ORIGINAL PAPER

Classifying Epilepsy Diseases Using Artificial Neural Networks and Genetic Algorithm

Sabri Koçer · M. Rahmi Canal

Abstract In this study, FFT analysis is applied to the EEG signals of the normal and patient subjects and the obtained FFT coefficients are used as inputs in Artificial Neural Network (ANN). The differences shown by the non-stationary random signals such as EEG signals in cases of health and sickness (epilepsy) were evaluated and tried to be analyzed under computer-supported conditions by using artificial neural networks. Multi-Layer Perceptron (MLP) architecture is used Levenberg-Marquardt (LM), Quickprop (QP), Delta-bar delta (DBD), Momentum and Conjugate gradient (CG) learning algorithms, and the best performance was tried to be attained by ensuring the optimization with the use of genetic algorithms of the weights, learning rates, neuron numbers of hidden layer in the training process. This study shows that the artificial neural network increases the classification performance using genetic algorithm.

Keywords EEG · Multilayer perceptron (MLP) · Genetic algorithm (GA)

Introduction

The EEG is still one of the most main tools to access to one of the most unknown and complex system is nature. There is no doubt that due to its complexity and ability to reflect underlying processes in the brain the EEG signal is theoretically the best physiological signal for extraction and comprehension of human behaviour. EEG involves the recording and analysis of electrical signals generated by the brain [1].

EEG is an important clinical tool for diagnosing and monitoring neurological disorders related to epilepsy. Epilepsy is characterized by sudden recurrent and transient disturbances of mental functions and/or movements of the body that result from excessive discharging of groups of brain cells. The term epilepsy does not refer to a specific disease but rather to a group of symptoms that have many causes. The characteristic activities observed in the scalp EEG of subjects with epilepsy are sharp transient waveforms. Such transient waveforms include spikes and sharp waves [2].

Epilepsy diagnosis is a more complicated problem due to overlapping symptomatology with other neurological disorders, low understanding of the exact mechanism responsible for epilepsy, and lack of knowledge about the possible manners of seizure progression. Usually, confirmation of the diagnosis involves a combination of the medical history of the patient and EEG interpretation by an expert neurologist. The development of accurate and reliable EEG-based automated tools is still in its infancy due to the lack of detectable markers. Although visible EEG markers such as interictal spikes are commonly employed by neurologists to identify epileptic/interictal EEG, they have a subtly variable shape which is not detected consistently by spike detection algorithms reported in the literature [3].

Neurologists usually diagnose the disease of EEG signals based on time domains. EEG signals, however, need to be reviewed in the frequency domain in order to make them more stationary and clear. Recent feature of the computers to record these signals and the development of spectral analysis methods have made it possible to benefit from the frequency elements during the finding of these pathological signals [4].

S. Koçer (✉) · M. R. Canal
Department of Electronics and Computer Education,
Faculty of Technical Education, Gazi University, Teknikokullar,
Ankara, Turkey
e-mail: skocer@gazi.edu.tr

Learning methods such as artificial neural networks are based on learning from examples. The main philosophy of learning from examples is to learn the relationships between the inputs and outputs of the event by using the examples that occurred about the event, and to determine the outputs of new examples to occur depending on these relationships. It is considered here that the relationship between the inputs and the outputs of examples about a particular event contains information that will represent the entirety of the event. It is assumed that different examples represent the events from different perspectives. Thus the event is learned from different perspectives by using different examples. Only the examples are shown to the computer. No other preliminary information are given. The system, which will perform the learning (the artificial neural network in this study) discovers the relationship by using its own algorithm [5–7].

Artificial neural networks (ANNs) have been used as computational tools for pattern classification including diagnosis of diseases because of the belief that they have greater predictive power than signal analysis techniques [8]. Artificial neural Networks (ANN) have been widely used for spike recognition [9–15].

Several neurological disorders are routinely examined by EEG analysis and the differentiation between physiological and pathological alterations requires the flexibility and excellent capability and recognition of various EEG-complexes. In this context, Schetinin has developed an algorithm to classify artifacts and normal segments in clinical EEGs. This method involves evolving cascade neural networks, ensuring a nearly minimal number of input and hidden neurons as well as connections. The algorithm was successfully applied, classifying correctly 96.69% of the testing segments [16].

In this study, different learning methods which is Levenberg-Marquardt (LM), Quickprop (QP), Delta-bar delta (DBD), Momentum and Conjugate gradient (CG) were used on multilayer perceptrons (MLP) architecture in the training of artificial neural networks; these methods were optimized with genetic algorithm; and their performances were compared.

The aim of using different learning algorithms is to achieve faster and more accurate results and to test different learning methods and networks for such applications.

Genetic algorithms can be used in the solution of complex optimization problems. Genetic optimization can also be used for determining the best parameters of ANN. It is known that especially the weights in ANN can vary depending on several parameters. It was observed that the genetic optimization of different learning algorithms and their parameters affects the network performance and increases the classification performance.

Material and method

Experimental protocol

The evaluation of several proposed EEG classification or recognition schemes was carried out on the basis of some signals experimentally acquired from human volunteers/samples. In some cases, the researchers acquired these signals from human volunteers, by setting up their own experimental set up. In many other cases, the researchers demonstrated the efficiacy of their proposed algorithms on the basis of popular benchmark EEG signals, freely available for downloading in the internet. In case of several research works regarding epilepsy recognition, the benchmark EEG signal database is available from the Department of Epileptology, University of Bonn, [17, 18]. This complete database contains five sets of 100 single-channel EEG signals, obtained with a sampling rate of 173.61 Hz and digitized with 12 bit resolution. The same amplifier system was employed for acquiring all EEG signals from human volunteers. The EEG signals were extracted for those portions which were devoid of any artifact e.g., due to eye movements or muscle activities [19]. The spectral bandwidth of the acquisition system was 0.5–85 Hz and the first step for processing of such EEG signals requires low pass filtering with a cut-off frequency of 40 Hz. Out of five sets of data, two sets were obtained from five healthy volunteers, relaxing in awake condition, with their eyes open and eyes closed and the EEG surface recordings were acquired using standard 10–20 international system of electrode placement.

In this research, the clinically significant epilepsy and seizure detection problem is modeled as a two-group classification problem. Epilepsy diagnosis is modeled as the classification of normal EEGs and interictal EEGs. Seizure detection is modeled as the classification of interictal and ictal EEGs. In order to improve the statistical significance, a large number of EEG data sets belonging to two subject groups are used: 1) healthy subjects (normal EEG), 2) epileptic subjects during a seizurefree interval (interictal EEG), and epileptic subjects during a seizure (ictal EEG).

EEG signals contain many concealed information about the brain functions. In order to use this information in research and patient diagnosis, the spectral analysis should be performed in real time with modern parametric methods, and automation should be initiated. Signal analysis is performed by the transfer of the signal itself or its display to other dimensions (time, frequency, time-scale, etc.). The purpose is to reach meaningful detail information, which can not be acquired from raw data, by transforming the signal without any information losses into one of these dimensions.

Parametric features used in the EEG signals

Using EEG signals as the control source are achieved by discriminating their patterns in specified epilepsy disorders. Basically, the entire task in EEG pattern classifications could be divided into three essential parts: data acquisition, feature extraction, and classification algorithm. Several features have been applied to the classifications of EEG signals. Our idea is to extract real EEG signals from the measured signals, so that we can expect two advantages [20].

- It can reduce the amount of data so that less memory is required for EEG signal storage.
- It can result in the improved performance in pattern recognition due to no possible confusion from useless data.

Spectral estimation by FFT-based methods

In order to make meaningful of EEG signal, spectral analysis should be applied to EEG signal. For this reason, FFT method can be applied to the EEG signal which is nonstationary since the algorithm of FFT is not complex. In order to take the FFT of a finite EEG signal, it must be framed with the powers of 2, such as 64, 128, and 256. Windowing technique is used to evaluate the frequency spectrum for the corresponding frame. By using windowing, the non-existing frequency components appearing in the spectrum are presented. In addition, zero padding is applied to the same signal after windowing process. This entails overhead on the process although it increases the readability of spectrum [20]. Fourier analysis is extremely useful for data analysis, as it breaks down a signal into constituent sinusoids of different frequencies. For sampled vector data, Fourier analysis is performed using the Discrete Fourier Transform (DFT). The Fast Fourier Transform (FFT) is an efficient algorithm for computing the DFT of a sequence; it is not a separate transform. It is particularly useful in areas such as signal and image processing, where it uses range from filtering, convolution, and frequency analysis to power spectrum estimation. Fast Fourier transforms:

$$
\begin{aligned}
X(k) &= \sum_{n=1}^{N} x(n)e^{-j2x(k-1)\left(\frac{n-1}{N}\right)} \quad 1 \le k \le N \\
x(k) &= \frac{1}{N}\sum_{k=1}^{N} X(n)e^{j2x(k-1)\left(\frac{n-1}{N}\right)} \quad 1 \le n \le N
\end{aligned}
\tag{1}
$$

where x is a length N discrete signal sampled at times t with spacing.

The number of signal samples required to form a frame depends heavily on how stationary the signal is. In general,

the EEG signal is non-stationary. When amplitude of EEG signal is very low, ANN is used to test diagnosed spectral curve which are estimated from the result of FFT analysis. On the other hand, very short frame lengths may yield statistically poor spectral resolution. Therefore, selection of frame length is an important factor in EEG spectral analysis. The frame length used in this study is 64. FFT coefficients of EEG signals are required to generate ANN as training and testing inputs.

ANN based classification algorithms

Artificial intelligence is an instrument, which can be used in medical field for decision-making support systems. Artificial neural networks are extensively used in the modeling of non-linear system [21, 22].

The MLP (Multilayer Perceptron) comprises an input layer, where all the input signals are lconnected to input nodes, followed by one or more hidden layers where several hidden layer nodes are connected to accommodate more nonlinearity which can, hopefully, help in determining an efficient multidimensional nonlinear mapping between input and output exemplars and this is followed by an output layer which produces the output of the neural network. Figure 1 shows the schematic form of the architecture of a typical m-input–noutput MLP, with a single hidden layer comprising neurons. The general form of an MLP is a fully connected one where each node or neuron in a given layer is connected to all nodes or neurons in the previous layer through some connecting weights. Each node, in its most general form, comprises two functions: integration function and activation function.

The integration function integrates or summates all the weighted inputs at the given node to produce an aggregated input for the activation function. Then the activation function applies nonlinearity on its aggregated input by employing a continuous nonlinear function. These non-
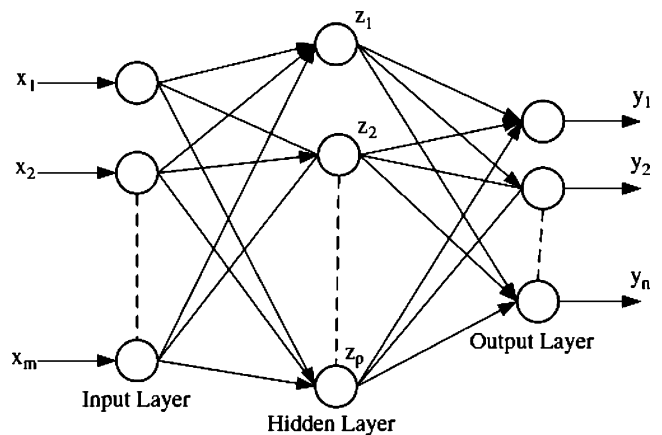

Fig. 1 MLP architecture

linearities can be popularly employed in terms of tanh functions, which are known as smooth functions that are differentiable everywhere. However, in case of some neurons, both the integration and activation function may not be present and, in those situations, in all probability, only the integration function will be present.

This network is trained in a supervised manner, in presence of ideal input–output exemplars, utilized in form of a training data set, which determines the suitable weights and biases of the network. In such cases, the most popular training algorithm is known as the error backpropagation algorithm, where the synaptic weights and biases are adjusted by backpropagating the error signal through different layers of the network in a chain form, with an objective of adjusting the free parameters of the network so that the actual response of the network approaches the ideal response in a statistical sense. This learning algorithm can be employed either in pattern mode, i.e., the weights and biases are adjusted every time an input exemplar is presented to the system, or in batch mode, (i.e., the weights and biases are adjusted every time all the input exemplars present in the training data set are presented once to the system) [21, 22].

Backpropagation with momentum (BPM)

It is a gradient descent method and the most commonly adopted ANN training algorithm [23]. It has local minima and slow convergence problems. It is an extended version of the BP algorithm. The weights and biases are updated according to gradient descent momentum and an adaptive learning rate [24].

Levenberg-Marquardt (LM)

This is a least-squares estimation method based on the maximum neighbourhood idea [25, 26]. The LM combines the best features of the Gauss–Newton technique and the steepest-descent method, but avoids many of their limitations. In particular, it generally does not suffer from the problem of slow convergence.

Scaled conjugate gradient (SCG)

This was developed by Moller, and designed to avoid the time consuming line search [27]. The basic idea is to combine the model-trust region approach and the conjugate gradient approach.

Delta bar delta

Delta-Bar-Delta learning rule was developed in 1988 by Jacobs in order to improve the convergence speed of the classical Delta rule [28].

Quickprop

The Quickprop implements Fahlman's quickprop algorithm. It is a gradient search procedure that has been shown to be very fast in a multitude of problems [29].

Cross-validation

Periodically the network is tested using the cross-validation, using other set of data than the one used for training and the performance must increase from test to test, if it doesn't the training is stopped. The cross-validation is a recommended criterion to stop training at the right moment [30].

Genetic algorithms (GA)

Holland (1975) described a technique, now known as the GA, which used concepts taken from the naturally occurring evolutionary process to solve problems by performing a highly parallel search. The GA begins by randomly generating an initial set or population of candidate solutions. Every population member is allocated a value that is a measure of its performance, known as the individual's fitness. A new population is then generated by applying the Darwinian principle of survival and reproduction of the fittest, making use of operators that are analogous to naturally occurring genetic operators such as sexual recombination (crossover) and mutation. The process is repeated over a number of iterations or generations in an attempt to evolve increasingly accurate solutions. As the individuals in the GA population are typically stored as fixed length character strings, a suitable encoding scheme must be devised before the algorithm can be applied to a problem [31].

GAs are algorithms used to find approximate solutions to difficult problems through application of the principles of evolutionary biology to computer science. They use biologically derived techniques such as inheritance, mutation, natural selection, and recombination to approximate an optimal solution to difficult problems. Genetic algorithms view learning as a competition among a population of evolving candidate problem solutions. A fitness function evaluates each solution to decide whether it will contribute to the next generation of solutions. Through operations analogous to gene transfer in sexual reproduction, the algorithm creates a new population of candidate solutions [32].

In manufacturing there are certain processes that are not possible to describe using analytical models for GA optimization. It has been hard to establish models that accurately correlate the process variables and performance of Levenberg-Marquardt (LM), Quickprop (QP), Delta-bar

delta (DBD), Momentum and Conjugate gradient (CG) learning algorithms. The present work describes the development and application of a hybrid MLP and GA methodology to model and optimize one or more parameters within the neural network. The most common parameters to optimize are the weights, the neuron number of hidden layers, and the learning rates. Many other network parameters are available for optimization.

If the search space consists of two or more dimensions, the gradient-dissent strategy may get caught in repeated cycles, where the local minima solution is found repeatedly. Use of MLP models for prediction of wide range of data is a difficult task. Large differential amplitudes of the solutions targeted at each and every output cause the error surface to be discontinuous and flat in certain regions. GA is a global search method that does not require the gradient data and locates globally optimum solution. The use of GA based learning methods is justified for learning tasks that require MLPs with hidden neurons for a non-linear data, which is the case in the present study.

The task of neural network training in MLP is a complicated process, in which a pattern set made up of pairs of inputs plus expected outputs is known beforehand, and used to compute the set of weights that makes the MLP to learn it. The architecture of the network and the weights are evolved by using error back propagation. The optimization of these weights improves the efficiency of the MLP model. In MLP-GA Hybrid model the concepts of GA are used for optimization of weights resulting to the minimization of error between actual output and MLP predicted output [33].

Receiver operating characteristic (ROC)

For comparison of the diagnostic accuracy of the different classification methods and groups, the concept of Receiver Operating Characteristic (ROC) analysis was used. ROC analysis is an appropriate means to display sensitivity and specificity relationships when a predictive output for two possibilities is continuous. In its tabular form the ROC analysis display true and false positive and negative totals and sensitivity and specificity for each listed cutoff value between 0 and 1. The ROC curves are a more complete representation of the classification performance than the report of a single pair of sensitivity and specificity values. In order to analyze the output data that are obtained from the application, sensitivity (true positive ratio) and specificity (true negative ratio) are calculated by using confusion matrix. Sensitivity value (true positive, same positive result with the diagnosis of expert physicians) is calculated by dividing the total of diagnosis numbers to total diagnosis numbers that are stated by the expert physicians [34].

Results and discussion

The actual purpose of our study is to classify the EEG signals for diagnosis purposes. For this purpose EEG signals of the patients and healthy person were recorded. EEG data was obtained from the database of Bonn University. The sampling frequency of the signals is 173.61 Hz. Spectral analysis was applied to each EEG segment of the data set using Welch Method. Welch method for evaluating the power spectrum divides the data in several segments, possibly overlapping, performs an FFT on each segment, computes the magnitude squared then averages these spectra.

Artificial neural networks (ANN) create complex and non-linear models, which create a relationship between the entries (independent variables of the system) and the outputs (dependent predictive variables). With the increase of the number of studies in this field, different network structures and learning algorithms were studied. Neurons are usually divided into layers in a feed-forward network. Feed-forward networks include multi layer perceptron (MLP) architecture [35].

In this study, 300 input data sets from healthy individuals and 200 input data sets from patients with epileptic disorders were used. Signals were sampled at 173.61 Hz and applied to a band-pass filter of 0.53–40 Hz. Signals were recorded for 23.6 s [17, 36].

Time-scale domain analysis of the mark can also be obtained practically with FFT, and thus some detail information, which can not be healthily selected from the row data in the signal, can be found without any signal information losses whatsoever. This is a method, which can be helpful in the solution of various problems. The benefits of the use of this method can be appreciated even more, when one thinks about the importance in terms of signal processing of the fact that the changeability feature of the signal's screening interval is quite effective for resolving the situations, where especially sharp bounces and discontinuities occur [7].

In Fig. 2, raw (unprocessed) EEG signals of healthy individuals as well as signals that are formed after the application of Fast Fourier Transform (FFT) are shown. In Fig. 3, however, raw (unprocessed) EEG signals of patients with epileptic disorders and spectral curve after the FFT analysis are shown.

A review of the FFT results reveals the connection between the deviations that occur in healthy individuals and the frequency levels that correspond to these deviations. As explained earlier with the information provided with regard to the EEG signals, the alpha signals are seen in the frequency of 8–12 Hz in normal individuals, and their amplitudes are not so large. This result is supportive of the FFT signal results. Similarly when the FFT results of sick individuals (epilepsy)

-a-
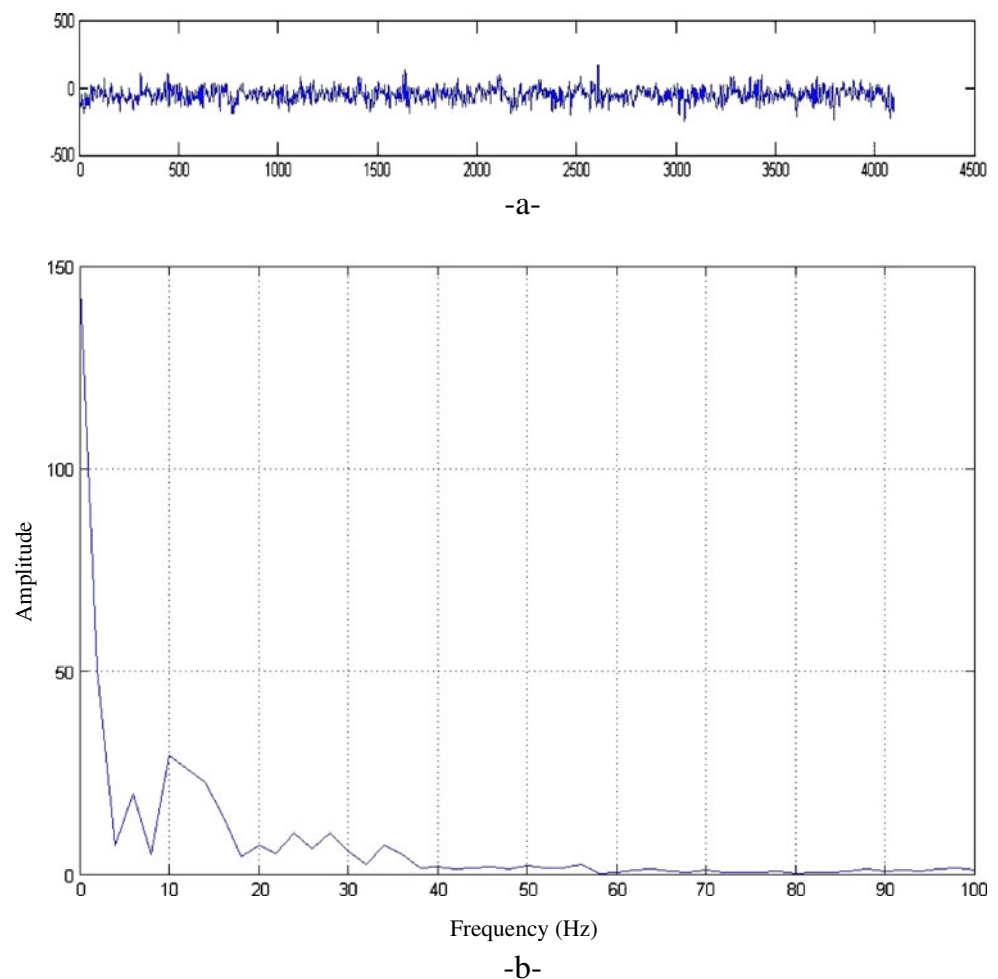
-b-

are reviewed, instant deviations occur in their amplitudes. Instant deviations that correspond to the same frequency compared to healthy individuals points out the difference of EEG mark. While no changes occur in the frequency of the signal in epilepsy, instant increases occur in the amplitude [37].

In order to ensure faster and more accurate diagnosis, FFT coefficients of these spectral curves will be used as the input vector to ANN. Of the EEG data, which were applied to the artificial neural network, 60% was used as training set, 25% was as test set and the rest was used as cross validation (CV) set. A neural network, which has MLP architecture with standard back propagation algorithm, was used in this study. The learning algorithms of Levenberg-Marquardt (LM), Quickprop (QP), Delta-bar delta (DBD), Momentum and Conjugate gradient (CG) were applied in the MLP architecture. The input vectors derived from the EEG data were selected randomly as training, test and CV sets. In all calculations regarding ANN, tangent hyperbolic was selected as the transfer function during application.

Recent developments in the theory of learning and the traditional information of data modeling have shown that while training after a critical point enable the results to get better, they also cause the test results to get worse.

Especially the data are trained more than intended. The training stops at the point, where maximum generalization is made in order to resolve this problem. This method is called the cross-validation method.

In order to achieve this; (1) a set that contains both the training set and the test set shall be created; (2) number of mistakes per each group shall be evaluated in the verification set after every 50 epochs; and the training set shall be trained; (3) the training shall be stopped if the error during the training in the verification set turn out to be higher than the last checked error value; and (4) the weights in the network are used, which the training had during the phase before the work.

Accordingly EEG data were trained by selecting 1,000 epochs; and an MSE and epoch value, where error is minimum, is achieved by using CV (cross-validation).

It is seen in Table 1 that the highest epoch value was achieved in QP (Quick prob) and the lowest epoch values in DBD (Delta bar delta) learning algorithms. After a review of the MSE values it is revealed that the highest MSE value was achieved in QP and the lowest MSE value in LM learning algorithms.

Genetic algorithms can be combined with neural networks to enhance their performance by taking some of the

Fig. 3 Raw EEG signal (**a**) and spectral curve (**b**) belonging to patient person
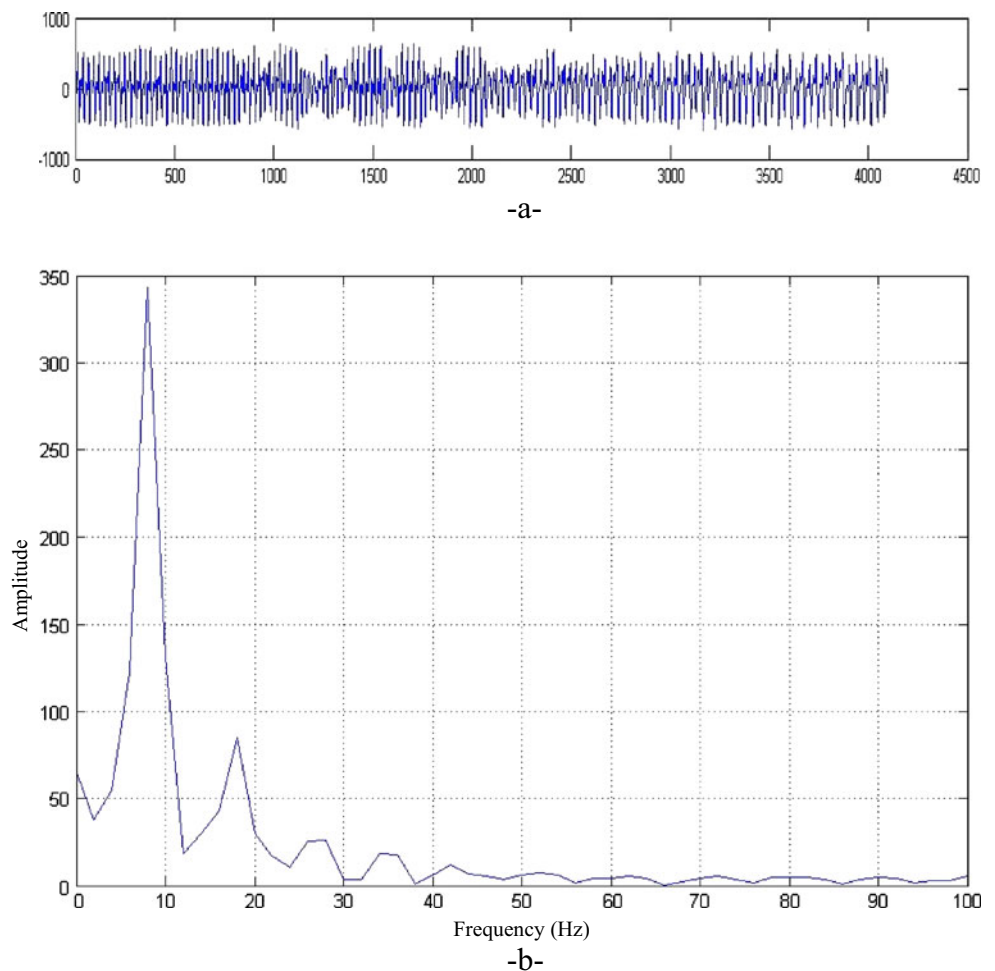
guesswork out of optimally choosing neural network parameters, inputs etc. They can be used to choose the best inputs to the neural network optimize the neural network parameters (such as the learning rates, neuron number of hidden layer, etc.), train the actual network weights.

Multi layer backpropagation network was used for training; and tangent hyperbolic function was used as the activation function. The weights were updated by using the back-propagation network and the genetic algorithm. The network training was completed with the achievement of acceptable error value. Optimization was performed with each standard of ANN architecture belonging to individual learning methods and with the genetic algorithm. The comparison table for performance values was given in Table 2.

After the review of Table 2, average highest success of 90% was achieved with CG and LM algorithms during the test phase before the application of genetic algorithm. Weight values of the input layer and concealed layers as well as the coefficients of each layer were optimized with GA after each cycle; and it was ensured that these optimized values are used for the next cycle. As can be seen in Table 2, it was observed that the classification periods of all learning algorithms increased. The highest success rate here (96.5% in average) was seen in LM algorithm. It is, however, determined that no significant changes were present in CG learning rule. The lowest MSE value of 0.048 here was achieved in LM algorithm. It is considered that the approach of value r caused the network

Table 1 The epoch and MSE values of learning methods

Learning method	LM	CG	QP	DBD	Momentum
Epoch #	276	536	971	72	187
Minimum MSE	0.08694004	0.07322647	0.44228187	0.38088033	0.31518395
Final MSE	0.08769681	0.11812871	0.69000029	0.74861335	0.38535204

Table 2 Comparison performance values for standard MLP algorithm and MLP algorithm using genetic algorithm optimization

	Learning method Performance	QP		CG		LM		DBD		Momentum	
		Normal	Patient	Normal	Patient	Normal	Patient	Normal	Patient	Normal	Patient
Standard MLP algorithm	MSE	0.198	0.195	0.085	0.085	0.085	0.085	0.244	0.240	0.214	0.216
	r	0.532	0.582	0.809	0.809	0.809	0.809	0.041	0.035	0.548	0.442
	Percent correct	14	97	88	92	88	92	10	100	24	97
MLP algorithm using genetic optimization	MSE	0.176	0.176	0.082	0.082	0.048	0.048	0.165	0.163	0.121	0.120
	r	0.570	0.570	0.817	0.817	0.907	0.907	0.585	0.593	0.730	0.732
	Percent correct	52	98	89	92	95	98	80	91	94	79

to access the training better. Similarly the highest success here was achieved with the LM learning rule.

Receiver Operation Characteristics (ROC) analysis was performed for all networks in order to verify the diagnosis performances of artificial neural networks. ROC analysis is used in order to determine the actual precision of medical diagnosis results. ROC analysis is a standard approach that is used in order to determine the sensitivity and the certainty of the diagnosis procedure. ROC curves are used for this purpose, which will define the relationship between the sensitivity and the certainty of the diagnosis.

In a Receiver Operating Characteristic (ROC) curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points. Each point on the ROC plot represents a

sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions) has a ROC plot that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore the closer the ROC plot is to the upper left corner, the higher the overall accuracy of the test.

For this purpose ROC curves were shown to each learning algorithm in Fig. 4. Area under the Curve (AUC), Standard Error (SE) and Confidence interval (CI) values for these curves were summarized in Table 3. According to Table 3, the highest AUC value (0.930) was achieved in CG learning rule, followed by Momentum, LM, QP and DBD learning period. After the review of the classification percentages in previous Table 2, it was observed that both LM and CG have same classification percentages.

Genetic algorithms (GAs) have been successfully implemented for optimizing ANN architectures [38–40]. In order to reduce the MSE at training and error for production data sets, it is proposed to train the developed network using genetic algorithms (GA). The advantage of ANN model with GA is that it optimizes the network weights to minimize the MSE during training of the network.
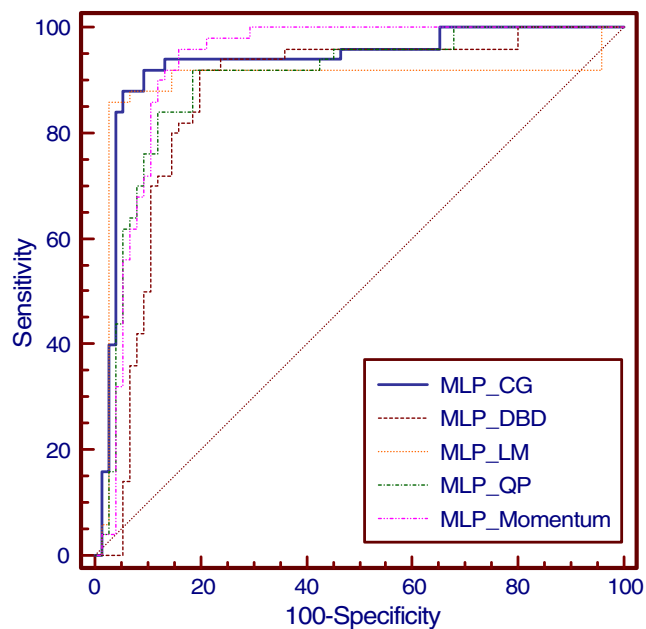

Fig. 4 ROC curves belonging to standard MLP algorithm

Table 3 Area under the curve (AUC), standard error (SE) vs. confidence interval (CI) for the ROC curves belonging to standard MLP algorithm

	AUC	SE	95% CI
MLP_CG	0.930	0.0263	0.870 to 0.968
MLP_DBD	0.863	0.036	0.790 to 0.918
MLP_LM	0.910	0.0321	0.856 to 0.962
MLP_QP	0.894	0.032	0.827 to 0.942
MLP_Momentum	0.924	0.0274	0.863 to 0.964

Fig. 5 ROC curves belonging to MLP algorithm using genetic algorithm optimization

Conclusion

Classification process is, at its most common sense, a decision-making mechanism. In this study, the differences shown by the non-stationary random EEG signals in cases of health and sickness (epilepsy) were evaluated and analyzed under computer-supported conditions by using artificial neural networks.

It is seen that the most efficient results in the study was achieved with the use of LM (Levenberg-Marquardt back-propagation) algorithm, which is optimized with the genetic algorithm. The general success here was 96.5%. In a separate evaluation for sick and healthy data, it is evident that the healthy data are superior. The success rate of healthy data is 95%, while the rate of sick data is 98%.

One of the major problems of patient diagnosis is being able to diagnose sick individuals as sick, and healthy individuals as healthy. A mistake in such a diagnosis is a situation, which is never wanted, since it involves risks for human health. For this purpose, EEG data set was evaluated with ANN structures in this study; ROC analysis was performed for all network structures that were used; sensitivity and clearance values were calculated; and the authenticity of the tests were controlled. ROC analysis results showed that the highest AUC value of 0.972. In this case it is safe to say that the most efficient result is achieved in the LM algorithm in terms of both the lowest MSE value and the success.

Appropriate ANN architectures can be selected by creating various ANN architectures in classification systems that are based on artificial neural networks. It was seen that the system performance changed depending on these MLP architectures. The weights in the architecture that was used were optimized during the training with genetic algorithm, and a major increase was observed in the system performance. A performance increase is also possible with the activation functions and parameters that are being used.

The numbers of neurons in hidden layer, the momentum and the learning rates have been determined using GA algorithm to minimize the time and effort required to find the optimal architecture and parameters of the back propagation based on MLP architectures. These values were optimized by using a generic algorithm, and an increase was observed in the classification performance.

Comparison of the results of GA-MLPs with the trial and error method indicates that GA approach is more efficient. In the other words, GA is found to be a good alternative over the trial and error approach to determine the optimal MLP architecture and internal parameters quickly and efficiently.

Optimization was carried out by using genetic algorithm in order to increase the classification performances of Levenberg-Marquardt (LM), Quickprop (QP), Delta-bar delta (DBD), and Momentum and Conjugate gradient (CG) learning algorithms. The performance increase with GA can be seen in ROC curves of each learning algorithm (Fig. 5). Area Under the Curve (AUC), Standard Error (SE) and Confidence Interval (CI) values for these curves were summarized in Table 4.

It is evident from the Table 4 that the highest AUC value was achieved in the LM learning rule, which is followed by CG, DBD, QP and Momentum learning algorithms, respectively. The AUC value (0.972), which was achieved with the LM learning algorithm here, is higher than the AUC value (0.930), which was achieved in CG, shown in Table 3. Classification performance and lowest ME value are also achieved with the LM algorithm.

Table 4 Area under the curve (AUC), standard error (SE) vs. confidence interval (CI) for the ROC curves belonging to MLP algorithm using genetic algorithm optimization

	AUC	SE	95% CI
MLP_CG	0.934	0.0255	0.876 to 0.971
MLP_DBD	0.947	0.0229	0.892 to 0.979
MLP_LM	0.972	0.0168	0.926 to 0.993
MLP_Momentum	0.922	0.0277	0.860 to 0.962
MLP_QP	0.928	0.0267	0.867 to 0.966

References

1. Novak, D., et al., *EEG and VEP signal processing*. Czech Technical University in Prague Depmment of Cybernetics, date unknown.

2. Hector, M. L., *EEG recording*. Butterworth & Co (Publishers) Ltd, 1980.

3. Ghosh-Dastidar, S., Hojjat, A., and Nahid, D., Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Trans. Biomed. Eng.* 55(2), 2008.

4. Yazgan, E., and Korurek, M., Tıp Elektronigi, İTÜ Matbaası, 95–220, 1996.

5. Vuckovic, A., Radivojevic, V. A., Chen, C. N., and Popovic, D., Automatic recognition of alertness and drowsiness from EEG by an artificial neural network. *Med. Eng. Phys.* 24:349–360, 2002.

6. Tsoi, A. C., So, D. S. C., and Sergejew, A., Classification of electroencephalogram using artificial neural networks. In: Cowan, J. D., Tesauro, G., and Alspector, J. (Eds.), *Advances in Neural Information Processing Systems*. Morgan Kaufmann, 6:1151–1158, 1994.

7. Tseng, S.-Y., Chen, R.-C., Chong, F.-C., and Kuo, T.-S., Evaluations of parametric methods in EEG signal analysis. *Med. Eng. Phys.* 17:71–78, 1995.

8. Miller, A. S., Blott, B. H., and Hames, T. K., Review of neural network applications in medical imaging and signal processing. *Med. Biol. Eng. Comput.* 30:449–64, 1992.

9. Webber, W. R. S., Lesser, R. P., Richardson, R. T., and Wilson, K., An approach to seizure detection using an artificial neural network (ANN). *Electroencephalogr. Clin. Neurophysiol.* 98:250–272, 1996.

10. Petrosian, A., Prokhorov, D., Homan, R., Dashei, R., and Wunsch, D., Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG. *Neurocomputing*. 30:201–218, 2000.

11. Sinha, R. K., Backpropagation artificial neural network to detect hyperthermic seizures in rats. *Online Journal of Health and Allied Sciences*. 4:1, 2002.

12. Kiymik, M. K., Akin, M., and Subasi, A., Automatic recognition of alertness level by using wavelets transform and artificial neural network. *J. Neurosci. Methods*. 139:231–240, 2004.

13. Sinha, R. K., Ray, A. K., and Agrawal, N. K., An artificial neural network to detect EEG seizures. *Neurol. India*. 52:399–400, 2004.

14. Subasi, A., Automatic detection of epileptic seizure using dynamic fuzzy neural networks. *Expert Syst. Appl.* 31:320–328, 2006.

15. Keshri, A. K., Sinha, R. K., Hatwal, R., and Das, B. N., Epileptic spike recognition in electroencephalogram using deterministic finite automata. *J. Med. Syst.* 33:173–179, 2009.

16. Olivier, C. L., Haas, K. J., *Burnham, intelligent and adaptive systems in medicine*. Taylor & Francis Group, 2008.

17. Andrzejak, R., Lehnertz, K., Rieke, C., et al., 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*. 64:061907, 2001.

18. Online. Available: http://www.meb.unibonn.de/epileptologie/cms/front/content.php?idcat=193&lang=3&changelang=3. Last accessed on: July 23, 2009.

19. Subasi, A., EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Syst. Appl.* 32:1084–1093, 2007.

20. Güler, N. F., Koçer, S., Classification of EMG signals using FFT and PCA. *J. Med. Syst.* 29(3), 2005.

21. Na, t-Ali (Ed.), *Advanced Biosignal Processing*. Springer-Verlag Berlin Heidelberg, 2009.

22. Acır, N., Oztura, I., Kuntalp, M., et al., Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks. *IEEE Trans. Biomed. Eng.* 52:30–40, 2005.

23. Rumelhart, D. E., and Mcclelland, J. L., *Parallel distributed processing: Explorations in the microstructure of cognition*. MIT Press, Cambridge, 1986. Vol. 1.

24. Hagan, M. T., Demuth, H. B., and Beale, M. H., *Neural network design*. PWS Publishing, Boston, 1996.

25. Levenberg, K., A method for the solution of certain nonlinear problems in least squares. *Q. Appl. Math.* 2:164–168, 1944.

26. Marquardt, D. W., An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math*. 11:431–441, 1963.

27. Moller, M. F., A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw.* 6:525–533, 1993.

28. Jacobs, R. A., Increased rates of convergence through learning rate adaptation. *Neural Netw.* 1 (4)295–307, 1988.

29. Fahlman, S. E., Faster-learning variations on back-propagation: An empirical study. In: Touretzky, D., Hinton, G, and Sejnowski, T., (Eds.), *Proceedings of the 1988 Connectionist Models Summer School*. Morgan Kaufmann, 38–51, 1989.

30. Tout, K., Sinno, N., and Mikati, M., Prediction of the Epileptic Events 'Epileptic Seizures' by Neural Networks and Expert Systems, Proceedıngs Of World Academy Of Scıence, Engıneerıng And Technology 31, 2008.

31. Hinchliffe, M., *Dynamic modelling using genetic programming*. Department of Chemical and Process Engineering University of New castle upon Tyne September 2001.

32. Abdella, M., Marwala, T., Treatment of missing data using neural networks and genetic algorithms, Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, July 31 - August 4, 2005.

33. Krishna Mohana Raoa, G., Rangajanardhaa, G., Hanumantha Raoc, D., and Sreenivasa Raoa, M., Development of hybrid model and optimization of surface roughness in electric discharge machining using artificial neural networks and genetic algorithm. *J. Mater. Process. Technol.* 209:1512–1520, 2009.

34. Güler, N. F., and Koçer, S., Use of support vector machinesneural network in diagnosis of neuromuscular disorders. *J. Med. Syst.* 29 (3), 2005.

35. Semmlow, L. J., Biosignal and Biomedical İmage Processing Matlab Based Application 198–206, 1998.

36. Webber, W. R. S., Lesser, R. P., Richardson, R. T., and Wılson, K., An approach to seizure detection using an artificial neural network (ANN). *Electroencephalogr. Clin. Neurophysiol.* 98:250–272, 1996.

37. Mutlu S., Application of artificial intelligence techniques to EEG signal, Master Thesis, Gazi Üniversitesi, Ankara, 2008.

38. Reeves, C. R., and Steele, N. C., Genetic algorithms and the design of artificial neural networks. *IEEE 'Microarch'* 15–20, 1991.

39. Wu, J.-X., Zhou, Z.-H., and Chen, Z.-Q., Ensemble of GA based Selective Neural Network Ensembles, Proceedings of the 8th International Conference on Neural Information Processing, Shanghai, China, 1477–1482, 2001.

40. Ozturk, N., Use of genetic algorithm to design optimal neural network structure. *Eng. Comput.* 7:979–997, 2001.