Computers in Biology and Medicine 38 (2008) 401 – 410

Computers in Biology and Medicine

www.intl.elsevierhealth.com/journals/cobm

Recurrent neural networks with composite features for detection of electrocardiographic changes in partial epileptic patients

Elif Derya Übeyli*

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

Received 20 September 2006; accepted 7 January 2008

Abstract

The aim of this study is to evaluate the diagnostic accuracy of the recurrent neural networks (RNNs) with composite features (wavelet coefficients and Lyapunov exponents) on the electrocardiogram (ECG) signals. Two types of ECG beats (normal and partial epilepsy) were obtained from the MIT-BIH database. The multilayer perceptron neural networks (MLPNNs) were also tested and benchmarked for their performance on the classification of the ECG signals. Decision making was performed in two stages: computing composite features which were then input into the classifiers and classification using the classifiers trained with the Levenberg–Marquardt algorithm. The research demonstrated that the wavelet coefficients and the Lyapunov exponents are the features which well represent the ECG signals and the RNN trained on these features achieved high classification accuracies.

Keywords: Recurrent neural networks (RNNs); Levenberg–Marquardt algorithm; Composite features; Wavelet coefficients; Lyapunov exponents; Electrocardiogram (ECG) signals

1. Introduction

Epileptic seizures are associated with several changes in autonomic functions, which may lead to cardiovascular, respiratory, gastrointestinal, cutaneous, and urinary manifestations [1–5]. Cardiovascular changes have received the most attention, because of their possible contribution to sudden unexplained death. Sudden unexplained death incidence varies from 1/100 patient years in patients with severe intractable epilepsy to 1/1000 patient years in patients with well-controlled epilepsy [6–10]. Studies have reported the importance of monitoring the electrocardiogram (ECG) signal during epileptic seizures, since the seizures can trigger high risk cardiac arrhythmias [1–5]. The ECG measures the electrical activity of the heart, and its morphology and timing provide information about the state of cardiac health [11–18]. Since seizures can occur at any time in an epileptic patient, the ECG may need to be recorded for several hours or days at a time, leading to an enormous quantity of data to be studied by physicians. To reduce the time and possibility of errors, automatic computer-based algorithms have been proposed to support or replace the diagnosis and analysis performed by the physician [11–18]. From the hours of ECG data, these algorithms can flag the periods when the patient is having a seizure and, eventually, determine from these periods if any cardiac arrhythmias occurred. This study provides a highly accurate algorithm for classifying non-arrhythmic ECG waveforms as normal or partial epileptic.

Numerous ECG beats detection algorithms such as derivative-based algorithms, algorithms based on digital filters, wavelet transform (WT), length and energy transform, artificial neural networks (ANNs), genetic algorithms, etc., are reported in the literature [11–18]. While many such algorithms are available, using the time-derivative, digital filters, WT, etc., discrimination of normal from partial epileptic ECG beats has been considered by researchers Übeyli and Güler [14–16]. Übeyli and Güler [14] computed Lyapunov exponents of the ECG signals (the dataset is the same with the dataset used in the present study), which were used as inputs of the multilayer perceptron neural networks (MLPNNs) trained with backpropagation, delta-bar-delta, extended delta-bar-delta, quick propagation, and Levenberg–Marquardt algorithms. The MLPNN

* Tel.: +90 312 2924080; fax: +90 312 2924091.
E-mail address: edubeyli@etu.edu.tr.

0010-4825/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2008.01.002

trained with the Levenberg–Marquardt algorithm classified the normal beats and partial epilepsy beats with the accuracy of 97.50% [14]. Güler and Übeyli [15] presented an approach based on adaptive neuro-fuzzy inference system (ANFIS) for detection of electrocardiographic changes in patients with partial epilepsy (the dataset is the same with the dataset used in the present study). Decision making was performed in two stages: feature extraction using the WT and the ANFIS trained with the backpropagation gradient descent method in combination with the least-squares method. The proposed ANFIS classified normal beats and partial epilepsy beats with the accuracy of 98.13% [15]. Güler and Übeyli [16] used the MLPNNs trained with the backpropagation, delta-bar-delta, extended delta-bar-delta, and quick propagation algorithms for detection of electrocardiographic changes in patients with partial epilepsy (the dataset is the same with the dataset used in the present study). At each level of wavelet decomposition, the absolute value of the detail signals were measured, and the two coefficients with the highest magnitude were retained. Thus, for $M = 4$ (number of decomposition level), there were $4 \times 2 = 8$ coefficients for each segment of the ECG signals which were used as the MLPNN inputs. The MLPNN trained with the quick propagation algorithm classified normal beats and partial epilepsy beats with the accuracy of 96.88% [16]. The present study indicated the usage of recurrent neural networks (RNNs) to improve the classification accuracy of normal and partial epilepsy ECG beats.

Automated ECG classification algorithms can be divided into three steps: pre-processing, feature extraction/selection, and classification. The techniques developed for automated electrocardiographic change detection transform the mostly qualitative diagnostic criteria into a more objective quantitative signal feature classification problem [11–18]. 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. Selection of the neural network inputs has two meanings: (1) which components of a pattern or (2) which set of inputs best represent a given pattern. Two diverse feature vectors were used in discrimination of the ECG signals, i.e., wavelet coefficients and Lyapunov exponents. Therefore, the RNNs employing composite features (wavelet coefficients and Lyapunov exponents) were implemented for automated electrocardiographic changes detection in partial epileptic patients.

The idea of using ANNs for pattern classification purposes has encountered, for a long time, the favor of many researchers [11–18]. Feedforward neural networks are a basic type of neural networks capable of approximating generic classes of functions, including continuous and integrable ones. One of the most frequently used feedforward neural network for pattern classification is the MLPNN which is trained to produce a spatial output pattern in response to an input spatial pattern [19,20]. The mapping performed is static; therefore, the network is inherently not suitable for processing temporal patterns. Attempts have been made to use the MLPNN to classify temporal patterns by transforming the temporal domain into a spatial domain.

An alternate neural network approach is to use RNNs which have memory to encode past history. Several forms of RNNs have been proposed and they may be classified as partially recurrent or fully recurrent networks [21–24]. RNNs can perform highly nonlinear dynamic mappings and thus have temporally extended applications, whereas multilayer feedforward networks are confined to performing static mappings. RNNs have been used in a number of interesting applications including associative memories, spatiotemporal pattern classification, control, optimization, forecasting, and generalization of pattern sequences [21–24]. In partially recurrent networks, partial recurrence is created by feeding back delayed hidden unit outputs or the outputs of the network as additional input units. The partially recurrent networks, whose connections are mainly feedforward, were used, but they include a carefully chosen set of feedback connections. One example of such a network is an Elman RNN which in principle is set up as a regular feedforward network [25].

The WT can be applied to extract the wavelet coefficients of discrete time signals. 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 [26]. Thus, the ECG signal, consisting of many data points, can be compressed into a few parameters. These parameters characterize the behavior of the ECG signal [12,15–17].

In recent years, there has been an increasing interest in applying techniques from the domains of nonlinear analysis and chaos theory in studying the behavior of a dynamical system from an experimental time series such as ECG signals [14,18,27,28]. The purpose of these studies is to determine whether dynamical measures especially Lyapunov exponents can serve as clinically useful parameters. Estimation of the Lyapunov exponents is more readily interpreted with respect to the presence of chaos. The positive Lyapunov exponents are the hallmark of chaos [28].

The evaluation of the classification capabilities of the Elman RNNs trained with Levenberg–Marquardt algorithm was performed on the ECG signals (normal and partial epilepsy ECG beats) from the MIT-BIH database [29]. As in traditional pattern recognition systems, the model consists of three main modules: a feature extractor that generates a feature vector from the ECG signals, feature selection that composes composite features (wavelet coefficients and Lyapunov exponents), and a feature classifier that outputs the class based on the composite features. A significant contribution of the work was the composition of composite features which were used to train novel classifier (RNN trained on composite feature) for the ECG signals. To evaluate performance of the classifiers, the classification accuracies, the central processing unit (CPU) times of training, and the receiver operating characteristic (ROC) curves of the classifiers were examined. The results demonstrated that significant improvement can be achieved in accuracy by using the RNNs compared to the feedforward neural network models (MLPNNs).

The outline of this study is as follows. In Section 2, a brief description of the used data is presented. In Section 3, the computation of the composite features (wavelet coefficients and Lyapunov exponents) of the ECG signals is explained in order to define the behavior of the signal under study. In Section 4, description of neural network models including RNNs that are considered in this study is presented. The Levenberg–Marquardt algorithm used for training the RNNs is also explained. In Section 5, the results of application of the RNNs to the ECG signals are presented. Finally, in Section 6 the study is concluded.

2. Data description

MIT-BIH database [29] is a large and growing archive of well-characterized digital recordings of physiologic signals and related data for use by the biomedical research community. MIT-BIH currently includes databases of multi-parameter cardiopulmonary, neural, and other biomedical signals from healthy subjects and patients with a variety of conditions with major public health implications, including sudden cardiac death, congestive heart failure, epilepsy, gait disorders, sleep apnea, and aging. The database of heart rate oscillations in partial epilepsy was studied in the present work. Two types (normal and partial epilepsy) of the ECG beats (180 records from each class) were obtained from the MIT-BIH database. Fig. 1 shows the waveforms of the two exemplary records of normal and partial epilepsy ECG beats (first records from two classes) from the MIT-BIH database [29].

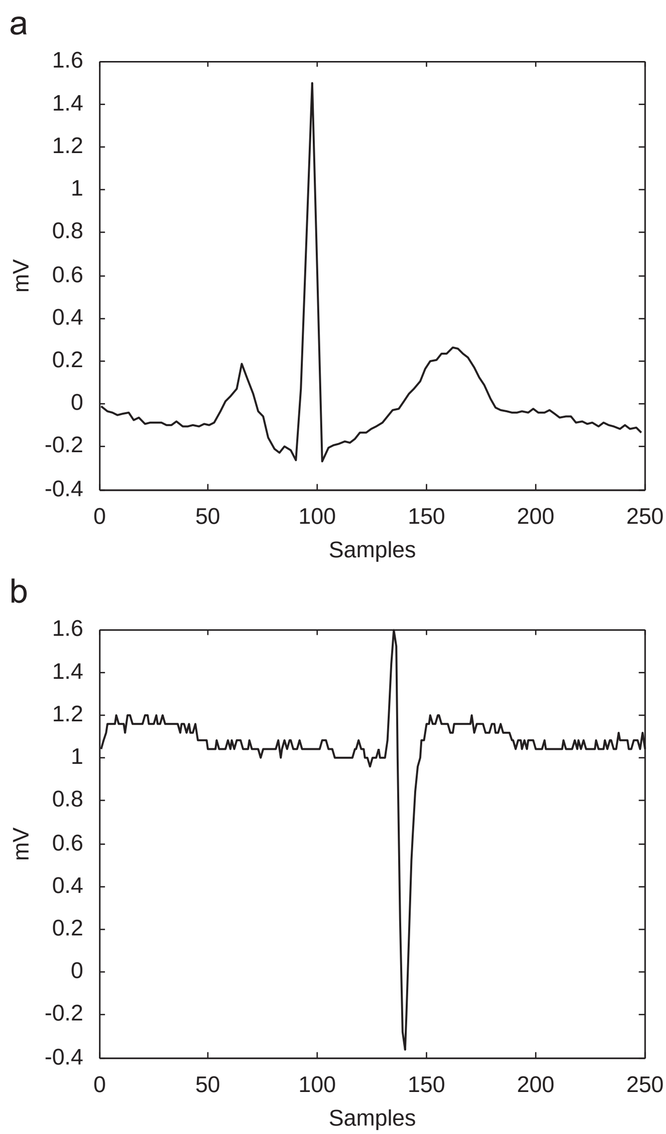

Fig. 1. Waveforms of the ECG beats: (a) normal beat (record 1) and (b) partial epilepsy beat (record 1).

Post-ictal heart rate oscillations were reported in a heterogeneous group of patients with partial epilepsy. This pattern is marked by the appearance of transient but prominent low-frequency heart rate oscillations (0.01–0.1 Hz) immediately following 5 of 11 seizures recorded in five patients. This finding may be a marker of neuroautonomic instability, and, therefore, may have implications for understanding perturbations of heart rate control associated with partial seizures. The preliminary report was based upon analysis of data from 11 partial seizures recorded in five women patients during continuous electroencephalographic/electrocardiographic/video monitoring. The patients ranged in age from 31 to 48 years old, were without clinical evidence of cardiac disease, and had partial seizures with or without secondary generalization from frontal or temporal foci. Recordings were made under a protocol approved by Beth Israel Deaconess Medical Center's Committee on Clinical Investigations [29].

Data were analyzed offline using customized software. Onset and offset of seizures were visually identified to the nearest 0.1 s by an experienced electroencephalographer (DLS) blinded with respect to the heart rate variability analysis. Continuous single-lead ECG signals were sampled at 200 Hz. From the digitized ECG recording, a heartbeat annotation file (a list of the type and time of occurrence of each heartbeat) was obtained using a version of the commercially available arrhythmia analysis software [29].

3. Computation of composite features

3.1. Computation of wavelet coefficients

The multi-scale feature of the WT allows the decomposition of a signal into a number of scales, each scale representing a particular coarseness of the signal under study. 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.

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

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

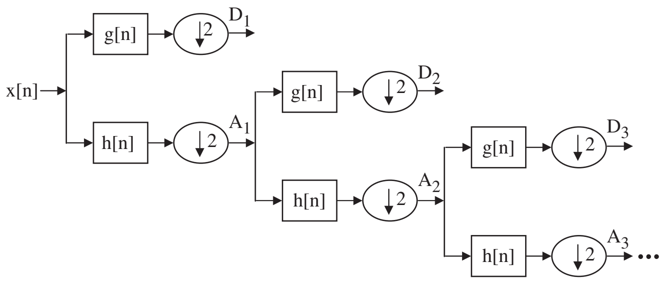

Fig. 2. Subband decomposition of discrete wavelet transform implementation; $g[n]$ is the high-pass filter, $h[n]$ is the low-pass filter.

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

$$G(z) = zH(-z^{-1}). \tag{2}$$

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

$$H_{i+1}(z) = H(z^{2^i})H_i(z),$$
$$G_{i+1}(z) = G(z^{2^i})H_i(z), \quad i = 0, \ldots, I-1, \tag{3}$$

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

$$h_{i+1}(k) = [h]_{\uparrow 2^i} * h_i(k),$$
$$g_{i+1}(k) = [g]_{\uparrow 2^i} * h_i(k), \tag{4}$$

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 $\varphi_{i,l}(k)$ and $\psi_{i,l}(k)$ can be defined as

$$\varphi_{i,l}(k) = 2^{i/2}h_i(k - 2^i l),$$
$$\psi_{i,l}(k) = 2^{i/2}g_i(k - 2^i l), \tag{5}$$

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

$$a_{(i)}(l) = x(k) * \varphi_{i,l}(k),$$
$$d_{(i)}(l) = x(k) * \psi_{i,l}(k), \tag{6}$$

where $a_{(i)}(l)$ and $d_i(l)$ are the approximation coefficients and the detail coefficients at resolution i, respectively [26].

The ECG 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. Selection of appropriate wavelet and the number of decomposition levels are 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 ECG 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 1 (db1) made it more suitable to detect changes of the ECG signals. Therefore, the wavelet coefficients were computed using the db1 in the present study. A rectangular window, which was formed by 256 discrete data, was selected so that it contained a single ECG beat. This size is long enough to cover most QRS signals. For each ECG beat, the detail wavelet coefficients (d^k, $k = 1, 2, 3, 4$) at the first, second, third, and fourth levels ($128 + 64 + 32 + 16$ coefficients) and the approximation wavelet coefficients (a^4) at the fourth level (16 coefficients) were computed. Then 256 wavelet coefficients were obtained for each ECG beat. The wavelet coefficients of the ECG signals formed the first diverse feature vector representing the ECG signals.

3.2. Computation of Lyapunov exponents

Consider two (usually the nearest) neighboring points in phase space at time 0 and at time t, distances of the points in the ith direction being $\|\delta x_i(0)\|$ and $\|\delta x_i(t)\|$, respectively. The Lyapunov exponent is then defined by the average growth rate λ_i of the initial distance:

$$\frac{\|\delta x_i(t)\|}{\|\delta x_i(0)\|} = 2^{\lambda_i t} \quad (t \to \infty)$$

or

$$\lambda_i = \lim_{t\to\infty} \frac{1}{t} \log_2 \frac{\|\delta x_i(t)\|}{\|\delta x_i(0)\|}. \tag{7}$$

The existence of a positive Lyapunov exponent indicates chaos [30,31]. This shows that any neighboring points with infinitesimal differences at the initial state abruptly separate from each other in the ith direction. In other words, even if the initial states are close, the final states are much different. This phenomenon is sometimes called sensitive dependence on initial conditions. Numerous methods for calculating the Lyapunov exponents have been developed during the past decade. Generally, the Lyapunov exponents can be estimated either from the equations of motion of the dynamic system (if it is known) or from the observed time series. The latter is what is of interest due to its direct relation to the work in this paper. The idea is based on the well-known technique of state space reconstruction with delay coordinates to build a system with Lyapunov exponents identical to that of the original system from which the present measurements have been observed. Generally, Lyapunov exponents can be extracted from observed signals in two different ways. The first is based on the idea of following the time-evolution of nearby points in the state space. This method provides an estimation of the largest Lyapunov exponent only. The second method is based on the estimation of local Jacobi matrices and is capable of estimating all the Lyapunov exponents. Vectors of all the Lyapunov exponents for particular systems are often called their Lyapunov spectra [14,30,31].

In the present study, the technique used in the computation of Lyapunov exponents was related with the Jacobi-based algorithms. Dynamical measures especially Lyapunov exponents can serve as clinically useful parameters and contain a significant amount of information about the signal. The calculations indicated that there are positive Lyapunov exponents, which confirm the chaotic nature of the ECG signals. After the required computations, 128 Lyapunov exponents were obtained for each ECG beat. The Lyapunov exponents of the ECG signals composed the second diverse feature vector representing the ECG signals.

4. Description of neural network models

4.1. Recurrent neural networks

Elman RNN was used in the applications and therefore in the following the Elman RNN is presented. An Elman RNN is a network which in principle is set up as a regular feedforward network. This means that all neurons in one layer are connected with all neurons in the next layer. An exception is the so-called context layer which is a special case of a hidden layer. Fig. 3 shows the architecture of an Elman RNN. The neurons in the context layer (context neurons) hold a copy of the output of the hidden neurons. The output of each hidden neuron is copied into a specific neuron in the context layer. The value of the context neuron is used as an extra input signal for all the neurons in the hidden layer one time step later. Therefore, the Elman network has an explicit memory of one time lag [25].

Similar to a regular feedforward neural network, the strength of all connections between neurons is indicated with a weight. Initially, all weight values are chosen randomly and are optimized during the stage of training. In an Elman network, the weights from the hidden layer to the context layer are set to one and are fixed because the values of the context neurons have to be copied exactly. Furthermore, the initial output weights

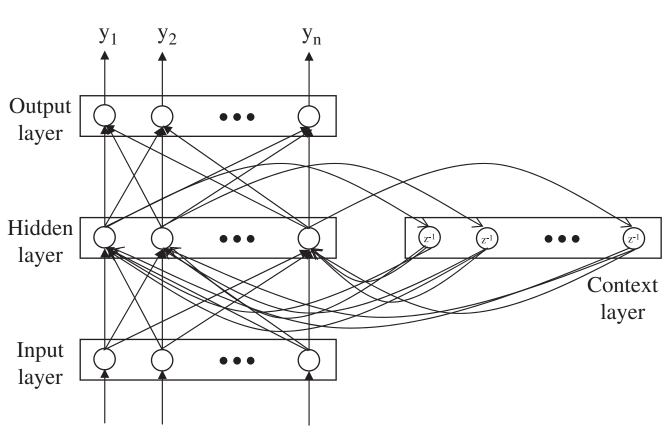

Fig. 3. A schematic representation of an Elman recurrent neural network. z^{-1} represents a one time step delay unit.

of the context neurons are equal to half the output range of the other neurons in the network. The Elman network can be trained with gradient descent backpropagation and optimization methods, similar to regular feedforward neural networks [32]. The 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 [19,20]. 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) have also been used for neural networks 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 [33,34] and can yield a good cost function compared with the other training algorithms.

4.2. 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:

$$E(\mathbf{w}) = \sum_{i=1}^{m} e_i^2(\mathbf{w}) = \| f(\mathbf{w})\|^2, \tag{8}$$

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:

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \delta\mathbf{w}_k, \tag{9}$$

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

$$\delta\mathbf{w}_k = -(J_k^{\mathrm{T}} f(\mathbf{w}_k))(J_k^{\mathrm{T}} J_k + \lambda\mathbf{I})^{-1}. \tag{10}$$

In Eq. (10), J_k is the Jacobian of f evaluated at $\mathbf{w}_k$, λ is the Marquardt parameter, $\mathbf{I}$ is the identity matrix [33,34]. 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. (10) 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. Results and discussion

5.1. Analysis of composite features

The detail wavelet coefficients at the first decomposition level of the two types of ECG beats are presented in Figs. 4(a) and (b), respectively. From these figures it is obvious that the detail wavelet coefficients of the two types of ECG beats are different from each other and therefore they can serve as useful parameters in discriminating the ECG signals. A smaller number of parameters called wavelet coefficients are obtained by the WT. These coefficients represent the ECG signals and therefore they are particularly important for recognition and diagnostic purposes. The Lyapunov exponents of the two types of ECG beats are shown in Figs. 5(a) and (b), respectively. One can see that the Lyapunov exponents of the two types of ECG beats differ significantly from each other so they can be used for representing the ECG signals. As it is seen from Figs. 5(a) and (b), there are positive Lyapunov exponents, which confirm the chaotic nature of the ECG signals. Lyapunov exponents are a quantitative measure for distinguishing among the various types of orbits based upon their sensitive dependence on the initial conditions, and are used to determine the stability of any steady-state behavior, including chaotic solutions. The reason why chaotic systems show aperiodic dynamics is that phase space trajectories that have nearly identical initial states will separate from each other at an exponentially increasing rate captured by the so-called Lyapunov exponent.

ANNs may offer a potentially superior method of biomedical signal analysis to the spectral analysis methods. 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 are

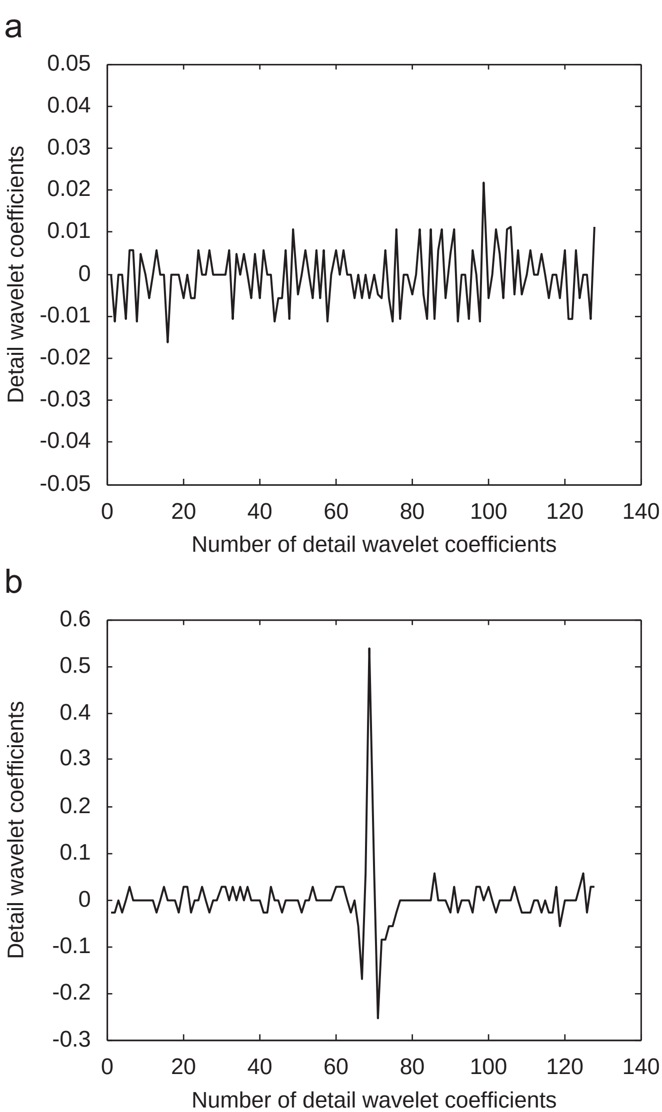

Fig. 4. The detail wavelet coefficients at the first decomposition level of the ECG beats: (a) normal beat and (b) partial epilepsy beat.

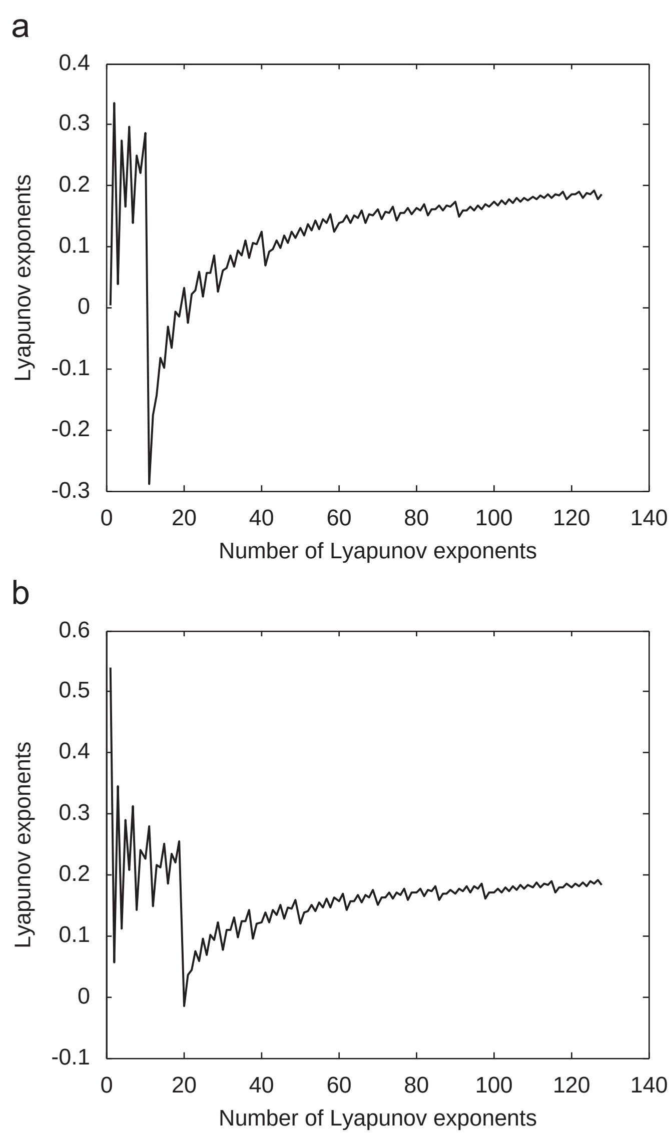

Fig. 5. Lyapunov exponents of the ECG beats: (a) normal beat and (b) partial epilepsy beat.

Table 1
The mean values of extracted features of all ECG records from two classes

ECG beat types	Extracted features	Wavelet coefficients					Lyapunov exponents
		Subbands					
		D_1	D_2	D_3	D_4	A_4	
Normal beat	Maximum	0.0225	0.0199	0.0376	0.0512	1.2031	0.3364
	Minimum	−0.0168	−0.0279	−0.0606	−0.0488	0.7576	−0.2870
	Mean	−0.0008	−0.0007	−0.0011	−0.0012	0.9521	0.1276
	Standard deviation	0.0062	0.0098	0.0164	0.0318	0.1378	0.0865
Partial epilepsy beat	Maximum	0.5379	0.9408	0.0841	2.0308	4.6808	0.5406
	Minimum	−0.2541	−0.6008	−0.5946	−0.2205	2.8906	−0.0144
	Mean	−0.0009	0.0006	−0.0366	0.1282	4.2431	0.1658
	Standard deviation	0.0601	0.1472	0.1215	0.5126	0.4092	0.0596

All standard deviations are less than 0.0021.

Table 2
The explanations of the diverse feature vectors

Feature vectors	Explanations
First diverse feature vector	20 wavelet coefficients of each beat (maximum of the wavelet coefficients in each subband, minimum of the wavelet coefficients in each subband, mean of the wavelet coefficients in each subband, standard deviation of the wavelet coefficients in each subband)
Second diverse feature vector	4 values of Lyapunov exponents of each beat (maximum of the Lyapunov exponents in each beat, minimum of the Lyapunov exponents in each beat, mean of the Lyapunov exponents in each beat, standard deviation of the Lyapunov exponents in each beat)

Table 3
The results of network architecture studies

Network architectures	Total classification accuracy (%)
RNNs	
$24 \times 30r \times 20r \times 10r \times 2^{a}$	94.44
$24 \times 20r \times 20r \times 2^{b}$	95.56
$24 \times 20r \times 15r \times 2^{b}$	95.56
$24 \times 30r \times 2^{c}$	96.11
$24 \times 25r \times 2^{c}$	97.22
$24 \times 20r \times 2^{c}$	98.33
MLPNNs	
$24 \times 30 \times 20 \times 10 \times 2^{a}$	88.89
$24 \times 20 \times 20 \times 2^{b}$	90.00
$24 \times 20 \times 15 \times 2^{b}$	90.56
$24 \times 30 \times 2^{c}$	91.67
$24 \times 20 \times 2^{c}$	92.22
$24 \times 25 \times 2^{c}$	92.78

"r" stands for a recurrent neuron.
[a]Design of RNNs and MLPNNs: number of input neurons × neurons in the first hidden layer × neurons in the second hidden layer × neurons in the third hidden layer × output neurons, respectively.
[b]Design of RNNs and MLPNNs: number of input neurons × neurons in the first hidden layer × neurons in the second hidden layer × output neurons, respectively.
[c]Design of RNNs and MLPNNs: number of input neurons × neurons in the first hidden layer × output neurons, respectively.

rapid. Therefore, the RNNs were trained on well-representing composite features (wavelet coefficients and Lyapunov exponents) for classification of the ECG signals.

Feature selection plays an important role in classifying systems such as neural networks. The feature selection process is performed on a set of predetermined features. Features are selected based on either (1) best representation of a given class of signals or (2) best distinction between classes. High-dimension of feature vectors increased computational complexity and the neural networks trained on these feature vectors produced lower accuracy. In order to reduce the dimensionality of the extracted diverse feature vectors, statistics over the set of the wavelet coefficients and Lyapunov exponents were used. The following statistical features were used in reducing the dimensionality of the extracted diverse feature vectors representing the ECG signals:

1. Maximum of the wavelet coefficients in each subband, maximum of the Lyapunov exponents in each beat.
2. Minimum of the wavelet coefficients in each subband, minimum of the Lyapunov exponents in each beat.
3. Mean of the wavelet coefficients in each subband, mean of the Lyapunov exponents in each beat.
4. Standard deviation of the wavelet coefficients in each subband, standard deviation of the Lyapunov exponents in each beat.

Table 1 presents mean values of the extracted features of all ECG records from two classes. From Table 1, one can see that the extracted diverse features of the two classes of ECG beats are different from each other. This result indicated that they can serve as useful parameters in classifying the ECG signals. Table 2 is a summary of the explanations of the diverse feature

Table 4
The values of the statistical parameters and the CPU times of training of the classifiers

Classifiers	Statistical parameters (%)			CPU time (min:s)
	Specificity	Sensitivity	Total classification accuracy	
RNN	98.89	97.78	98.33	12:34
MLPNN	92.22	93.33	92.78	17:05

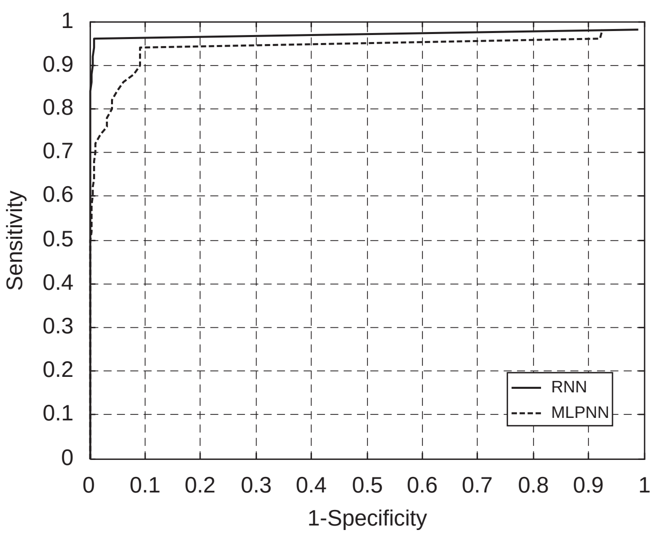

Fig. 6. ROC curves of the classifiers.

vectors. The diverse feature vectors were computed by the usage of the MATLAB software package.

5.2. Implementation of RNNs

The Elman RNNs proposed for classification of the ECG signals were implemented by using the MATLAB software package (MATLAB version 7.0 with neural networks toolbox). The RNN architecture used for classification of the ECG signals is shown in Fig. 3. The key design decisions for the neural networks used in classification are the architecture and the training process. ANN architectures are derived by trial and error and the complexity of the neural network is characterized by the number of hidden layers. There is no general rule for selection of appropriate number of hidden layers. The most popular approach to finding the optimal number of hidden layers is by trial and error. Different network architectures were tested and the architecture studies confirmed that for the ECG signals, RNN with one hidden layer consisting of 20 recurrent neurons trained on a composite feature vector (24 inputs) results in higher classification accuracy. Apart from one hidden layered network architecture, although extensively with larger networks with two and three hidden layers were tested (the RNNs used had the $24 \times 30r \times 20r \times 10r \times 2$ and $24 \times 20r \times 20r \times 2$ architecture, where "r" stands for a recurrent neuron), a significant improvement in training and testing performance cannot be obtained. The number of output was 2 and samples with target outputs normal beat and partial epilepsy beat were given the binary target values of (0,1), (1,0), respectively. In this application, in the hidden layers and the output layer sigmoidal function was used as activation function. The sigmoidal function with the range between 0 and 1 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.

In order to compare performance of the different classifiers, for the same classification problem the MLPNN, which is the most commonly used feedforward neural networks, was also implemented. The single hidden layered (25 hidden neurons) MLPNN was used to classify the ECG signals based on a composite feature vector (24 inputs). In the hidden layers and the output layers, the activation function was the sigmoidal function. Different experiments were performed during implementation of the classifiers, and the number of hidden neurons was determined by taking into consideration the classification accuracies. The results of architecture studies of the implemented RNNs and the MLPNNs are presented in Table 3.

The adequate functioning of neural networks depends on the sizes of the training set and test set. In this study, training and test sets were formed by 360 vectors (180 vectors from each class) of 24 dimensions (dimension of the extracted feature vectors). The 180 vectors (90 vectors from each class) of 24 dimensions were used for training and the 180 vectors (90 vectors from each class) of 24 dimensions were used for testing. A practical way to find a point of better generalization is to use a small percentage (around 20%) of the training set for cross-validation. For obtaining a better network generalization 36 vectors (18 vectors from each class) of training set, which were selected randomly, were used as cross validation set.

The training holds the key to an accurate solution, so the criterion to stop training must be very well described. 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. Training of the RNN was performed with the Levenberg–Marquardt algorithm in 350 epochs. The MLPNN trained with the Levenberg–Marquardt algorithm had a slow convergence and mean square error (MSE) converged to a small constant of approximately zero in 1200 epochs. Thus, the convergence rate of the RNN presented in this study was found to be higher than that of the MLPNN.

5.3. Classification errors and ROC analysis

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 correct classified normal beats/number of total normal beats;

sensitivity: number of correct classified partial epilepsy beats/number of total partial epilepsy beats;

total classification accuracy: number of correct classified beats/number of total beats.

The values of the statistical parameters (specificity, sensitivity, and total classification accuracy) and the CPU times of training (for Pentium 4, 3.00 GHz) of the two classifiers are presented in Table 4. ROC plots provide a view of the whole spectrum of sensitivities and specificities because all possible sensitivity/specificity pairs for a particular test are graphed. The performance of a test can be evaluated by plotting an ROC curve for the test, and therefore ROC curves were used to describe the performance of the classifiers. A good test is one for which sensitivity rises rapidly and 1-specificity hardly increases at all until sensitivity becomes high. ROC curves, which are shown in Fig. 6, demonstrate the performances of the classifiers on the test files. The classification results presented in Table 4 and Fig. 6 (classification accuracies, CPU times of training, ROC curves) denote that the RNN trained on composite feature vectors obtains higher accuracy than that of the MLPNN.

6. Conclusions

The RNNs trained on composite features (wavelet coefficients and Lyapunov exponents) were applied to the ECG beats classification problem. The RNNs were chosen because they can implement extremely nonlinear decision boundaries and possess memory of the state, which is crucial for the considered task. The wavelet coefficients and dynamical measures especially Lyapunov exponents can serve as clinically useful parameters and contain a significant amount of information about the signal. Therefore, the wavelet coefficients and the Lyapunov exponents become a natural complement to the applications of the RNNs. In order to evaluate the used classifiers, the classification accuracies, the CPU times of training, and ROC curves of the classifiers were considered. The overall results of the RNN were better when they were trained on the computed composite features for each ECG beat. The performance of the MLPNN was not as high as the RNN. 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 results of the present study demonstrated that the RNN can be used in classification of the ECG beats by taking into consideration the misclassification rates.

Conflict of interest statement

None declared.

References

[1] R.C. Frysinger, J. Engel, R.M. Harper, Interictal heart rate patterns in partial seizure disorders, Neurology 43 (1993) 2136–2139.
[2] T. Tomson, M. Ericson, C. Ihrman, L.E. Lindblad, Heart rate variability in patients with epilepsy, Epilepsy Res. 30 (1998) 77–83.
[3] F. Leutmezer, C. Schernthaner, S. Lurger, K. Pötzelberger, C. Baumgartner, Electrocardiographic changes at the onset of epileptic seizures, Epilepsia 44 (3) (2003) 348–354.
[4] M. Zijlmans, D. Flanagan, J. Gotman, Heart rate changes and ECG abnormalities during epileptic seizures: prevalence and definition of an objective clinical sign, Epilepsia 43 (8) (2002) 847–854.
[5] R. Rocamora, M. Kurthen, L. Lickfett, J. von Oertzen, C.E. Elger, Cardiac asystole in epilepsy: clinical and neurophysiologic features, Epilepsia 44 (2) (2003) 179–185.
[6] G. Kenneback, M. Ericson, T. Tomson, L. Bergfeldt, Changes in arrythmia profile and heart rate variability during abrupt withdrawal of antiepileptic drugs. Implications for sudden death, Seizure 6 (5) (1997) 369–375.
[7] J.E. Leestma, J.F. Annegers, M.J. Brodie, S. Brown, P. Schraeder, D. Siscovich, B.B. Wannamaker, P.S. Tennis, M.A. Cierpiel, N.L. Earl, Sudden unexplained death in epilepsy: observations from a large clinical development program, Epilepsia 38 (1) (1997) 47–55.
[8] J.E. Leestma, M. Kalelkar, S. Teas, G. Jay, J. Hughes, Sudden unexpected death associated with seizures: analysis of 66 cases, Epilepsia 25 (1984) 84–88.
[9] T.S. Walczak, I.E. Leppik, M. D'Amelio, J. Rarick, E. So, P. Ahman, K. Ruggles, C.D. Cascino, J.F. Annegers, W.A. Hauser, Incidence and risk factors in sudden unexpected death in epilepsy, Neurology 56 (2001) 519–525.
[10] C. Stollberger, J. Finsterer, Cardiorespiratory findings in sudden unexplained/unexpected death in epilepsy (SUDEP), Epilepsy Res. 59 (1) (2004) 51–60.
[11] A.S. Miller, B.H. Blott, T.K. Hames, Review of neural network applications in medical imaging and signal processing, Med. Biol. Eng. Comput. 30 (1992) 449–464.
[12] S.C. Saxena, V. Kumar, S.T. Hamde, Feature extraction from ECG signals using wavelet transforms for disease diagnostics, Int. J. Syst. Sci. 33 (13) (2002) 1073–1085.
[13] S.Y. Foo, G. Stuart, B. Harvey, A. Meyer-Baese, Neural network-based EKG pattern recognition, Eng. Appl. Artif. Intell. 15 (2002) 253–260.
[14] E.D. Übeyli, İ. Güler, Detection of electrocardiographic changes in partial epileptic patients using Lyapunov exponents with multilayer perceptron neural networks, Eng. Appl. Artif. Intell. 17 (6) (2004) 567–576.
[15] İ. Güler, E.D. Übeyli, Application of adaptive neuro-fuzzy inference system for detection of electrocardiographic changes in patients with partial epilepsy using feature extraction, Expert Syst. Appl. 27 (3) (2004) 323–330.
[16] İ. Güler, E.D. Übeyli, An expert system for detection of electrocardiographic changes in patients with partial epilepsy using wavelet-based neural networks, Expert Syst. 22 (2) (2005) 62–71.
[17] K. Sternickel, Automatic pattern recognition in ECG time series, Comput. Methods Prog. Biomed. 68 (2002) 109–115.
[18] M.I. Owis, A.H. Abou-Zied, A.-B.M. Youssef, Y.M. Kadah, Study of features based on nonlinear dynamical modeling in ECG arrhytmia detection and classification, IEEE Trans. Biomed. Eng. 49 (7) (2002) 733–736.
[19] S. Haykin, Neural Networks: A Comprehensive Foundation, Macmillan, New York, 1994.
[20] B.B. Chaudhuri, U. Bhattacharya, Efficient training and improved performance of multilayer perceptron in pattern classification, Neurocomputing 34 (2000) 11–27.
[21] E.D. Übeyli, Recurrent neural networks employing Lyapunov exponents for analysis of Doppler ultrasound signals, Expert Sys. Appl., 2008, in press.
[22] E.W. Saad, D.V. Prokhorov, D.C. Wunsch II, Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks, IEEE Trans. Neural Networks 9 (6) (1998) 1456–1470.

[23] L. Gupta, M. McAvoy, Investigating the prediction capabilities of the simple recurrent neural network on real temporal sequences, Pattern Recognition 33 (12) (2000) 2075–2081.

[24] L. Gupta, M. McAvoy, J. Phegley, Classification of temporal sequences via prediction using the simple recurrent neural network, Pattern Recognition 33 (10) (2000) 1759–1770.

[25] J.L. Elman, Finding structure in time, Cognitive Sci. 14 (2) (1990) 179–211.

[26] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, IEEE Trans. Inform. Theory 36 (5) (1990) 961–1005.

[27] A. Casaleggio, S. Braiotta, Estimation of Lyapunov exponents of ECG time series—the influence of parameters, Chaos Solitons Fractals 8 (10) (1997) 1591–1599.

[28] R. Silipo, G. Deco, R. Vergassola, H. Bartsch, Dynamics extraction in multivariate biomedical time series, Biol. Cybern. 79 (1998) 15–27.

[29] I.C. Al-Aweel, K.B. Krishnamurthy, J.M. Hausdorff, J.E. Mietus, J.R. Ives, A.S. Blum, D.L. Schomer, A.L. Goldberger, Postictal heart rate oscillations in partial epilepsy, Neurology 53 (7) (1999) 1590–1592.

[30] S. Haykin, X.B. Li, Detection of signals in chaos, Proc. IEEE 83 (1) (1995) 95–122.

[31] H.D.I. Abarbanel, R. Brown, M.B. Kennel, Lyapunov exponents in chaotic systems: their importance and their evaluation using observed data, Int. J. Mod. Phys. B 5 (9) (1991) 1347–1375.

[32] F.J. Pineda, Generalization of back-propagation to recurrent neural networks, Phys. Rev. Lett. 59 (19) (1987) 2229–2232.

[33] M.T. Hagan, M.B. Menhaj, Training feedforward networks with the Marquardt algorithm, IEEE Trans. Neural Networks 5 (6) (1994) 989–993.

[34] R. Battiti, First- and second-order methods for learning: between steepest descent and Newton's method, Neural Comput. 4 (1992) 141–166.

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; 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 90 journals and 37 conference papers on her research areas.