ELSEVIER

Journal of Neuroscience Methods 148 (2005) 167–176

JOURNAL OF NEUROSCIENCE METHODS

www.elsevier.com/locate/jneumeth

Automatic seizure detection in EEG using logistic regression and artificial neural network

Ahmet Alkan [a,*], Etem Koklukaya [b], Abdulhamit Subasi [a]

[a] *Department of Electrical and Electronics Engineering, Kahramanmaras Sutcu Imam University, 46050-9 Kahramanmaraş, Turkey*
[b] *Department of Electrical and Electronics Engineering, Sakarya University, 54187 Sakarya, Turkey*

Received 31 January 2005; received in revised form 25 April 2005; accepted 26 April 2005

Abstract

The detection of epileptiform discharges in the EEG is an important component in the diagnosis of epilepsy. In this study, multiple signal classification (MUSIC), autoregressive (AR) and periodogram methods were used to get power spectra in patients with absence seizure. The EEG power spectra were used as an input to a classifier. We introduce two fundamentally different approaches for designing classification models (classifiers); the traditional statistical method based on logistic regression (LR) and the emerging computationally powerful techniques based on artificial neural networks (ANNs). LR as well as multilayer perceptron neural network (MLPNN) based classifiers were developed and compared in relation to their accuracy in classification of EEG signals. 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 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.

Keywords: EEG; Spectral analysis; Epileptic seizure; Logistic regression; Multilayer perceptron neural network

1. Introduction

Electroencephalography is an important clinical tool for the evaluation and treatment of neurophysiologic disorders related to epilepsy. Careful analyses of the EEG records can provide valuable insight and improved understanding of the mechanisms causing epileptic disorders. About 1% of the people in the world suffer from epilepsy and about 30% of epileptics are not helped by medication (Adeli et al., 2003).

Research is needed for better understanding of the mechanisms causing epileptic disorders. The detection of epileptiform discharges occurring in the EEG between seizures is an important component in the diagnosis of epilepsy. EEG signals contain a wide range of frequency components. However, the range of clinical and physiological interests is between 0.5 and 30 Hz. This range is classified approximately in a number of frequency bands as follows: δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz) (Adeli et al., 2003; Gevins and Remond, 1987; Schneble and Matthes, 1996).

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. Periodogram method used for calculating the power density of the frequency components in a signal is based on Fourier conversion. Non-parametric power spectrum estimation methods can be calculated using FFT and understood easily compared with the parametric methods. But these methods are powerful tool for the analysis of stationary signals and need long duration data records for suitable frequency resolution. For shorter signals, non-parametric spectral estimation is less reliable. In non-parametric applications, FFT is applied to windowed data which assumes all data zero except window. Also some spec-

* Corresponding author. Tel.: +90 505 234 80 17; fax: +90 344 219 10 52.
E-mail address: aalkan2004@yahoo.com (A. Alkan).

0165-0270/$ – see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.jneumeth.2005.04.009

tral losses occur because of window and these losses mask weak signals in the data.

Parametric power spectrum estimation methods reduce the spectral loss problems and gives better frequency resolution. Also AR method has an advantage over FFT that, it needs shorter duration data records than FFT. Autoregressive (AR) model is suitable for signals which have sudden peaks in their frequency spectrums. Inversely, moving average (MA) model is suitable for signals which have no sharp peaks in their frequency spectrums. But the autoregressive moving average (ARMA) model is suitable for signals which have both type peaks in their frequency spectrums. AR model parameters can be calculated by solving a set of linear equations but calculating ARMA parameters requires the solution of a set of nonlinear equations, which is computationally expensive. For this reason, AR models are more commonly used than ARMA model (Akin and Kiymik, 2000; Cohen, 2000; Guler et al., 2001; Marchant, 2003). Since, EEG signals contain peaks at some frequencies AR model can be used by using Burg or Levinson–Durbin Algorithm. Multiple signal classification (MUSIC) method have been used for estimating frequencies of different signals, and can be used for obtaining PSD estimates of various signals (Übeyli and Guler, 2003).

The STFT and wavelet are popular methods of analyzing nonstationary signals. Since STFT is simple and computationally efficient and yields reliable time–frequency plots for slowly varying signals. The major drawback is that there is a compromise between time and frequency resolution of the decomposition. On the other hand, wavelets handle the time–frequency resolution compromise in a different manner to the STFT. Rather than having uniform time and frequency resolution across the time–frequency plane, wavelets offer good time resolution at high frequencies and good frequency resolution at low frequencies. This property can be very useful in the detection of short-time transients, such as high frequency waves (Marchant, 2003). But AR and MUSIC methods are faster than Continuous Wavelet transform techniques, especially in real time applications.

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 (Basheer and Hajmeer, 2000; Fausett, 1994; Hajmeer and Basheer, 2003; Haselsteiner and Pfurtscheller, 2000; Sun and Sclabassi, 2000). Pradhan et al. (1996) uses the raw EEG as an input to a neural network while Weng and Khorasani (1996) uses the features proposed by Gotman and Wang (1991) with an adaptive structure neural network, but his results show a poor false detection rate.

In this work, the MUSIC, AR and periodogram methods were used to analyze epileptiform discharges in recorded brain waves (EEG) for patient with absence seizure (petit mal). Neurologists make the absence seizure epileptic diagnosis primarily through visual identification of the so-called 3 Hz spike and wave complex (Kiymik et al., 2004; Subasi, 2005). Power spectral densities (PSDs) of EEG signals obtained from healthy and unhealthy (epileptic patient) subjects were computed by MUSIC, AR and periodogram methods. Then these PSDs were used as an input to classification system. For developing classifiers for the detection of epileptic seizure in multi-channel EEG the traditional method of LR to the more advanced neural network techniques. In the neural network techniques, the multilayer perceptron neural network (MLPNN) was used with backpropagation training algorithm. The choice of this network was based on the fact that it is the most popular type of ANNs. In these methods we used PSDs of EEG signals as an input to classification system with two discrete outputs: epileptic seizure or not.

2. Materials and method

2.1. Subjects and data acquisition

The EEG data used in our study was 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 our study we used 11 subjects, five of them epileptic patients. Twenty absence seizures (petit mal) from five epileptic patients admitted for video-EEG monitoring were analyzed. EEG signals recorded from a patient with absence seizure epileptic discharge is shown in Fig. 1, and normal EEG signal is shown in Fig. 2. The different stages of EEG signals were determined by two physicians. EEG data were acquired with Ag/AgCl disk electrodes placed using the 10–20 international electrode placement system. For this study, four channel recordings containing epileptiform events (spikes, spike and waves) were digitized at 200 samples per second using 12-bit resolution.

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. 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.

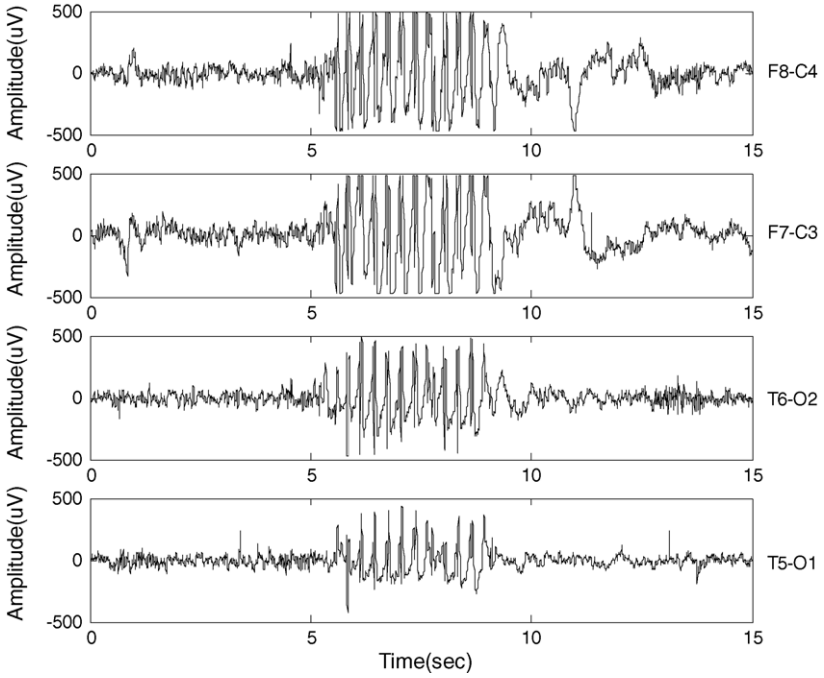

Fig. 1. Epileptic EEG signal.

2.2. Periodogram method

FFT methods such as periodogram are defined as non-parametric methods. The periodogram method used for determining the power density of the frequency components in a signal is based on Fourier conversion. For obtaining the power spectrum of an EEG signal with periodogram, EEG signals were divided into frames as 64, 128, 256, which are the powers of 2. To increase the resolution, zero-padding process was performed, because the resolution of FFT is inverse at the time when the sampled data were presented.

It is impossible to work with an infinite length signal practically. Therefore the signal is windowed with a proper window. When the bioelectric signals are studied, especially

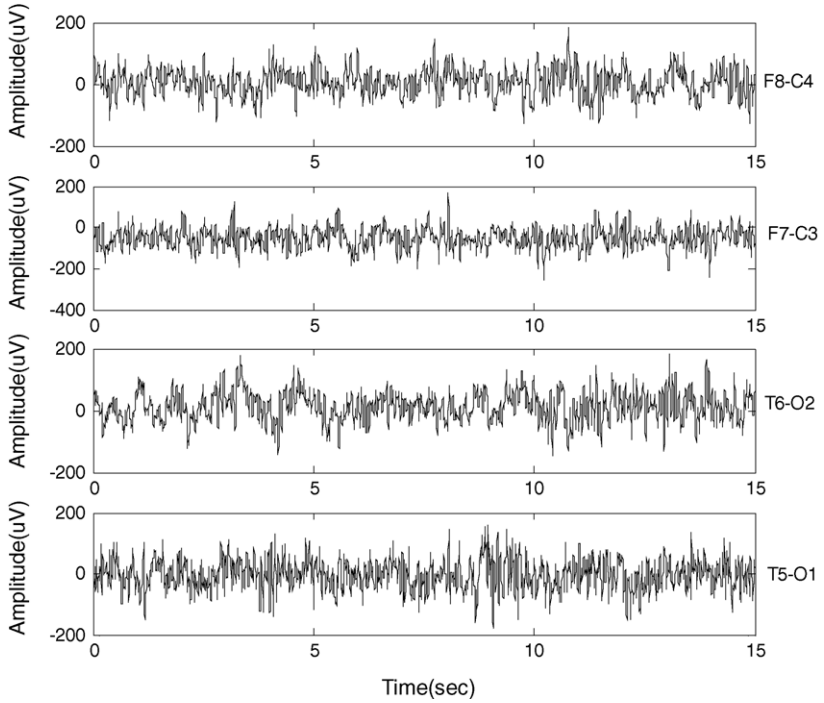

Fig. 2. Normal EEG signal.

better appropriate windows are rectangular and Hanning windows on the spectral analysis of these signals (Akin and Kiymik, 2000).

Periodogram spectral estimator is one of the FFT methods and relies on the definition of the power spectrum density (PSD) given by

$$P(f) = \lim_{N\to\infty} E\left\{\frac{1}{N}\left|\sum_{n=1}^{N} x(n)\,\mathrm{e}^{-2\pi fn}\right|^2\right\} \quad (1)$$

By neglecting the expectation operator and the limit operation in Eq. (1), which cannot be performed when the only available information on the signal consists of the samples $\{x(n)\}_{n=1}^{N}$ the periodogram spectral estimator is defined as

$$(\hat{P}_{\mathrm{PER}}(f) = \lim_{N\to\infty} \frac{1}{N}\left|\sum_{n=1}^{N} x(n)\,\mathrm{e}^{-2\pi fn}\right|^2 \quad (2)$$

Then, evaluation of $P_{\mathrm{PER}}(f)$ at the frequency samples basically reduces to the computation of the following discrete Fourier transform (DFT):

$$X(k) = \sum_{n=1}^{N} x(n)\,\mathrm{e}^{-\mathrm{j}(2\pi/N)nk}, \quad k = 0, \ldots, N-1 \quad (3)$$

where $X(k)$ is expressed as the discrete Fourier coefficient, N the length of available data and $x(n)$ is the input signal on the time domain. The procedure that computes Eq. (3) is called as FFT algorithm. The estimated periodogram can be computed by use of the DFT, which in turn is efficiently computed by the FFT algorithm (Proakis and Manolakis, 1996).

2.3. AR method for spectral analysis

The model-based (parametric) methods are based on modelling the data sequence $x(n)$ as the output of a linear system characterized by a rational system. In the model-based methods, the spectrum estimation procedure consists of two steps. The parameters of the method are estimated from given data sequence $x(n)$, $0 \le n \le N-1$. Then from these estimates, the PSD estimate is computed. AR method is the most frequently used parametric method because estimation of the AR parameters can be done easily by solving linear equations. In the AR method, data can be modelled as output of a causal, all-pole, discrete filter whose input is white noise.

To obtain stable and high performance AR method, some factors must be taken into consideration such as selection of the optimum estimation method, selection of the model order, the length of the signal which will be modelled, and the level of stationary of the data (Akin and Kiymik, 2000; Guler et al., 2001).

The AR spectral estimation methods are based on estimation of either the AR parameters or the reflection coefficients. The method developed by Burg for AR parameter estimation is based on the minimization of the forward and backward prediction errors and on estimation of the reflection coefficient. The forward and backward prediction errors for a pth-order model are defined as

$$\hat{e}_{\mathrm{f},p}(n) = x(n) + \sum_{i=1}^{p} \hat{a}_{p,i}x(n-i), \quad n = p+1, \ldots, N \quad (4)$$

$$\hat{e}_{\mathrm{b},p}(n) = x(n-p) + \sum_{i=1}^{p} \hat{a}^*_{p,i}x(n-p+i),$$
$$n = p+1, \ldots, N \quad (5)$$

The AR parameters related to the reflection coefficient $\hat{k}_p$, are given by

$$\hat{a}_{p,i} = \begin{cases} \hat{a}_{p-1,i} + \hat{k}_p\hat{a}^*_{p-1,p-i}, & i = 1, \ldots, p-1 \\ \hat{k}_{p,} & i = p \end{cases} \quad (6)$$

The Burg method considers the recursive-in-order estimation of $\hat{k}_p$, given that the AR coefficients for order $p-1$ have been computed. The reflection coefficient estimate is given by

$$\hat{k}_p = -\frac{2\sum_{n-p+1}^{N}\hat{e}_{\mathrm{f},p-1}(n)\hat{e}^*_{\mathrm{b},p-1}(n-1)}{\sum_{n-p+1}^{N}[|\hat{e}_{\mathrm{f},p-1}(n)|^2 + |\hat{e}^*_{\mathrm{b},p-1}(n-1)|^2]} \quad (7)$$

The prediction errors satisfy the following recursive-in-order expressions:

$$\hat{e}_{\mathrm{f},p}(n) = \hat{e}_{\mathrm{f},p-1}(n) + \hat{k}_p\hat{e}_{\mathrm{b},p-1}(n-1) \quad (8)$$

$$\hat{e}_{\mathrm{b},p}(n) = \hat{e}_{\mathrm{b},p-1}(n-1) + \hat{k}^*_p\hat{e}_{\mathrm{f},p-1}(n) \quad (9)$$

and these expressions are used to develop a recursive-in-order algorithm for estimating the AR coefficients. From the estimates of the AR parameters, PSD estimation is defined as

$$\hat{P}(f) = \frac{\hat{e}_p}{\left|1 + \sum_{k=1}^{p}\hat{a}_p(k)\,\mathrm{e}^{-\mathrm{j}2\pi jk}\right|^2} \quad (10)$$

where $\hat{e}_p = \hat{e}_{\mathrm{f},p} + \hat{e}_{\mathrm{b},p}$ is the total least-squares error (Cohen, 2000; Guler et al., 2001).

2.4. MUSIC method

Subspace-based methods are used for estimating frequencies and powers of signals from noisy measurements. These methods are based on an eigen-decomposition of the correlation matrix of the noisy signal. Even when the signal-to-noise ratio (SNR) is low, the subspace-based methods give frequency spectra of high resolution. These methods are best suited to signals that can be assumed to be composed of several special sinusoids buried in noise (Friedlander and Weiss, 1994; Porat and Friedlander, 1988; Stoica and Nehorai, 1990; Weiss and Friedlander, 1994). In this study, subspace-based MUSIC method had been selected to generate the PSD estimates. The polynomial $A(f)$ which contains zeros on the unit

circle can then be used to estimate the PSD:

$$A(f) = \sum_{k=0}^{m} a_k \, e^{-j2\pi fk} \tag{11}$$

where $A(f)$ represents the desired polynomial, a_k represents coefficients of the desired polynomial, and m represents the order of the eigenfilter, $A(f)$.

The polynomial can also be expressed in terms of the autocorrelation matrix R of the input signal. Assuming that the noise is white:

$$R = E\{x(n)^* x(n)^T\} = \sigma v^2 \tag{12}$$

where $x(n)$ is the observed signal, S represents the signal direction matrix of dimension $(m+1) \times L$ and L is the dimension of the signal subspace, R the autocorrelation matrix of dimension $(m+1) \times (m+1)$, P the signal power matrix of dimension $(L) \times (L)$, σv^2 represents the noise power, * represents the complex conjugate, # represents the complex conjugate transposed, T shows the matrix transposed. S the signal direction matrix is expressed as

$$S = [\, Sw_1 \quad Sw_2 \quad \cdots \quad Sw_L \,] \tag{13}$$

where $w_1, w_2, \ldots, w_L$ represent the signal frequencies:

$$Sw_i = [\, 1 \quad e^{jw_i} \quad e^{j2w_i} \quad \cdots \quad e^{jmw_i} \,]^T \quad i = 1, 2, \ldots, L \tag{14}$$

In practice, it is common to construct the estimated autocorrelation matrix $\hat{R}$ from the autocorrelation lags:

$$\hat{R}(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n+k)x(n), \quad k = 0, 2, \ldots, m \tag{15}$$

where k is the autocorrelation lag index and N is the number of the signal samples. Then, the estimated autocorrelation matrix becomes:

$$\hat{R}(k) = \begin{bmatrix} \hat{R}(0) & \hat{R}(1) & \hat{R}(2) & \cdot & \cdot & \hat{R}(m) \\ \hat{R}(1) & & & \cdot & \cdot & \hat{R}(m-1) \\ \hat{R}(2) & & \cdot & & & \hat{R}(m-2) \\ & & \cdot & \cdot & & \cdot \\ & & \cdot & & \cdot & \cdot \\ \hat{R}(m) & \hat{R}(m-1) & \cdot & \cdot & \cdot & \hat{R}(0) \end{bmatrix} \tag{16}$$

Multiplying by the eigenvector of the autocorrelation matrix a, Eq. (7) can be rewritten as

$$\hat{R}_a = \mathrm{SPS}\, a + \sigma v^2 a \tag{17}$$

where a represents the eigenvector of the estimated autocorrelation matrix $\hat{R}$ and a is expressed as $[\, a_0 \quad a_1 \quad \cdots \quad a_m \,]^T$.

The subspace-based method uses only the eigenvector corresponding to the minimum eigenvalue to construct the desired polynomial (Eq. (11)) and to calculate the spectrum. Thus, it determines a, such that $S^{\#} a = 0$. The eigenvector a can then be considered to lie in the noise subspace, and Eq. (17) reduces to

$$\hat{R}_a = \sigma v^2 a \tag{18}$$

under the constraint $a^{\#} a = 1$ where σv^2 is the noise power which in the subspace-based method is the same as the minimum eigenvalue corresponding to the eigenvector a.

In principle, under the assumption of white noise all noise subspace eigenvalues should be equal:

$$\lambda_1 = \lambda_2 = \cdots = \lambda_k = \sigma v^2, \tag{19}$$

where λ_i represents the noise subspace eigenvalues, $i = 1, 2, \ldots, K$ and K represents the dimension of the noise subspace.

The multiple signal classification (MUSIC) method is a noise subspace frequency estimator. This method proposed by Schmidt (1986) eliminates the effects of spurious zeros by using the averaged spectra of all of the eigenvectors corresponding to the noise subspace. The resultant PSD is determined from

$$P_{\mathrm{MUSIC}}(f) = \frac{1}{(1/K)\sum_{i=0}^{k-1} |A_i(f)|^2} \tag{20}$$

where K represents the dimension of noise subspace, $A_i(f)$ represents the desired polynomial that corresponds to all the eigenvectors of the noise subspace (Friedlander, 1990; Friedlander and Weiss, 1994; Porat and Friedlander, 1988; Proakis and Manolakis, 1996; Stoica and Nehorai, 1990).

2.5. Logistic regression

Logistic regression (LR) is a widely used statistical modelling technique in which the probability, P_1, of dichotomous outcome event is related to a set of explanatory variables in the form:

$$\mathrm{logit}(P_1) = \ln\left(\frac{P_1}{1-P_1}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \tag{21}$$

In Eq. (21), β_0 is the intercept and $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients associated with the explanatory variable $x_1, x_2, \ldots, x_n$. 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 (e.g., 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. (21), 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, LR 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. LR 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. (21) as

$$P_1(x) = \frac{1}{1 + e^{-\text{logit}(P_1(x))}} = \frac{1}{1 + e^{-\beta_0 + \sum_{i=1}^{n} \beta_i x_i}} \quad (22)$$

Unlike OLR, LR will force the probability values (P_1) to lie between 0 and 1 ($P_1 \to 0$ as the right-hand side of Eq. (22) approaches $-\infty$, and $P_1 \to 1$ as it approaches $+\infty$). Commonly, the maximum likelihood estimation (MLE) method is used to estimate the coefficients $\beta_0, \beta_1, \ldots, \beta_n$ in the LR equation (Hajmeer and Basheer, 2003; Hosmer and Lemeshow, 1989; Schumacher et al., 1996; Vach et al., 1996).

2.6. Artificial neural networks

Artificial neural networks (ANNs) 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 (Basheer and Hajmeer, 2000; Dreiseitl and Ohno-Machado, 2002; Fausett, 1994). 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. 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 (Chaudhuri and Bhattacharya, 2000; Fausett, 1994). In the present study, MLPNN consisted of one input layer, one hidden layer with 36 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, in our study we used the back propagation training algorithm (Dreiseitl and Ohno-Machado, 2002; Fausett, 1994; Kiymik et al., 2004).

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:

$$\text{sensitivity} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (23)$$

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:

$$\text{specificity} = \text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\% \quad (24)$$

Neural network and LR 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 cut-off 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 cut-off 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

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, we have used periodogram, AR and MUSIC methods to compute PSDs of EEG signals. Then these PSDs of EEG signals were used as an input to LR and MLPNN that could be used to detect epileptic seizure.

In Fig. 3, periodogram, AR and MUSIC PSDs of an epileptic EEG signal is given. If frequency spectrum of this figure is examined visually, it can be seen that there are more misleading peaks in periodogram PSD. When we compare these spectrums, it is seen that MUSIC method's spectrum has got sharper peaks and less misleading peaks than periodogram and AR PSDs.

Fig. 4 shows PSDs of a normal EEG signal. If these two spectrums are examined visually, although periodogram spectrum has got wide and misleading peaks, AR spectrum has got a smooth spectrum. We can say that MUSIC spectrum has got sharper and clearer peaks than periodogram and AR PSDs. From these spectrums, delta activity, alpha activity, and beta activity can be seen easily. These results show the nature of a normal EEG signal (δ, θ, α, β bands).

3.1. 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 LR and neural network classifiers, 300 examples were randomly taken from the 500 examples and used for deriving the LR 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 1.

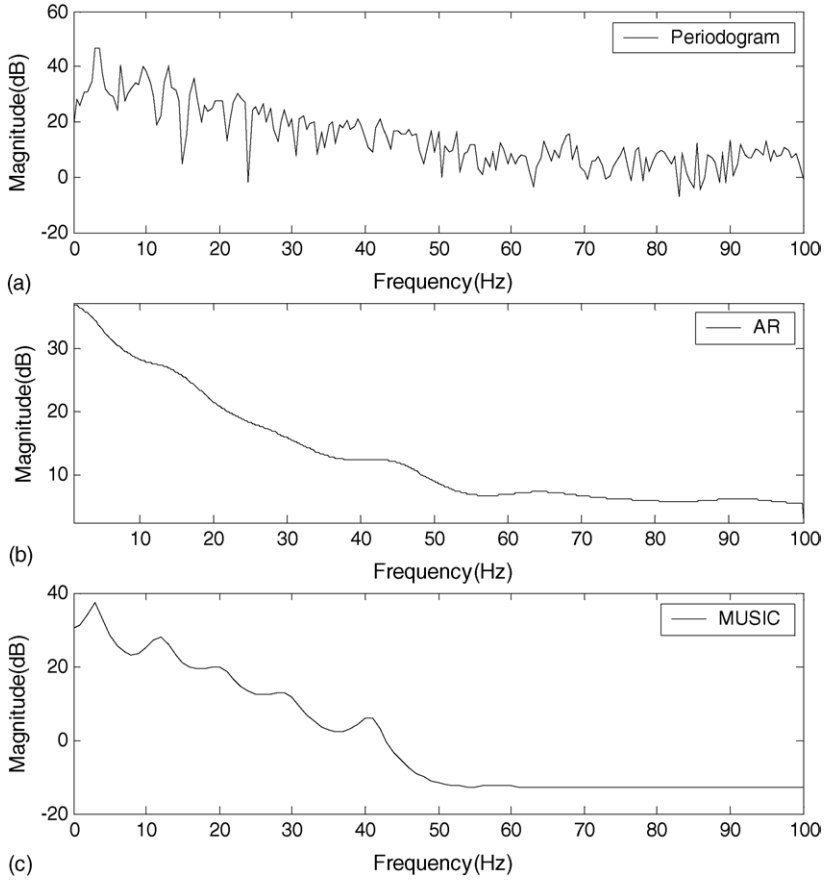

Fig. 3. PSDs of epileptic EEG signal: (a) periodogram, (b) AR, and (c) MUSIC method.

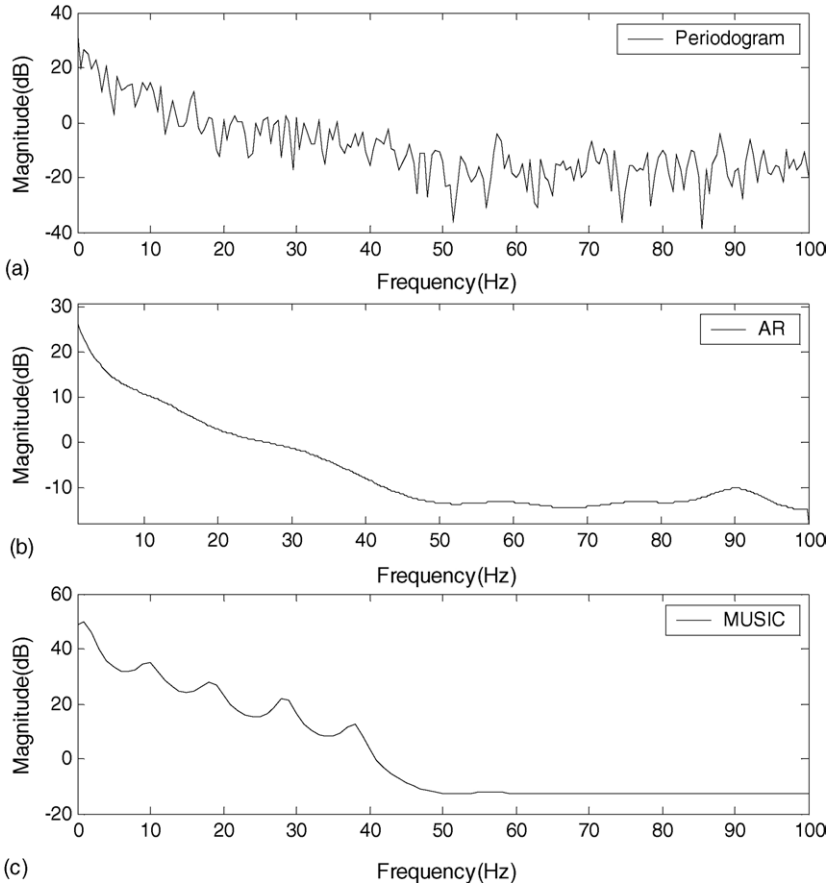

Fig. 4. PSDs of normal (healthy) EEG signal: (a) Periodogram, (b) AR, and (c) MUSIC method.

The MLPNN was designed with PSDs of EEG signal in the input layer and the output layer consisted of one node representing whether epileptic seizure was detected or not. A value of "0" was used when the experimental investigation indicated a normal 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.

Additionally, because the problem involves classification into two classes, accuracy, sensitivity and specificity 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 were used to develop the MLPNNs.

Table 1
Class distribution 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

3.2. Applying test data

LR 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 LR 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 LR-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.

Firstly, we used periodogram, AR and MUSIC spectrum of EEG signals for LR and MLPNN classification. The procedure was repeated on EEG recordings of all subjects (healthy and epileptic patients). The correct classification results for periodogram, AR and MUSIC are shown in Table 2. As seen in Table 2, periodogram spectrum gave the poorest result, AR is better than periodogram, and MUSIC is the best of all.

Table 3 shows a summary of the performance measures by using MUSIC preprocessing method. It is obvious from Table 3 that the MLPNN is ranked first in terms of its classification accuracy of the EEG signals epileptic/normal

Table 2
Comparison of correct classification performance for the periodogram, AR and MUSIC methods

Preprocessing method	Logistic regression (%)	MLPNN (%)
Periodogram	88.5	88.5
AR	89.5	90.5
MUSIC	90.5	92

Table 3
Comparison of LR and MLPNN for MUSIC preprocessing method

Classifier type	Correctly classified (%)	Specificity (%)	Sensitivity (%)	Area under ROC curve
Logistic regression	90.5	92.6	87.9	0.897
MLPNN	92	93.6	90	0.908

data 92%, while the LR-based classifier had lower accuracy (90.5%) compared to the MLPNN counterparts. The MLPNN was able to accurately predict (detect) epileptic cases 90% of sensitivity, while the LR-based classifiers indicated a detection accuracy of only 87.9%.

Also, the area under ROC curves for the classifiers (LR and MLPNN) are given in Table 3. When the area under the ROC curve in Table 3 is examined, the MLPNN has achieved an acceptable classification success with the value 0.908. However, the area under the curve has been found to be 0.897 in the LR analysis. Thus, it can be seen clearly that the performance of the MLPNN is better than the LR model.

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. Conclusion

In this paper, two approaches to develop classifiers for identifying epileptic seizure were discussed. One approach is based on the traditional method of statistical LR analysis where logistic regression equations were developed. The other approach is based on the neural network technology. Using PSDs of EEG signals, two 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 (ROC) curves of the two classifiers and two scalar performance measures derived from the confusion matrices; namely specificity and sensitivity. The MLPNN identified accurately all the epileptic and normal cases. Out of the 200 epileptic/normal cases, the LR-based classifier misclassified a total of 11 cases, while the MLPNN misclassified seven cases.

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 LR, MLPNN include their robustness to noisy data (with outliers) which can severely hamper many types of most traditional statistical methods. Other advantages of MLPNNs over LR that a 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.

References

Adeli H, Zhou Z, Dadmehr N. Analysis of EEG records in an epileptic patient using wavelet transform. J Neurosci Meth 2003;123:69–87.

Akin M, Kiymik MK. Application of periodogram and AR spectral analysis to EEG signals. J Med Syst 2000;24(4):247–56.

Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. J Microbiol Meth 2000;43:3–31.

Chaudhuri BB, Bhattacharya U. Efficient training and improved performance of multilayer perceptron in pattern classification. Neurocomputing 2000;34:11–27.

Cohen A. Biomedical signals: origin and dynamic characteristics; frequency-domain analysis. In: Joseph D, editor. The biomedical engineering handbook. 2nd ed. Bronzino: CRC Press/LLC; 2000.

Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002;35:352–9.

Fausett L. Fundamentals of neural networks architectures, algorithms, and applications. Englewood Cliffs, NJ: Prentice-Hall; 1994.

Friedlander B. A sensitivity analysis of the MUSIC algorithm. IEEE Trans Acoust Speech Signal Process 1990;38(10):1740–51.

Friedlander B, Weiss AJ. Effects of model errors on waveform estimation using the MUSIC algorithm. IEEE Trans Signal Process 1994;42(1):147–55.

Gevins A, Remond A. Handbook of electroencephalography and clinical neurophysiology, vol. I: methods of analysis of brain electrical and magnetic signals. Amsterdam: Elsevier; 1987.

Gotman J, Wang L. State-dependent spike detection: concepts and preliminary results. Electr Clin Neurophysiol 1991;79:11–9.

Guler I, Kiymik MK, Akin M, Alkan A. AR spectral analysis of EEG signals by using maximum likelihood estimation. Comput Biol Med 2001;31:441–50.

Hajmeer MM, Basheer IA. Comparison of logistic regression and neural network-based classifiers for bacterial growth. Food Microbiol 2003;20:43–55.

Haselsteiner E, Pfurtscheller G. Using time-dependent neural networks for EEG classification. IEEE Trans Rehabil Eng 2000;8:457–63.

Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley; 1989.

Kiymik MK, Akin M, Subasi A. Automatic recognition of alertness level by using wavelet transform and artificial neural network. J Neurosci Meth 2004;139:231–40.

Marchant BP. Time frequency analysis for biosystems engineering review paper. Biosyst Eng 2003;85(3):261–81.

Porat B, Friedlander B. Analysis of the asymptotic relative efficiency of the MUSIC algorithm. IEEE Trans Acoust Speech Signal Process 1988;36(4):532–44.

Pradhan N, Sadasivan PK, Arunodaya GR. Detection of seizure activity in EEG by an artificial neural network: a preliminary study. Comput Biomed Res 1996;29:303–13.

Proakis JG, Manolakis DG. Digital signal processing principles, algorithms and applications. New Jersey: Prentice-Hall; 1996.

Schmidt RO. Multiple emitter location and signal parameter estimation. IEEE Trans Antennas Propag 1986;AP-34(3):276–80.

Schneble H, Matthes A. Epilepsy in children. Germany: Kork Epilepsy Center; 1996. pp. 3–30.

Schumacher M, Robner R, Vach W. Neural networks and logistic regression: part I. Comput Stat Data Anal 1996;21:661–82.

Stoica P, Nehorai A. MUSIC, maximum likelihood, and Cramer–Rao bound: further results and comparisons. IEEE Trans Acoust Speech Signal Process 1990;12:2140–50.

Subasi A. Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. Expert Syst Appl 2005;28:701–11.

Sun M, Sclabassi RJ. The forward EEG solutions can be computed using artificial neural networks. IEEE Trans Biomed Eng 2000;47:1044–50.

Übeyli ED, Guler I. Comparison of eigenvector methods with classical and model-based methods in analysis of internal carotid arterial Doppler signals. Comput Biol Med 2003;33:473–93.

Vach W, Robner R, Schumacher M. Neural networks and logistic regression: part II. Comput Stat Data Anal 1996;2:683–701.

Weiss AJ, Friedlander B. Effects of modeling errors on the resolution threshold of the MUSIC algorithm. IEEE Trans Signal Process 1994;42(6):1519–26.

Weng W, Khorasani K. An adaptive structure neural network with application to EEG automatic seizure detection. Neural Networks 1996;9:1223–40.