8) were also tried. Average efficiency obtained for each wavelet when EEG signals were classified using various ANN structures. It can be seen that the Daubechies wavelet offers better efficiency than the others and db4 is marginally better than db2 and db8. Hence, db4 wavelet is chosen for this application.

The testing performance of the neural network diagnostic system is found to be satisfactory and we think that this system can be used in clinical studies in the future 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 faster. A "black box" device that may be developed as a result of this study may provide feedback to the neurologists for classification of the EEG signals quickly and accurately by examining the EEG signals with real-time implementation.

4. Summary and conclusions

Diagnosing epilepsy is a difficult task requiring observation of the patient, an EEG, and gathering of additional clinical information. An artificial neural network that classifies subjects as having or not having an epileptic seizure provides a valuable diagnostic decision support tool for neurologists treating potential epilepsy, since differing etiologies of seizures result in different treatments.

In this study, classification of EEG signals was examined. Delta, theta, alpha and beta subfrequencies of the EEG signals were extracted by using LBDWT. The LBDWT coefficients of EEG signals were used as an input to LR and MLPNN that could be used to detect epileptic seizure. This process is realized by online data acquisition system. Depending on these sub-frequencies, classifiers have been developed and trained. We have presented new alternative method based on lifting-based wavelet filters for decomposition of the EEG records of the 3-Hz spike and slow wave epileptic discharges. The capability of this mathematical microscope to analyze different scales of neural rhythms is shown to be a powerful tool for investigating small-scale oscillations of the brain signals. However, to utilize this mathematical microscope effectively, the best suitable wavelet basis function has to be identified for the particular application. Lifting-based wavelets are experimentally found to be very appropriate and faster for wavelet analysis of spike and wave EEG signals. It also needs less computational power than CWT.

In this paper, two approaches to develop classifiers for identifying epileptic seizure were discussed. One approach is based on the traditional method of statistical logistic regression analysis where logistic regression equations were developed. The other approach is based on the neural network technology, mainly using MLPNN trained by the backpropagation and L—M algorithm. Using LBDWT of EEG signals, three classifiers were constructed and cross-compared in terms of their accuracy relative to the observed epileptic/normal patterns. The comparisons were based on analysis of the receiving operator characteristic curves of the three classifiers and two scalar performance measures derived from the confusion matrices; namely specifity and sensitivity. The MLPNN trained with L—M algorithm identified accurately all the epileptic and normal cases. Out of the 100 epileptic/normal cases, the LR-based classifier misclassified a total of 11 cases; MLPNN trained with backpropagation misclassified 8 cases, while the MLPNN trained with L—M misclassified 7 cases.

If we compare our method to Petrosian et al. [22], since they used only one-channel and wavelet decomposed low-pass and high-pass subsignals, their method is not as effective as our method. Because we used four channel of EEG and we divided these signals into five subbands frequencies and used four of these subband frequencies (D3—D5 and A5) as an input to classifier.

Essentially, MLPNNs require deciding on the number of hidden layers, number of nodes in each hidden layer, number of training iteration cycles, choice of activation function, selection of the optimal learning rate and momentum coefficient, as well as other parameters and problems pertaining to convergence of the solution. Compared to logistic regression, MLPNN are easier to build, as for developing logistic regression equations one starts with no knowledge as to the best combination of the parameters or the shape and degree of nonlinearity required to produce an optimal model, with this difficulty increasing by increasing the number of independent parameters. Other advantages of MLPNNs over logistic regression include their robustness to noisy data (with outliers), which can severely hamper many types of most traditional statistical methods. Finally, the fact that an MLPNN-based classifier can be developed quickly makes such classifiers efficient tools that can be easily re-trained, as additional data become available, when implemented in the hardware of EEG signal processing systems.

With specificity and sensitivity values both above 90%, the MLPNN classification may be used as an important diagnostic decision support mechanism to assist physicians in the treatment of epileptic patients.

Appendix A. Lifting-based wavelet transform

Lifting provides a framework that allows the construction certain biorthogonal wavelets and can be generalized to the second-generation setting. First generation families can be built with the lifting framework. Wavelet filters can be decomposed into lifting step, which leads to write transform in the polyphase form then lifting can be made using matrices with Laurent polynomial elements. A lifting step, then, becomes supposedly elementary matrix, which is a triangular matrix (lower or triangular) with all diagonal elements unity. In the simplest form of lifting scheme, the lifting scheme corresponds to a factorization of the polyphase matrix for the wavelet filters [17,18,30,31].

The Classical wavelet transform (or subband coding or multi resolution analysis) is performed using a filter bank in Fig. 6a and can be made using FIR filters.

The analyzing filters are shown by $\tilde{h}$ and $\tilde{g}$, i.e., with a tilde, while the synthesizing filters are denoted by a plain h and g. In the first step, the input

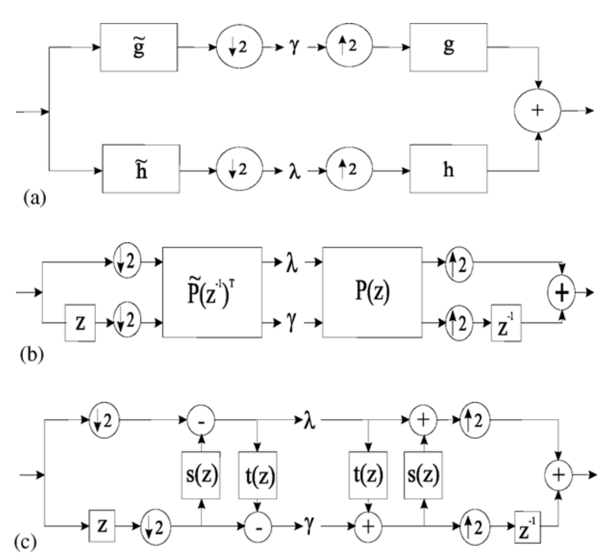

Fig. 6 (a) Two-channel filter bank with analysis filters $\tilde{g}$ and $\tilde{h}$ and synthesis filters g and h; (b) polyphase representation of wavelet transform; (c) left side of the figure is the forward wavelet transform using lifting, right side is the inverse wavelet transform using lifting.

signal is convolved with a high pass filter $\tilde{g}$ and a low pass filter $\tilde{h}$. Since these convolutions yield a result with a size equal to that of the input signal, this convolution process doubles the total number of data. Therefore, sub-sampling follows the low-pass filter $\tilde{h}$ and the high-pass filter $\tilde{g}$. To recover the input signal, inverse transform is performed by first inserting a zero between two elements and then convolution using two synthesis filters h (low-pass) and g (high-pass) [17,18,30,31].

For filter bank in Fig. 6a the conditions for perfect reconstruction are given by

$$\begin{aligned} h(z)\tilde{h}(z^{-1}) + g(z)\tilde{g}(z^{-1}) &= 2 \\ h(z)\tilde{h}(-z^{-1}) + g(z)\tilde{g}(-z^{-1}) &= 0 \end{aligned} \tag{A.1}$$

the polyphase matrix can be defined as

$$\tilde{P}(z) = \begin{bmatrix} \tilde{h}_e(z) & h_0(z) \\ \tilde{g}_e(z) & g_0(z) \end{bmatrix} \tag{A.2}$$

At this phase the wavelet transform is performed by the polyphase matrix. If $\tilde{h}_e(z)$ and $g_0(z)$ are set to unity and both $h_0(z)$ and $\tilde{g}_e(z)$ are zero, $\tilde{P}(z)$ becomes a unity matrix, then the wavelet transform is referred to as the Lazy wavelet transform. The Lazy wavelet transform does nothing but splits the input signal into even and odd components. $P(z)$ is defined in the similar way. The wavelet transform now is represented schematically in Fig. 6b.

As can be seen in this figure the condition for perfect reconstruction is now given by

$$\begin{aligned} P(z)\tilde{P}(z^{-1})^{\mathrm{T}} &= I \\ P(z)^{-1} &= \tilde{P}(z^{-1})^{\mathrm{T}} \end{aligned} \tag{A.3}$$

$$\tilde{P}(z^{-1})^{\mathrm{T}} = P(z)^{-1} = \frac{1}{(h_e(z)g_0(z) - h_0(z)g_e(z))} \times \begin{bmatrix} g_0(z) & -g_e(z) \\ -h_0(z) & h_e(z) \end{bmatrix} \tag{A.4}$$

It is assumed that the determinant of $P(z) = 1$

$$\begin{aligned} \tilde{h}_e(z) &= g_0(z^{-1}) \\ \tilde{h}_0(z) &= -g_e(z^{-1}) \\ \tilde{g}_e(z) &= -h_0(z^{-1}) \\ \tilde{g}_0(z) &= h_e(z^{-1}) \end{aligned} \tag{A.5}$$

The lifting theorem indicates that any other finite filter g^{new} complementary to h is of the form

$$g^{\mathrm{new}}(z) = g(z) + h(z)s(z^2) \tag{A.6}$$

where $s(z^2)$ is a Laurent polynomial conversely any filter of this form is complementary to h. If $g^{\mathrm{new}}(z)$ is written in polyphase form then the new polyphase matrix reads out as follows:

$$P^{\mathrm{new}}(z) = \begin{bmatrix} h_e(z) & h_e(z)s(z) + g_e(z) \\ h_0(z) & h_0(z)s(z) + g_0(z) \end{bmatrix} = P(z)\begin{bmatrix} 1 & s(z) \\ 0 & 1 \end{bmatrix} \tag{A.7}$$

Similarly, we can use the lifting theorem to create the filter $\tilde{h}^{\mathrm{new}}(z)$ complementary to $\tilde{g}(z)$

$$\tilde{h}^{\mathrm{new}}(z) = \tilde{h}(z) - \tilde{g}(z)\tilde{s}(z^{-2}) \tag{A.8}$$

The dual polyphase matrix is given by

$$\tilde{P}^{\mathrm{new}}(z) = \tilde{P}(z)\begin{bmatrix} 1 & 0 \\ -s(z^{-1}) & 1 \end{bmatrix} \tag{A.9}$$

From all the given equations, how things work in the lifting scheme is clear. A procedure starts with a Lazzy wavelet then both the polyphase matrices are equal to the unit matrix. After applying a primal- and/or a dual lifting step to the Lazzy wavelet we get a new wavelet transform that is a little more sophisticated. In other words, we have lifted the wavelet transform to higher level of sophistication. Many lifting steps can be performed to build highly sophisticated wavelet transforms.

Any two-band FIR filter bank can be factored in a set of lifting steps using Euclidean algorithm. Polyphase matrix is factored in a cascade of triangular submatrices, where each submatrix corresponds to a lifting or a dual lifting step [17,18,30,31].

Polyphase matrix $\tilde{P}(z)$ of filter bank from Fig. 6c is factored in triangular submatrices:

$$\tilde{P}(z) = \prod_{i=1}^{m} \begin{bmatrix} 1 & 0 \\ -s_i(z^{-1}) & 1 \end{bmatrix} \begin{bmatrix} 1 & -t_i(z^{-1}) \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} K_2 & \\ & K_1 \end{bmatrix} \tag{A.10}$$

in a similar way polyphase matrix $P(z)$ is factored into lifting steps

$$P(z) = \prod_{i=1}^{m} \begin{bmatrix} 1 & s_i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t_i(z) & 1 \end{bmatrix} \begin{bmatrix} K_1 & \\ & K_2 \end{bmatrix} \tag{A.11}$$

References

[1] I. Guler, M.K. Kiymik, M. Akin, A. Alkan, AR spectral analysis of EEG signals by using maximum likelihood estimation, Comput. Biol. Med. 31 (2001) 441—450.

[2] H. Adeli, Z. Zhou, N. Dadmehr, Analysis of EEG records in an epileptic patient using wavelet transform, J. Neurosci. Methods 123 (2003) 69—87.

[3] O.A. Rosso, M.T. Martin, A. Plastino, Brain electrical activity analysis using wavelet-based informational tools, Physica A 313 (2002) 587—608.
[4] N. Hazarika, J.Z. Chen, A.C. Tsoi, A. Sergejew, Classification of EEG signals using the wavelet transform, Signal Process. 59 (1) (1997) 61—72.
[5] S.V. Patwardhan, A.P. Dhawan, P.A. Relue, Classification of melanoma using tree structured wavelet transforms, Comput. Methods Programs Biomed. 72 (2003) 223—239.
[6] M.L. Van Quyen, J. Foucher, J.P. Lachaux, E. Rodriguez, A. Lutz, J. Martinerie, F.J. Varela, Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony, J. Neurosci. Methods 111 (2001) 83—98.
[7] S. Soltani, P. Simard, D. Boichu, Estimation of the self-similarity parameter using the wavelet transform, Signal Process. 84 (2004) 117—123.
[8] R.Q. Quiroga, M. Schurmann, Functions and sources of event-related EEG alpha oscillations studied with the wavelet transform, Clin. Neurophysiol. 110 (1999) 643—654.
[9] Z. Zhang, H. Kawabata, Z.Q. Liu, Electroencephalogram analysis using fast wavelet transform, Comput. Biol. Med. 31 (2001) 429—440.
[10] E. Basar, M. Schurmann, T. Demiralp, C. Basar-Eroglu, A. Ademoglu, Event-related oscillations are 'real brain responses'—wavelet analysis and new strategies, Int. J. Psychophysiol. 39 (2001) 91—127.
[11] A. Folkers, F. Mosch, T. Malina, U.G. Hofmann, Realtime bioelectrical data acquisition and processing from 128 channels utilizing the wavelet-transformation, Neurocomputing 52—54 (2003) 247—254.
[12] O.A. Rosso, S. Blanco, A. Rabinowicz, Wavelet analysis of generalized tonic—clonic epileptic seizures, Signal Process. 83 (2003) 1275—1289.
[13] V.J. Samar, A. Bopardikar, R. Rao, K. Swartz, Wavelet analysis of neuroelectric waveforms: a conceptual tutorial, Brain Lang. 66 (1999) 7—60.
[14] Y.U. Khan, J. Gotman, Wavelet based automatic seizure detection in intracerebral electroencephalogram, Clin. Neurophysiol. 114 (2003) 898—908.
[15] R.Q. Quiroga, O.W. Sakowitz, E. Basar, M. Schurmann, Wavelet transform in the analysis of the frequency composition of evoked potentials, Brain Res. Protoc. 8 (2001) 16—24.
[16] A.B. Geva, D.H. Kerem, Forecasting generalized epileptic seizures from the eeg signal by wavelet analysis and dynamic unsupervised fuzzy clustering, IEEE Trans. Biomed. Eng. 45 (10) (1998) 1205—1216.
[17] E. Ercelebi, Electrocardiogram signals de-noising using lifting-based discrete wavelet transform, Comput. Biol. Med. 34 (6) (2004) 479—493.
[18] E. Ercelebi, Second generation wavelet transform-based pitch period estimation and voiced/unvoiced decision for speech signals, Appl. Acoustics 64 (2003) 25—41.
[19] N. Pradhan, P.K. Sadasivan, G.R. Arunodaya, Detection of seizure activity in EEG by an artificial neural network: a preliminary study, Comput. Biomed. Res. 29 (1996) 303—313.
[20] W. Weng, K. Khorasani, An adaptive structure neural network with application to EEG automatic seizure detection, Neural Netw. 9 (1996) 1223—1240.
[21] J. Gotman, Automatic recognition of epileptic seizures in the EEG, Electroencephalogr. Clin. Neurophysiol. 54 (1982) 530—540.
[22] A. Petrosian, D. Prokhorov, R. Homan, R. Dashei, D. Wunsch, Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG, Neurocomputing 30 (2000) 201—218.
[23] A.J. Gabor, R.R. Leach, F.U. Dowla, Automated seizure detection using a self-organizing neural network, Electroencephalogr. Clin. Neurophysiol. 99 (1996) 257—266.
[24] E. Haselsteiner, G. Pfurtscheller, Using time-dependent neural Networks for EEG classification, IEEE Trans. Rehab. Eng. 8 (2000) 457—463.
[25] B.O. Peters, G. Pfurtscheller, H. Flyvbjerg, Automatic differentiation of multichannel EEG signals, IEEE Trans. Biomed. Eng. 48 (2001) 111—116.
[26] H. Qu, J. Gotman, A Patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: possible use as a warning device, IEEE Trans. Biomed. Eng. 44 (1997) 115—122.
[27] C. Robert, J.F. Gaudy, A. Limoge, Electroencephalogram processing using neural Networks, Clin. Neurophysiol. 113 (2002) 694—701.
[28] M. Sun, R.J. Sclabassi, The forward EEG solutions can be computed using artificial neural networks, IEEE Trans. Biomed. Eng. 47 (2000) 1044—1050.
[29] W.R.S. Webber, R.P. Lesser, R.T. Richardson, K. Wilson, An approach to seizure detection using an artificial neural network (ANN), Electroencephalogr. Clin. Neurophysiol. 98 (1996) 250—272.
[30] W. Sweldens, The lifting scheme: a custom-design construction of biorthogonal wavelets, Appl. Comput. Harmon. Anal. 3 (2) (1996) 186—200.
[31] W. Sweldens, The lifting scheme: a construction of second generation wavelets, SIAM J. Math. Anal. 29 (2) (1997) 511—546.
[32] D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, Wiley, New York, 1989.
[33] M. Schumacher, R. Robner, W. Vach, Neural networks and logistic regression: Part I, Comput. Stat. Data Anal. 21 (1996) 661—682.
[34] W. Vach, R. Robner, M. Schumacher, Neural networks and logistic regression: part II, Comput. Stat. Data Anal. 21 (1996) 683—701.
[35] M. Hajmeer, M.I.A. Basheer, Comparison of logistic regression and neural network-based classifiers for bacterial growth, Food Microbiol. 20 (2003) 43—55.
[36] S. Dreiseitl, L. Ohno-Machado, Logistic regression and artificial neural network classification models: a methodology review, J. Biomed. Inform. 35 (2002) 352—359.
[37] I.A. Basheer, M. Hajmeer, Artificial neural networks: fundamentals, computing, design, and application, J. Microbiol. Methods 43 (2000) 3—31.
[38] B.B. Chaudhuri, U. Bhattacharya, Efficient training and improved performance of multilayer perceptron in pattern classification, Neurocomputing 34 (2000) 11—27.
[39] L. Fausett, Fundamentals of Neural Networks Architectures, Algorithms, and Applications, Prentice Hall, Englewood Cliffs, NJ, 1994.
[40] S. Haykin, Neural Networks: A Comprehensive Foundation, Macmillan, New York, 1994.
[41] M.T. Hagan, M.B. Menhaj, Training feedforward networks with the Marquardt algorithm, IEEE Trans. Neural Netw. 5 (6) (1994) 989—993.