Digital Signal Processing 19 (2009) 297–308

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Combined neural network model employing wavelet coefficients for EEG signals classification

Elif Derya Übeyli *

Department of Electrical and Electronics Engineering, Faculty of Engineering, TOBB Ekonomi ve Teknoloji Üniversitesi, 06530 Söğütözü, Ankara, Turkey

A R T I C L E   I N F O

Article history:
Available online 17 July 2008

Keywords:
Combined neural network model
EEG signals classification
Diagnostic accuracy
Discrete wavelet transform

A B S T R A C T

This paper illustrates the use of combined neural network model to guide model selection for classification of electroencephalogram (EEG) signals. The EEG signals were decomposed into time–frequency representations using discrete wavelet transform and statistical features were calculated to depict their distribution. The first-level networks were implemented for the EEG signals classification using the statistical features as inputs. To improve diagnostic accuracy, the second-level networks were trained using the outputs of the first-level networks as input data. Three types of EEG signals (EEG signals recorded from healthy volunteers with eyes open, epilepsy patients in the epileptogenic zone during a seizure-free interval, and epilepsy patients during epileptic seizures) were classified with the accuracy of 94.83% by the combined neural network. The combined neural network model achieved accuracy rates which were higher than that of the stand-alone neural network model.

1. Introduction

The electroencephalogram (EEG), a highly complex signal, is one of the most common sources of information used to study brain function and neurological disorders [1]. Large amounts of data are generated by EEG monitoring systems for electroencephalographic changes, and their complete visual analysis is not routinely possible. Computers have long been proposed to solve this problem and thus, automated systems to recognize electroencephalographic changes have been under study for several years [1–4]. There is a strong demand for the development of such automated devices, due to the increased use of prolonged and long-term video EEG recordings for proper evaluation and treatment of neurological diseases and prevention of the possibility of the analyst missing (or misreading) information. Various methodologies of automated diagnosis have been adopted. However, the entire process can generally be subdivided into a number of disjoint processing modules: segment detection, feature extraction/selection, and classification (Fig. 1). The techniques developed for automated EEG event detection systems work by transforming the mostly qualitative diagnostic criteria into a more objective quantitative signal feature classification problem. In order to address this problem, the techniques such as the analysis of EEG signals for detection of electroencephalographic changes using the autocorrelation function, frequency domain features, time–frequency analysis, and wavelet transform (WT) have been used [1,5–7].

Artificial neural networks (ANNs) have been used in a great number of medical diagnostic decision support system applications because of the belief that they have great predictive power [8–10]. Many authors have shown that combining the predictions of several models often results in a prediction accuracy that is higher than that of the individual models [11–14]. The general framework for predicting using an ensemble of models consists of two levels and is often referred to as

* Fax: +90 312 2924091.
E-mail address: edubeyli@etu.edu.tr.

1051-2004/$ – see front matter
doi:10.1016/j.dsp.2008.07.004

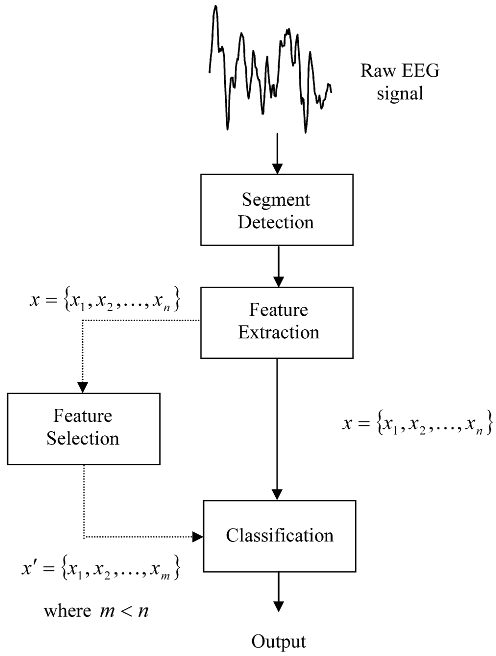

Fig. 1. Functional modules in a typical computerized electroencephalographic system.

stacked generalization [11]. In the first level, various learning methods are used to learn different models from the original data set. The predictions of the models from the first level along with the corresponding target class of the original input data are then used as inputs to learn a second level model.

The WT is designed to address the problem of nonstationary signals. It involves representing a time function in terms of simple, fixed building blocks, termed wavelets. These building blocks are actually a family of functions which are derived from a single generating function called the mother wavelet by translation and dilation operations. The main advantage of the WT is that it has a varying window size, being broad at low frequencies and narrow at high frequencies, thus leading to an optimal time–frequency resolution in all frequency ranges. Furthermore, owing to the fact that windows are adapted to the transients of each scale, wavelets lack of the requirement of stationarity. The property of time and frequency localization is known as compact support and is one of the most attractive features of the WT. The WT of a signal is the decomposition of the signal over a set of functions obtained after dilatation and translation of an analyzing wavelet [15–20]. The EEG signals, consisting of many data points, can be compressed into a few features by performing spectral analysis of the signals with the WT. These features characterize the behavior of the EEG signals. Using a smaller number of features to represent the EEG signals is particularly important for recognition and diagnostic purposes [1,17,18]. Therefore, the EEG signals were decomposed into time–frequency representations using discrete wavelet transform (DWT). Wavelet coefficients were used as feature vectors identifying characteristics of the signal that were not apparent from the original time domain signal.

In the present study, the EEG signals were classified using the combined neural network model. In the development of combined neural network for classification of the EEG signals, for the first-level models three sets of neural networks were used since there were three diagnostic classes (EEG signals recorded from healthy volunteers with eyes open, epilepsy patients in the epileptogenic zone during a seizure-free interval, and epilepsy patients during epileptic seizures). Networks in each set were trained so that they are likely to be more accurate for one type of EEG signal than the other EEG signals.

The outline of this study is as follows. In Section 2, the sets of the EEG signals used in the study are briefly described. In Section 3, spectral analysis of signals using DWT is explained in order to extract features characterizing the behavior of the signal under study. In Section 4, descriptions of neural network models including multilayer perceptron neural network (MLPNN) and combined neural network topology are presented. In Section 5, the results of the experiments involving the application of combined neural network model to the EEG signals are presented. Finally, in Section 6 the study is concluded.

2. Data selection

The data described in Ref. [21], which is publicly available, was used. In this section, only a short description is presented and refer to Ref. [21] for further details. The complete dataset consists of five sets (denoted A–E), each containing 100 single-channel EEG signals of 23.6 s. Each signal has been selected after visual inspection for artifacts and has passed a weak stationarity criterion. Sets A and B have been taken from surface EEG recordings of five healthy volunteers with eyes open and closed, respectively. Signals in two sets have been measured in seizure-free intervals from five patients in the epileptogenic zone (D) and from the hippocampal formation of the opposite hemisphere of the brain (C). Set E contains

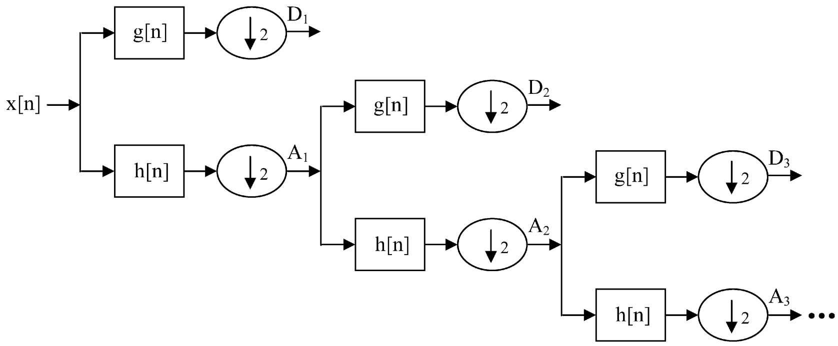

Fig. 2. Subband decomposition of discrete wavelet transform implementation; $g[n]$ is the high-pass filter, $h[n]$ is the low-pass filter.

seizure activity, selected from all recording sites exhibiting ictal activity. Sets A and B have been recorded extracranially, whereas sets C, D, and E have been recorded intracranially. In the present applications, performance degraded for a more detailed classification which further dissociated between sets A (healthy volunteer, eyes open) and B (healthy volunteer, eyes closed), and sets D (epileptogenic zone) and C (hippocampal formation of opposite hemisphere). Therefore, in the present study three dataset (A, D, E) of the complete dataset was classified.

3. Spectral analysis using discrete wavelet transform

Predicting the onset of epileptic seizure is an important and difficult biomedical problem, which has attracted substantial attention of the intelligent computing community over the past two decades [1–4]. The combined neural network model employing signal wavelet decomposition was applied to the problem. The WT provides very general techniques, which can be applied to many tasks in signal processing. One very important application is the ability to compute and manipulate data in compressed parameters, which are often called features [1,17,18]. Thus, the EEG signal, consisting of many data points, can be compressed into a few parameters. These parameters characterize the behavior of the EEG signal. This feature of using a smaller number of parameters to represent the EEG signal is particularly important for recognition and diagnostic purposes.

The WT can be categorized into continuous and discrete. Continuous wavelet transform (CWT) is defined by

$$\mathrm{CWT}(a,b)=\int_{-\infty}^{+\infty} x(t)\psi^*_{a,b}(t)\,dt, \tag{1}$$

where $x(t)$ represents the analyzed signal, a and b represent the scaling factor (dilatation/compression coefficient) and translation along the time axis (shifting coefficient), respectively, and the superscript asterisk denotes the complex conjugation. $\psi_{a,b}(\cdot)$ is obtained by scaling the wavelet at time b and scale a:

$$\psi_{a,b}(t)=\frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), \tag{2}$$

where $\psi(t)$ represents the wavelet [15,19,20].

Continuous, in the context of the WT, implies that the scaling and translation parameters a and b change continuously. However, calculating wavelet coefficients for every possible scale can represent a considerable effort and result in a vast amount of data. Therefore discrete wavelet transform (DWT) is often used. The WT can be thought of as an extension of the classic Fourier transform, except that, instead of working on a single scale (time or frequency), it works on a multiscale basis. The procedure of multiresolution decomposition of a signal $x[n]$ is schematically shown in Fig. 2. Each stage of this scheme consists of two digital filters and two downsamplers by 2. The first filter, $g[\cdot]$ is the discrete mother wavelet, high-pass in nature, and the second, $h[\cdot]$ is its mirror version, low-pass in nature. The downsampled outputs of first high-pass and low-pass filters provide the detail, D_1 and the approximation, A_1, respectively. The first approximation, A_1 is further decomposed and this process is continued as shown in Fig. 2 [15,16].

All wavelet transforms can be specified in terms of a low-pass filter h, which satisfies the standard quadrature mirror filter condition:

$$H(z)H\left(z^{-1}\right)+H(-z)H\left(-z^{-1}\right)=1, \tag{3}$$

where $H(z)$ denotes the z-transform of the filter h. Its complementary high-pass filter can be defined as

$$G(z)=zH\left(-z^{-1}\right). \tag{4}$$

A sequence of filters with increasing length (indexed by i) can be obtained:

$$H_{i+1}(z)=H\left(z^{2^i}\right)H_i(z), \qquad G_{i+1}(z)=G\left(z^{2^i}\right)H_i(z), \quad i=0,\dots,I-1, \tag{5}$$

with the initial condition $H_0(z) = 1$. It is expressed as a two-scale relation in time domain

$$h_{i+1}(k) = [h]_{\uparrow 2^i} * h_i(k), \qquad g_{i+1}(k) = [g]_{\uparrow 2^i} * h_i(k), \tag{6}$$

where the subscript $[\cdot]_{\uparrow m}$ indicates the up-sampling by a factor of m and k is the equally sampled discrete time.

The normalized wavelet and scale basis functions $\varphi_{i,l}(k)$, $\psi_{i,l}(k)$ can be defined as

$$\varphi_{i,l}(k) = 2^{i/2} h_i\bigl(k - 2^i l\bigr), \qquad \psi_{i,l}(k) = 2^{i/2} g_i\bigl(k - 2^i l\bigr), \tag{7}$$

where the factor $2^{i/2}$ is an inner product normalization, i and l are the scale parameter and the translation parameter, respectively. The DWT decomposition can be described as

$$a_{(i)}(l) = x(k) * \varphi_{i,l}(k), \qquad d_{(i)}(l) = x(k) * \psi_{i,l}(k), \tag{8}$$

where $a_{(i)}(l)$ and $d_i(l)$ are the approximation coefficients and the detail coefficients at resolution i, respectively [15].

4. Description of neural network models

Neural networks have been successfully used in a variety of medical applications [8–10]. Recent advances in the field of neural networks have made them attractive for analyzing signals. The application of neural networks has opened a new area for solving problems not resolvable by other signal processing techniques [1,3,13,14]. In contrast to the conventional spectral analysis methods, ANNs not only model the signal, but also make a decision as to the class of signal. Another advantage of ANN analysis over existing methods of biomedical signals analysis is that, after an ANN has trained satisfactorily and the values of the weights and biases have been stored, testing and subsequent implementation is rapid. During implementation of the combined neural networks the predictions of the networks in the first-level were combined by a second-level neural network. The results of the present study showed that significant improvement is achieved in accuracy by applying neural networks as the second-level model compared to the stand-alone MLPNNs.

4.1. Multilayer perceptron neural network

The MLPNN is a nonparametric technique for performing a wide variety of detection and estimation tasks [22–24]. In the MLPNN, each neuron j in the hidden layer sums its input signals x_i after multiplying them by the strengths of the respective connection weights w_{ji} and computes its output y_j as a function of the sum:

$$y_j = f\Bigl(\sum w_{ji} x_i\Bigr), \tag{9}$$

where f is activation function that is necessary to transform the weighted sum of all signals impinging onto a neuron. The activation function (f) can be a simple threshold function, or a sigmoidal, hyperbolic tangent, or radial basis function. In the present study, in the hidden layer and the output layer, the activation function f was the sigmoidal function.

The sum of squared differences between the desired and actual values of the output neurons E is defined as:

$$E = \frac{1}{2} \sum_j (y_{dj} - y_j)^2, \tag{10}$$

where y_{dj} is the desired value of output neuron j and y_j is the actual output of that neuron. Each weight w_{ji} is adjusted to reduce E as rapidly as possible. How w_{ji} is adjusted depends on the training algorithm adopted [22–24].

4.2. Combined neural network models

Combined neural network models often result in a prediction accuracy that is higher than that of the individual models. This construction is based on a straightforward approach that has been termed stacked generalization. Training data that are difficult to learn usually demonstrate high dispersion in the search space due to the inability of the low-level measurement attributes to describe the concept concisely. Because of the complex interactions among variables and the high degree of noise and fluctuations, a significant number of data used for applications are naturally available in representations that are difficult to learn. Transforming the data into a more appropriate representation can facilitate the learning process. For instance, using a smaller number of parameters, which are often called features, to represent the signal under study is particularly important for recognition and diagnostic purposes. Given any set of features for data representation, it is therefore important to estimate the difficulty of learning the underlying concepts using that training data. The learning system should then seek to transform the representations into a space that is easier for learning purposes [25–27].

Piramuthu et al. [28] show that the degree of difficulty in training a neural network is inherent in the given set of training examples. By developing a technique for measuring this learning difficulty, they devise a feature construction methodology that transforms the training data and attempts to improve both the classification accuracy and computational times of ANN algorithms. The fundamental notion is to organize data by intelligent preprocessing, so that learning is facilitated.

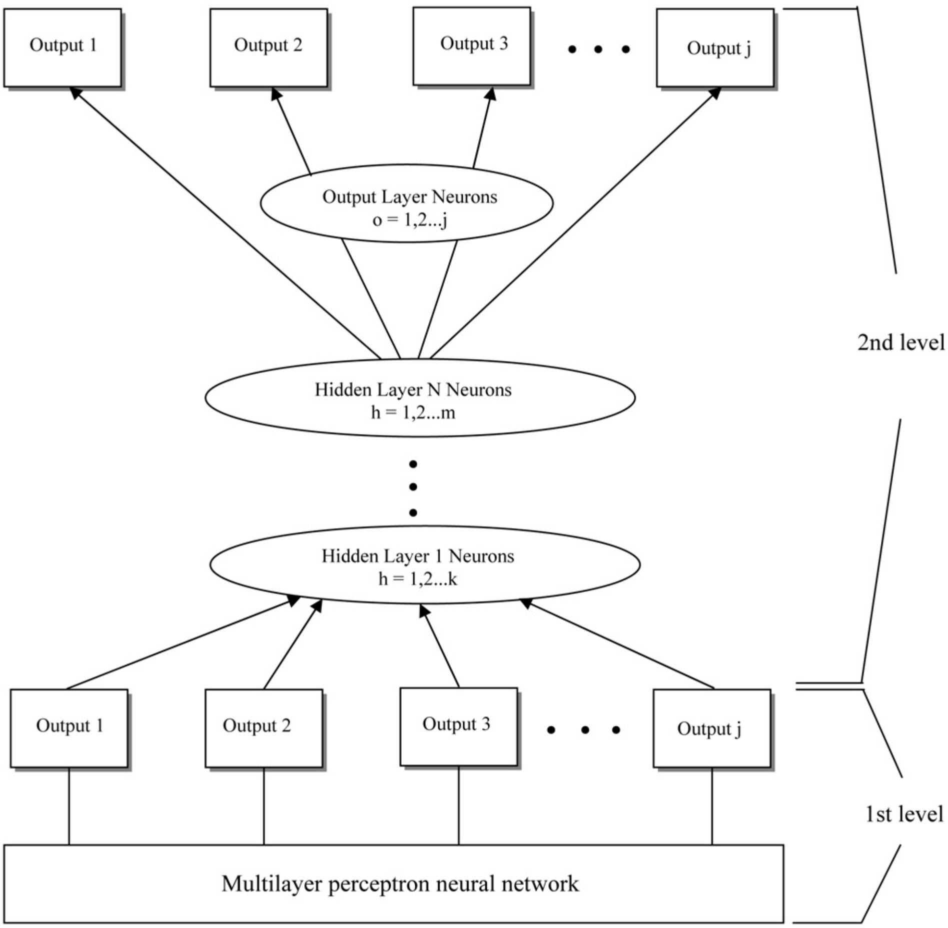

Fig. 3. A second-level neural network is used to combine the predictions of the first-level neural networks.

The stacked generalization concepts formalized by Wolpert [11] consider these ideas and refer to schemes for feeding information from one set of generalizers to another before forming the final predicted value (output). The unique contribution of stacked generalization is that the information fed into the net of generalizers comes from multiple partitionings of the original learning set. The stacked generalization scheme can be viewed as a more sophisticated version of cross validation and has been shown experimentally to effectively improve generalization ability of ANN models over using stand-alone neural networks.

Essentially a generalizer is a mapping of a set of m pairs $\{x_k \in R^n, y_k \in R^p\}$, $1 \leqslant k \leqslant m$, together with an unseen instance, $\{x_{k+1} \in R^n\}$ into a prediction $\{y_{k+1} \in R^p\}$. For time series analysis, $p = 1$ and the R^{n+1} space inhabited by the original training set is labeled the first-level space. Any generalizer derived from the first-level space is termed a first-level generalizer. Similarly, the training set at the next-level, which may be the output of the first-level generalizer plus other input–output pairs constitutes the second-level space. Its generalizer is a second-level generalizer. This pattern proceeds in like manner through succeeding spaces [25–27].

The combined neural network model used in the present study is shown in Fig. 3. The MLPNNs were used at the first-level and second-level for the implementation of the combined neural network proposed in this study. This configuration occurred on the theory that MLPNN has features such as the ability to learn and generalize, smaller training set requirements, fast operation, ease of implementation. In both the first-level and second-level analysis, the Levenberg–Marquardt training algorithm was used. 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 [22–24]. 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 have been proposed. Optimization methods such as second-order methods (conjugate gradient, quasi-Newton, Levenberg–Marquardt) 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 [29–32]. The combined neural network model proposed for classification of the EEG signals was implemented by using the MATLAB software package (MATLAB version 6.5 with neural networks toolbox).

4.3. Levenberg–Marquardt algorithm

Essentially, the Levenberg–Marquardt algorithm is a least-squares estimation algorithm based on the maximum neighborhood idea. Let $E(\mathbf{w})$ be an objective error function made up of m individual error terms $e_i^2(\mathbf{w})$ as follows:

$$E(\mathbf{w}) = \sum_{i=1}^{m} e_i^2(\mathbf{w}) = \left\| f(\mathbf{w}) \right\|^2, \tag{11}$$

where $e_i^2(\mathbf{w}) = (\mathbf{y}_{di} - \mathbf{y}_i)^2$ and $\mathbf{y}_{di}$ is the desired value of output neuron i, $\mathbf{y}_i$ is the actual output of that neuron.

It is assumed that function $f(\cdot)$ and its Jacobian J are known at point $\mathbf{w}$. The aim of the Levenberg–Marquardt algorithm is to compute the weight vector $\mathbf{w}$ such that $E(\mathbf{w})$ is minimum. Using the Levenberg–Marquardt algorithm, a new weight vector $\mathbf{w}_{k+1}$ can be obtained from the previous weight vector $\mathbf{w}_k$ as follows:

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \delta\mathbf{w}_k, \tag{12}$$

where $\delta\mathbf{w}_k$ is defined as

$$\delta\mathbf{w}_k = -\left(J_k^T f(\mathbf{w}_k)\right)\left(J_k^T J_k + \lambda\mathbf{I}\right)^{-1}. \tag{13}$$

In Eq. (13), J_k is the Jacobian of f evaluated at $\mathbf{w}_k$, λ is the Marquardt parameter, $\mathbf{I}$ is the identity matrix [29–32]. The Levenberg–Marquardt algorithm may be summarized as follows:

(i) compute $E(\mathbf{w}_k)$,
(ii) start with a small value of λ ($\lambda = 0.01$),
(iii) solve Eq. (13) for $\delta\mathbf{w}_k$ and compute $E(\mathbf{w}_k + \delta\mathbf{w}_k)$,
(iv) if $E(\mathbf{w}_k + \delta\mathbf{w}_k) \geqslant E(\mathbf{w}_k)$, increase λ by a factor of 10 and go to (iii),
(v) if $E(\mathbf{w}_k + \delta\mathbf{w}_k) < E(\mathbf{w}_k)$, decrease λ by a factor of 10, update $\mathbf{w}_k : \mathbf{w}_k \leftarrow \mathbf{w}_k + \delta\mathbf{w}_k$ and go to (iii).

5. Experimental results

5.1. Feature extraction using discrete wavelet transform

The EEG signals can be considered as a superposition of different structures occurring on different time scales at different times. One purpose of wavelet analysis is to separate and sort these underlying structures of different time scales. It is known that the WT is better suited to analyzing nonstationary signals, since it is well localized in time and frequency. The property of time and frequency localization is known as compact support and is one of the most attractive features of the WT. The main advantage of the WT is that it has a varying window size, being broad at low frequencies and narrow at high frequencies, thus leading to an optimal time–frequency resolution in all frequency ranges. Therefore, spectral analysis of the EEG signals was performed using the DWT as described in Section 3.

Selection of appropriate wavelet and the number of decomposition levels is very important in analysis of signals using the WT. The number of decomposition levels is chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the signal that correlate well with the frequencies required for classification of the signal are retained in the wavelet coefficients. In the present study, the number of decomposition levels was chosen to be 4. Thus, the EEG signals were decomposed into the details D_1–D_4 and one final approximation, A_4. Usually, tests are performed with different types of wavelets and the one which gives maximum efficiency is selected for the particular application. The smoothing feature of the Daubechies wavelet of order 2 made it more suitable to detect changes of the EEG signals. Therefore, the wavelet coefficients were computed using the Daubechies wavelet of order 2 in the present study. The wavelet coefficients were computed using the MATLAB software package.

Selection of the ANN inputs is the most important component of designing the neural network based on pattern classification since even the best classifier will perform poorly if the inputs are not selected well. Input selection has two meanings: (1) which components of a pattern, or (2) which set of inputs best represent a given pattern. The computed discrete wavelet coefficients provide a compact representation that shows the energy distribution of the signal in time and frequency. Therefore, the computed detail and approximation wavelet coefficients of the EEG signals were used as the feature vectors representing the signals. A rectangular window, which was formed by 256 discrete data, was selected so that it contained a single EEG segment. For each EEG segment, the detail wavelet coefficients (d^k, $k = 1, 2, 3, 4$) at the first, second, third and fourth levels ($129 + 66 + 34 + 18$ coefficients) and the approximation wavelet coefficients (a^4) at the fourth-level (18 coefficients) were computed. Then 265 wavelet coefficients were obtained for each EEG segment. In order to reduce the

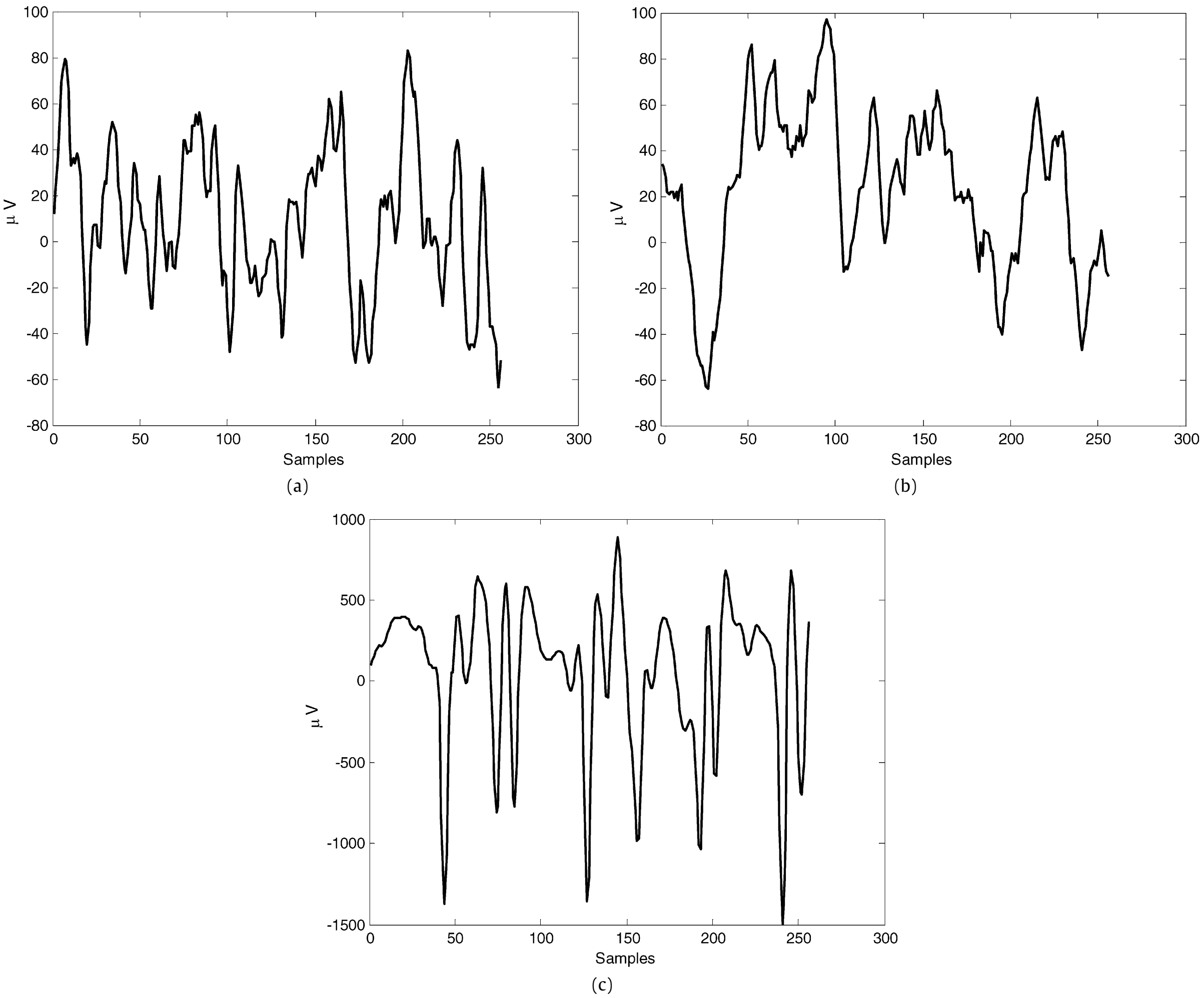

Fig. 4. Waveforms of the EEG segments: (a) set A (EEG signals recorded from healthy volunteers with eyes open), (b) set D (EEG signals recorded from epilepsy patients in the epileptogenic zone during a seizure-free interval), (c) set E (EEG signals recorded from epilepsy patients during epileptic seizures).

dimensionality of the extracted feature vectors, statistics over the set of the wavelet coefficients were used. The following statistical features were used to represent the time–frequency distribution of the EEG signals:

1. Maximum of the wavelet coefficients in each subband.
2. Minimum of the wavelet coefficients in each subband.
3. Mean of the wavelet coefficients in each subband.
4. Standard deviation of the wavelet coefficients in each subband.

5.2. Application of combined neural network model to EEG signals

The waveforms of three different EEG segments classified in the present study are shown in Figs. 4a–4c. For the three diagnostic classes (set A—EEG signals recorded from healthy volunteers with eyes open, set D—EEG signals recorded from epilepsy patients in the epileptogenic zone during a seizure-free interval, and set E—EEG signals recorded from epilepsy patients during epileptic seizures) training and test sets were formed by 1200 vectors (400 vectors from each class) of 20 dimensions (extracted feature vectors). The detail wavelet coefficients at the first decomposition level of the three types of EEG segments are given in Figs. 5a–5c, respectively. Table 1 presents the extracted features of exemplary records from three classes of the signals under study. It can be noted that the detail wavelet coefficients of the three types of EEG segments are different from each other. In order to extract features, the wavelet coefficients corresponding to the D_1–D_4 and A_4 frequency bands of the three types of EEG segments were computed.

Automated diagnostic systems aim to enhance the ability to detect pathological structures in medical examinations and to support evaluation of pathological findings during the diagnostic procedure. The techniques developed for automated electroencephalographic change detection transform the mostly qualitative diagnostic criteria into a more objective quan-

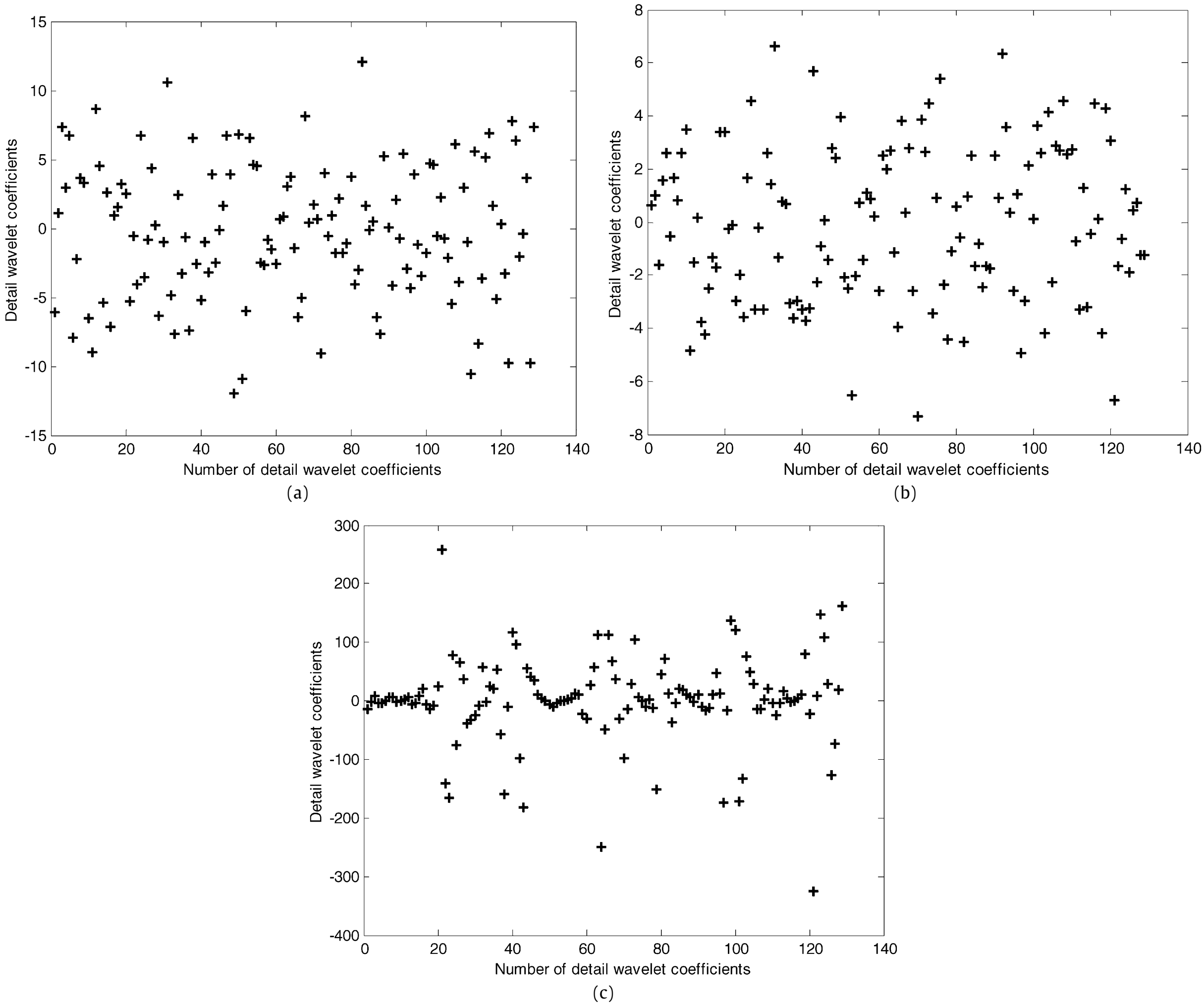

Fig. 5. The detail wavelet coefficients at the first decomposition level of the EEG segments: (a) set A (EEG signals recorded from healthy volunteers with eyes open), (b) set D (EEG signals recorded from epilepsy patients in the epileptogenic zone during a seizure-free interval), (c) set E (EEG signals recorded from epilepsy patients during epileptic seizures).

titative signal feature classification problem. For pattern processing problems to be tractable requires the conversion of patterns to features, which are condensed representations of patterns, ideally containing only salient information. Therefore, the combined neural networks and MLPNNs employing wavelet coefficients were implemented for automated electroencephalographic changes detection.

The combined neural network topology used for classification of the EEG segments is shown in Fig. 6. Three sets of neural networks for the first-level models were trained since there were three diagnostic classes (healthy segments, seizure free epileptogenic zone segments, epileptic seizure segments). Networks in each set were trained so that they are likely to be more accurate for one type of EEG segment than the other EEG segments. The network topology was the MLPNN with a single hidden layer. Each network had 20 input neurons, equal to the number of input feature vectors. The feature vectors were calculated for the frequency bands D_1–D_4 and A_4 as explained in Section 5.1. The number of hidden neurons was 20 and the number of output was 3. Samples with target outputs set A (EEG signals recorded from healthy volunteers with eyes open), set D (EEG signals recorded from epilepsy patients in the epileptogenic zone during a seizure-free interval), and set E (EEG signals recorded from epilepsy patients during epileptic seizures) were given the binary target values of $(0,0,1)$, $(0,1,0)$, and $(1,0,0)$, respectively. Second-level neural network was trained to combine the predictions of the first-level networks. The second-level network had 9 inputs which correspond to the outputs of the three groups of the first-level networks. The targets for the second-level network were the same as the targets of the original data. The number of outputs was three and the number of hidden neurons was chosen to be 25.

In order to compare performance of the different classifiers, for the same classification problem the stand-alone MLPNN which is the most commonly used feedforward neural networks was implemented. The single hidden layered (20 hidden neurons) MLPNN was used to classify the EEG signals based on a feature vector (20 inputs). Different experiments were

Table 1
The extracted features of three exemplary records from three classes

Dataset	Extracted features	Wavelet coefficients				
		Subbands				
		D_1	D_2	D_3	D_4	A_4
Set A	Maximum	12.0394	31.3064	75.7695	120.0146	192.6771
	Minimum	−12.0140	−42.0737	−92.3744	−105.3666	−172.4994
	Mean	−0.2611	0.1775	1.6022	2.1703	34.4130
	Standard deviation	4.9689	14.8416	41.1865	60.3469	96.4623
Set D	Maximum	26.0292	117.9646	32.3480	88.2469	320.4451
	Minimum	−20.6820	−82.1600	−61.5424	−89.1512	−175.7673
	Mean	−0.1935	0.1121	−2.2112	−2.6360	94.1584
	Standard deviation	4.3874	19.2455	20.1756	43.6354	126.3576
Set E	Maximum	258.0806	644.3659	1524.4000	1420.1000	1639.2000
	Minimum	−325.4508	−1074.6000	−1508.9000	−1107.0000	−1917.6000
	Mean	−0.1337	0.1052	65.5614	−77.2298	281.4010
	Standard deviation	75.1448	303.6744	716.0870	614.2615	1138.5000

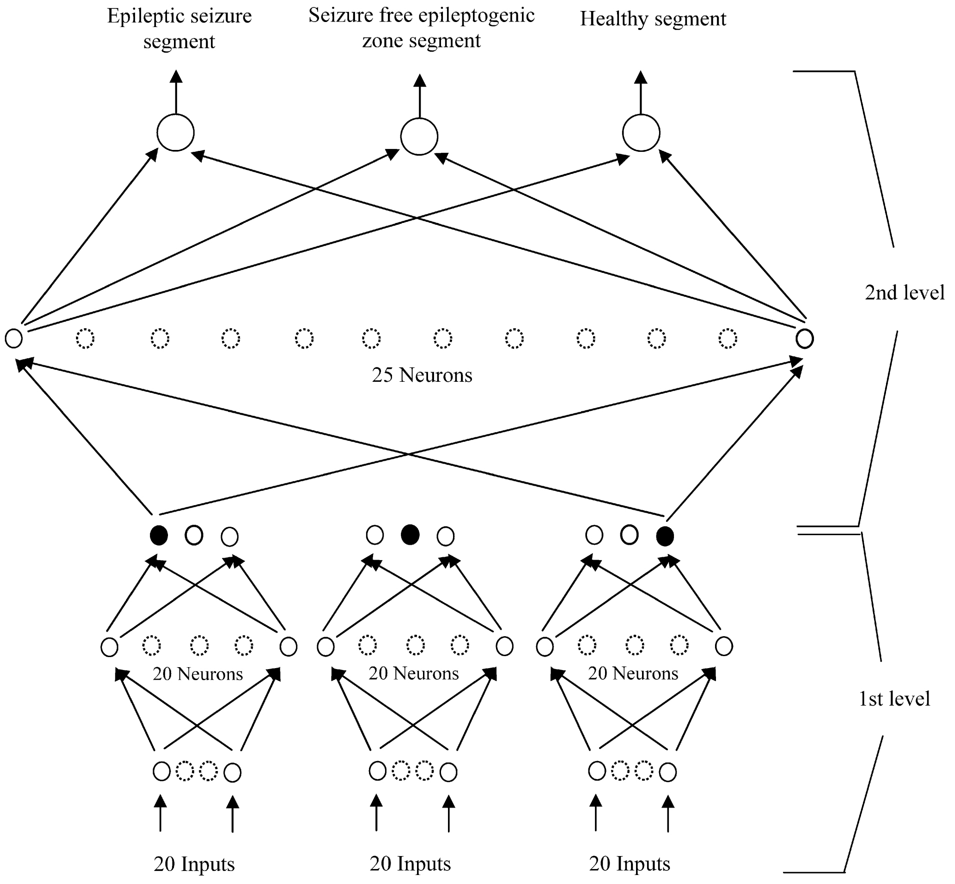

Fig. 6. A combined neural network topology used for classification of the EEG segments.

performed during implementation of these classifiers and the number of hidden neurons was determined by taking into consideration the classification accuracies. In the hidden layers and the output layers, the activation function was the sigmoidal function. The sigmoidal function with the range between zero and one introduces two important properties. First, the sigmoid is nonlinear, allowing the network to perform complex mappings of input to output vector spaces, and secondly it is continuous and differentiable, which allows the gradient of the error to be used in updating the weights.

The adequate functioning of neural networks depends on the sizes of the training set and test set. In the combined neural network and MLPNN, the 600 vectors (200 vectors from each class) were used for training and the 600 vectors (200 vectors from each class) were used for testing. A practical way to find a point of better generalization is to use a small

Table 2
Confusion matrix

Output/desired	Result (set A—healthy segment)	Result (set D—seizure free epileptogenic zone segment)	Result (set E—epileptic seizure segment)
Result (set A—healthy segment)	192	8	5
Result (set D—seizure free epileptogenic zone segment)	6	189	7
Result (set E—epileptic seizure segment)	2	3	188

Table 3
The classification accuracies

Statistical parameters	Values
Specificity	96.00%
Sensitivity (seizure free epileptogenic zone segments)	94.50%
Sensitivity (epileptic seizure segments)	94.00%
Total classification accuracy	94.83%

percentage (around 20%) of the training set for cross validation. For obtaining a better network generalization 120 vectors (40 vectors from each class) of training set, which were selected randomly, were used as cross validation set. Beside this, in order to enhance the generalization capability of the combined neural network, the training and the test sets were formed by data obtained from different subjects. For all of the segments, waveform variations were observed among the vectors belonging to the same class.

The training holds the key to an accurate solution, so the criterion to stop training must be very well described. When the network is trained too much, the network memorizes the training patterns and does not generalize well. Cross validation is a highly recommended criterion for stopping the training of a network. When the error in the cross validation increases, the training should be stopped because the point of best generalization has been reached. In both the first-level and second-level, training of neural networks was done in 700 epochs since the cross validation errors began to rise at 700 epochs. Since the values of mean square errors (MSE) converged to small constants approximately zero in 700 epochs, training of the neural networks with the Levenberg–Marquardt algorithm was determined to be successful. However, the stand-alone MLPNN (wavelet coefficients used as inputs) trained with the backpropagation algorithm had a slow convergence and MSE converged to a small constant of approximately zero in 4500 epochs. Thus, the convergence rate of the combined neural network presented in this study was found to be higher than that of the stand-alone MLPNN.

In classification, the aim is to assign the input patterns to one of several classes, usually represented by outputs restricted to lie in the range from 0 to 1, so that they represent the probability of class membership. While the classification is carried out, a specific pattern is assigned to a specific class according to the characteristic features selected for it. In this application, there were three classes: set A (EEG signals recorded from healthy volunteers with eyes open), set D (EEG signals recorded from epilepsy patients in the epileptogenic zone during a seizure-free interval), and set E (EEG signals recorded from epilepsy patients during epileptic seizures). Classification results of the combined neural network were displayed by a confusion matrix. The confusion matrix showing the classification results of the combined neural network is given in Table 2.

The test performance of the classifiers can be determined by the computation of specificity, sensitivity and total classification accuracy. The specificity, sensitivity and total classification accuracy are defined as:

Specificity: number of true negative decisions/number of actually negative cases;

Sensitivity: number of true positive decisions/number of actually positive cases;

Total classification accuracy: number of correct decisions/total number of cases.

A true negative decision occurs when both the classifier and the physician suggested the absence of a positive detection. A true positive decision occurs when the positive detection of the classifier coincided with a positive detection of the physician.

The classification accuracies (specificity, sensitivity, total classification accuracy) on the test sets are presented in Table 3. As it is seen from Table 3, the combined neural network classified healthy segments, seizure free epileptogenic zone segments, epileptic seizure segments with the accuracy of 96.00%, 94.50%, 94.00%, respectively. The healthy segments, seizure free epileptogenic zone segments, epileptic seizure segments were classified with the accuracy of 94.83%. The total classification accuracy of the stand-alone MLPNN (trained with the backpropagation algorithm, wavelet coefficients used as inputs) was 84.83%. Thus, the accuracy rates of the combined neural network structure presented for this application were found to be higher than that of the stand-alone MLPNN. This may be attributed to several factors including the training algorithms, estimation of the network parameters and the scattered and mixed nature of the features. The behavior of each classifier provides valuable insights to the properties of the feature space and from these insights it may be possible to implement a classification model that will give perfect classification results on the EEG signals.

Table 4
Total classification accuracy obtained for each wavelet when the EEG segments were classified using the combined neural network

Wavelet type	Total classification accuracy
sym6	90.17%
sym10	90.67%
coif4	91.33%
coif2	91.83%
db1	93.67%
db6	94.33%
db2	94.83%

The classification accuracy which is defined as the percentage ratio of the number of segments correctly classified to the total number of segments considered for classification depends on the type of wavelet chosen for the application. Daubechies wavelet of order 2 (db2) was used and found to yield good results in classification of the EEG segments. Since there is no analytical method determining the best wavelet for a particular data type, the optimal wavelet should be determined experimentally. In order to investigate the effect of other wavelets on classifications accuracy, tests were carried out using other wavelets also. Apart from db2, Symmlet of order 6 (sym6), Symmlet of order 10 (sym10), Coiflet of order 2 (coif2), Coiflet of order 4 (coif4), Daubechies of order 1 (db1), Daubechies of order 6 (db6) were also tried. Total classification accuracy obtained for each wavelet when the EEG segments were classified using the combined neural network, is presented in Table 4. Since no detection parameters other than the wavelet type changed, the results give a good indication on the performance of the wavelet function for the signals under study. It can be seen that the Daubechies wavelet offers better accuracy than the others, and db2 is marginally better than db1 and db6. Hence db2 wavelet was chosen for this application.

6. Conclusions

This paper presented the use of neural networks to combine the predictions of an ensemble neural networks for classification of the EEG signals. Toward achieving classification of the EEG signals, three sets of neural networks were trained. Networks in each group were trained by the Levenberg–Marquardt algorithm with different targets. The learning targets were modified so that the trained networks would predict one particular EEG signal with higher accuracy than the other types of EEG signals. Improvement in accuracy was obtained by training new neural networks to combine the predictions of the original networks. The combined neural network used for classification of the EEG signals was trained, cross validated and tested with the extracted features using discrete wavelet transform of the EEG signals. The accuracy rates achieved by the combined neural network model presented for classification of the EEG signals were found to be higher than that of the stand-alone MLPNN trained with the backpropagation algorithm (wavelet coefficients used as inputs).

References

[1] 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.
[2] E.D. Übeyli, Analysis of EEG signals using Lyapunov exponents, Neural Network World 16 (3) (2006) 257–273.
[3] A. Saastamoinen, T. Pietila, A. Varri, M. Lehtokangas, J. Saarinen, Waveform detection with RBF network—Application to automated EEG analysis, Neurocomputing 20 (1–3) (1998) 1–13.
[4] E.D. Übeyli, Analysis of EEG signals by combining eigenvector methods and multiclass support vector machines, Comput. Biol. Med. 38 (1) (2008) 14–22.
[5] E.D. Übeyli, Fuzzy similarity index employing Lyapunov exponents for discrimination of EEG signals, Neural Network World 16 (5) (2006) 421–431.
[6] W.R.S. Webber, B. Litt, R.P. Lesser, R.S. Fisher, I. Bankman, Automatic EEG spike detection: What should the computer imitate? Electroencephalogr. Clin. Neurophysiol. 87 (6) (1993) 364–373.
[7] A.J. Gabor, M. Seyal, Automated interictal EEG spike detection using artificial neural networks, Electroencephalogr. Clin. Neurophysiol. 83 (5) (1992) 271–280.
[8] D. West, V. West, Model selection for a medical diagnostic decision support system: A breast cancer detection case, Artif. Intell. Med. 20 (3) (2000) 183–204.
[9] D. West, V. West, Improving diagnostic accuracy using a hierarchical neural network to model decision subtasks, Int. J. Med. Inform. 57 (1) (2000) 41–55.
[10] H. Kordylewski, D. Graupe, K. Liu, A novel large-memory neural network as an aid in medical diagnosis applications, IEEE Trans. Inform. Technol. Biomed. 5 (3) (2001) 202–209.
[11] D.H. Wolpert, Stacked generalization, Neural Networks 5 (1992) 241–259.
[12] M. Taniguchi, V. Tresp, Combining regularized neural networks, in: Proceedings of the ICANN'97, in: Lecture Notes in Comput. Sci., vol. 1327, Springer, Berlin, 1997, pp. 349–354.
[13] E.D. Übeyli, İ. Güler, Improving medical diagnostic accuracy of ultrasound Doppler signals by combining neural network models, Comput. Biol. Med. 35 (6) (2005) 533–554.
[14] İ. Güler, E.D. Übeyli, ECG beat classifier designed by combined neural network model, Pattern Recogn. 38 (2) (2005) 199–208.
[15] I. Daubechies, The wavelet transform, time–frequency localization and signal analysis, IEEE Trans. Inform. Theory 36 (5) (1990) 961–1005.
[16] S. Soltani, On the use of the wavelet decomposition for time series prediction, Neurocomputing 48 (2002) 267–277.
[17] M. Akay, Wavelet applications in medicine, IEEE Spectrum 34 (5) (1997) 50–56.
[18] M. Unser, A. Aldroubi, A review of wavelets in biomedical applications, Proc. IEEE 84 (4) (1996) 626–638.
[19] M. Misiti, Y. Misiti, G. Oppenheim, J.-M. Poggi, Wavelet Toolbox, The MathWorks Inc., 1997.
[20] K.N. Le, K.P. Dabke, G.K. Egan, Hyperbolic wavelet power spectra of nonstationary signals, Opt. Eng. 42 (10) (2003) 3017–3037.

[21] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C.E. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, Phys. Rev. E 64 (2001) 061907.
[22] I.A. Basheer, M. Hajmeer, Artificial neural networks: Fundamentals, computing, design, and application, J. Microbiol. Methods 43 (1) (2000) 3–31.
[23] B.B. Chaudhuri, U. Bhattacharya, Efficient training and improved performance of multilayer perceptron in pattern classification, Neurocomputing 34 (2000) 11–27.
[24] S. Haykin, Neural Networks: A Comprehensive Foundation, Macmillan, New York, 1994.
[25] V. Hansen, R.D. Nelson, Data mining of time series using stacked generalizers, Neurocomputing 43 (2002) 173–184.
[26] K.M. Ting, I.H. Witten, Issues in stacked generalization, J. Artif. Intell. Res. 10 (1999) 271–289.
[27] D.V. Sridhar, E.B. Bartlett, R.C. Seagrave, An information theoretic approach for combining neural network process models, Neural Networks 12 (6) (1999) 915–926.
[28] S. Piramuthu, H. Ragavan, M.J. Shaw, Using feature construction to improve the performance of neural networks, Manage. Sci. 44 (3) (1998) 416–430.
[29] M.T. Hagan, M.B. Menhaj, Training feedforward networks with the Marquardt algorithm, IEEE Trans. Neural Networks 5 (6) (1994) 989–993.
[30] R. Battiti, First- and second-order methods for learning: Between steepest descent and Newton's method, Neural Comput. 4 (1992) 141–166.
[31] R. Fletcher, Practical Methods of Optimization, Wiley, New York, 1987.
[32] J.J. More, The Levenberg–Marquardt Algorithm: Implementation and Theory, Springer, Berlin, 1978.

Elif Derya Übeyli (http://edubeyli.etu.edu.tr/) is an Associate Professor at the Department of Electrical and Electronics Engineering, TOBB University of Economics and Technology. She obtained Ph.D. degree in Electronics and Computer Technology from the Gazi University in 2004. She has worked on variety of topics including biomedical signal processing, neural networks, optimization and artificial intelligence. She has worked on several projects related with biomedical signal acquisition, processing and classification. Dr. Übeyli has served (or is currently serving) as a program organizing committee member of the national and international conferences. She is editorial board member of several scientific journals (Journal of Engineering and Applied Sciences; International Journal of Soft Computing; Research Journal of Applied Sciences; Research Journal of Medical Sciences; Scientific Journals International/Electrical, Mechanical, Manufacturing, and Aerospace Engineering; The Open Medical Informatics Journal; Bulletin of the International Scientific Surgical Association; Nano Science & Nano Technology: An Indian Journal). She is Associate Editor of Expert Systems. She is serving as a guest editor to the Expert Systems on a special issue on "Advances in Medical Decision Support Systems." Moreover, she is voluntarily serving as a technical publication reviewer for many respected scientific journals and conferences. She has also published 101 journal and 38 conference papers on her research areas.