Computer Methods and Programs in Biomedicine 66 (2001) 167–181

Computer Methods and Programs in Biomedicine

www.elsevier.com/locate/cmpb

ECG beat classification by a novel hybrid neural network

Zümray Dokur *, Tamer Ölmez

Istanbul Technical University, Department of Electronics and Communication Engineering, 80626 Maslak, Istanbul, Turkey

Received 26 November 1999; received in revised form 13 July 2000; accepted 15 July 2000

Abstract

This paper presents a novel hybrid neural network structure for the classification of the electrocardiogram (ECG) beats. Two feature extraction methods: Fourier and wavelet analyses for ECG beat classification are comparatively investigated in eight-dimensional feature space. ECG features are determined by dynamic programming according to the divergence value. Classification performance, training time and the number of nodes of the multi-layer perceptron (MLP), restricted Coulomb energy (RCE) and a novel hybrid neural network are comparatively presented. In order to increase the classification performance and to decrease the number of nodes, the novel hybrid structure is trained by the genetic algorithms (GAs). Ten types of ECG beats obtained from the MIT-BIH database and from a real-time ECG measurement system are classified with a success of 96% by using the hybrid structure. © 2001 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: ECG beat classification; Neural networks; Wavelet; Genetic algorithms

1. Introduction

Electrocardiogram signal being the recording of the heart's electrical activity provides valuable clinical information about heart's performance. Many algorithms have been proposed over years for electrocardiogram (ECG) beat detection and classification. In a clinical setting, such as an intensive care unit, it is essential for automated systems to accurately detect and classify electrocardiographic signals on a real time basis.

It is observed that ECG feature extraction methods can be divided into three functional groups: direct methods; transformation methods; and characteristic parameter estimation methods.

1.1. Direct methods

The magnitudes of the samples are used to perform the classification [1–10]. The dimension of the feature vectors is kept at a high value to include enough information about P, Q, R, S and T characteristic points. In most studies of this group, dimension reduction methods are used to simplify the classifier structure [2,11,12]. Since feature vectors are only formed by the magnitudes of the ECG signals, computational cost is low. However, feature vectors are affected by the de-

* Corresponding author. Tel.: +90-212-2853643; fax: +90-212-2853565.

E-mail address: zumray@ehb.itu.edu.tr (Z. Dokur).

PII: S0169-2607(00)00133-4

termination of the R-peak position. If the R-peak position is not correctly determined, the magnitudes of the ECG signals in the feature space will shift and vectors will scatter in the feature space. In addition, noise directly affects the elements of the feature vectors.

1.2. Transformation methods

The original samples are subjected to a transformation, and the classification is performed in the new domain. Mathematical transformations are applied to the signals to obtain a further information from the signal that is not readily available in the raw signal. Since feature vectors are formed by using transforms, computational cost is high. Transformation methods prevent the scattering of vectors in the feature space.

We observed that there exist two promising methods in the literature to extract features from the ECG: (i) Wavelet transform [13–16]; and (ii) Fourier transform [14,17–22]. In the literature, there is another transform known as Karhunen–Loeve. This method is well suited for signal compression. However, in terms of classification, discriminatory information between classes is not preserved with this transform [23].

1.3. Characteristic parameter estimation methods

A preprocessing is employed to extract some features that are used during the classification. The methods in this group determine the characteristic points like P, Q, R, S and T [24–27]. Feature vectors are formed by time intervals of the characteristic points and signal magnitudes at these points. The dimension of the feature vectors is low. However, these parameters are affected by noise in human body and in the instrumentation system, and also they show variation in different patients.

In the literature, it is observed that the number of classes that ECG beats are separated into is usually four [3,4,10,16] or five [1,8]. It is important to increase the number of classes to enhance the ECG data analysis. However, as the number of classes increases, performance of the class separation decreases [8]. It is observed that some studies tend to develop new neural network structures to increase both the number of classes and the classification performance [7].

In this study, a novel hybrid neural network structure and feature extraction method are presented. Classification performance and the number of classes are increased by using the feature extraction method based on wavelet transform and the hybrid neural network structure together.

2. Methods

Decision making is performed in three stages: normalization process; feature extraction method; and artificial neural network as a classifier.

A window, which is formed by 256 discrete data, is selected so that it contains a single ECG beat. For ten classes, training set is formed by choosing 1500 vectors (150 vectors from each class) of 256 dimensions from the MIT-BIH database [28]. Vectors in the training set are normalized, and then features for each normalized vector are extracted. After the feature vectors are formed, they are applied to the input of the neural networks for training or classification purposes. Fig. 1 shows the blocks used in the decision making.

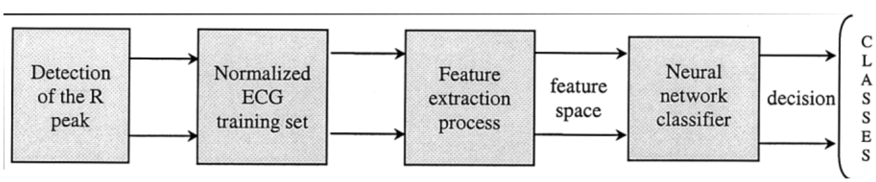

Fig. 1. Decision making blocks.

2.1. Normalization process

Feature extraction processes are affected by the peak-to-peak magnitudes, the offset of the signal, and R-peak position in the windowed ECG. These effects are due to physiology, sex and age of the patient, and parameters of the measurement system. In the normalization, three processes are realized:

1. A rectangular window is formed so that a single ECG beat is contained in this window. By adjusting the ECG signal, the position of the R-peak in the QRS complex is centered in the window.
2. Peak-to-peak magnitudes of the ECG signal are normalized to a value of 1 mV. Thus, it is provided that classification decision does not depend on the maximum amplitude of the ECG records.
3. Mean value of the ECG signal in the window is fixed to zero value. Thus, offset is removed from the signal.

The dependence of the feature extraction method to the offset and the peak-to-peak magnitude of the signal is decreased by the normalization.

2.1.1. Feature extraction methods

In the study, the best ECG features among all the extracted ones are searched by dynamic programming (DP) according to the divergence value. Divergence analysis gives information about the distribution of the class vectors in the feature space. As the class vectors arbitrarily scatter, the divergence value decreases. Divergence value is computed as follows [23]:

$$W_i^j = \sum_n (\underline{K}_i^{nj} - \underline{\hat{\mu}}_i^j)(\underline{K}_i^{nj} - \underline{\hat{\mu}}_i^j)^T$$

$$\underline{K}_{15}^T = [k_1, k_2, \ldots\ldots, k_{15}] \quad \underline{\hat{\mu}}_i = \sum_{k=1}^{C} \underline{\hat{\mu}}_i^k$$

$$W_i = \sum_{k=1}^{C} W_i^k \quad B_i = \sum_{k=1}^{C} (\underline{\hat{\mu}}_i - \underline{\hat{\mu}}_i^k)(\underline{\hat{\mu}}_i - \underline{\hat{\mu}}_i^k)^T$$

$$D_i = tr((W_i)^{-1} B_i) \tag{1}$$

where $\underline{K}_i^{nj}$ is the i-dimensional nth feature vector of the jth class, $\underline{\hat{\mu}}_i^j$ is the mean vector of the feature vectors of the jth class, W_i^j is the within class scatter matrix of the jth class, B_i is the between class scatter matrix, C is the number of classes, D_i is the divergence value at i-dimensions, and tr(·) is the trace operation.

Divergence analysis determines the best features, which increase the classification performance and also it gives information to determine the dimension of the feature vectors. In a previous study [14], two different sets of feature vectors with 15 dimensions were formed by Fourier and wavelet analyses. Two sets were comparatively investigated to find the solution that increased the classification performance. Each set contained 1000 vectors with 15 dimensions. In this study, the numbers of the vectors in the training and test sets are increased to 1500 and the dimension of the feature vectors is decreased to a value of 8. Moreover, these sets are enhanced by data obtained from 20 different records of the MIT-BIH database and a real-time ECG measurement system. The best features are determined by applying divergence analysis to the new training set of 1500 vectors.

In the first feature extraction method, frequency analysis involves the discrete Fourier transform (DFT) of a window of data of length 256 around the R-peak. From the calculated 40 DFT coefficients a subset of the best eight coefficients is searched by using DP according to the divergence values. The frequency-based elements of each feature vector are as follows:

k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8
y_1	y_{11}	y_6	y_7	y_2	y_{35}	y_{10}	y_4

$$K = [k_1, k_2, \ldots, k_8] \qquad Y = [y_1, y_2, \ldots, y_{40}]$$

$$\text{frequency} = i \times 360/256 \qquad (y_i,\ i = 1, 2, \ldots, 40) \tag{2}$$

where Y represents the modules of the DFT coefficients, K is the eight-dimensional new feature vector, 360 is the sampling frequency of the ECG signals in the MIT-BIH database.

With the features shown above, divergence value is found as 2.52. Low divergence value indicates that vectors formed by the DFT of the ECG signals scatter in the feature space.

Second feature extraction method involves the discrete wavelet transform (DWT) of a window of data of length 256 around the R-peak. Feature vectors are formed by using Daubechies-2 wavelets [29]. For each ECG beat, wavelet detail coefficients (d^k, $k=2,3,4$) at the second, third and fourth levels ($66+34+18$ coefficients), wavelet approximation coefficients (a^4) at the fourth level (18 coefficients), and autocorrelations (r_{a0}–r_{a10}) of these approximation coefficients are computed. From the 147 ($66+34+18+18+11$) elements a subset of the best eight features that give the highest divergence value is searched by DP in the same way as done in the previous analysis. Divergence analysis on this feature set showed the fourth level wavelet approximation coefficients and their autocorrelations (r_{aj}) to be more advantageous than the detail coefficients. The use of autocorrelation values in wavelet analysis decreases the dependence of the elements of the feature vectors on the R-peak position. Since the module values of the Fourier transform are not affected by the misdetections (shifts of R-peaks positions) of the R-peaks positions in the ECG recordings, autocorrelation is not used in the first feature extraction method. With the features found in the order as shown below, divergence value is maximized at 7.21 value.

k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8
r_{a1}	r_{a2}	a^4_{17}	a^4_9	a^4_{18}	a^4_7	a^4_{14}	a^4_6

$$K=[k_1,\ k_2,\ \ldots,\ k_8] \qquad R=[r_{a0},\ r_{a1},\ \ldots,\ r_{a10}]$$

$$A=[a^4_1,\ a^4_2,\ \ldots,\ a^4_{18}] \tag{3}$$

where A represents the approximation coefficients at the fourth level, R represents the autocorrelations of these coefficients, K is the eight-dimensional new feature vector.

In these analyses, it is observed that only k_8–k_{15} elements of the feature vectors are different from the k_8–k_{15} elements of the vectors found in the previous study [14]. Simulation results show that the order of the k_1–k_7 elements of the feature vectors is not affected by changing the training set.

3. Artificial neural networks

Constitution of the right feature space for the physical process under study is a common problem in connection with classification. There is not a unique feature extraction method that can suit all kinds of data items. Divergence analysis showed that both of the methods in the preceding section do not provide enough information to separate ten classes. In this case, the solution of the problem is searched in the classifier structures, and artificial neural networks (ANNs) are used as classifiers to increase the classification performance. There are four reasons to use an artificial neural network as a classifier: (1) Weights representing the solution are found by iteratively training; (2) ANN has a simple structure for physical implementation; (3) ANN can easily map complex class distributions; and (4) Generalization property of the ANN produces appropriate results for the input vectors that are not present in the training set.

In the study, three neural networks are comparatively investigated for ECG beat classification: MLP, RCE and a novel hybrid neural network which is called intersecting spheres (InS) network. In order to examine the classification ability of the networks, a two-dimensional sample space is formed for test and observation purposes. Analysis in two dimensions gives a visual information about the representation of the class boundaries by the nodes of the ANNs. The sample space also gives information about how the training algorithms of the classical networks direct the nodes in the feature space. Fig. 2 shows the two-dimensional sample space. In this space, there are three classes: white, light gray and dark gray (black colored region does not belong to any class). Training set is formed by selecting 100 vectors from each class in the 2D space. A sample space with a low divergence value is specially selected to test the classification abilities of the networks. Fig. 2 shows that feature vectors scatter in the sample space. This analysis is used in the development of new neural network structures.

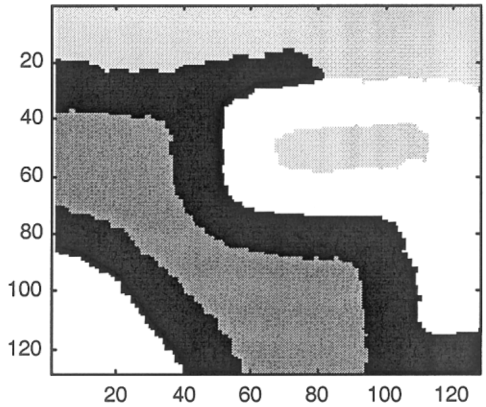

Fig. 2. Two dimensional sample feature space.

Multi-layer perceptron [30] is frequently used in biomedical signal processing [31]. It is observed that MLP has three disadvantages: (1) back-propagation algorithm takes too long time during learning; (2) the number of nodes in the hidden layers must be defined before the training; the structure is not automatically determined by the training algorithm; and (3) Back-propagation algorithm may be caught by local minima, which decreases network performance.

RCE networks [32] have been quite successful in classification tasks. They use hyperspheres (HSs) to determine class boundaries. For this purpose, the vectors in the training set are used in a supervised learning scheme to determine the centers and the radii of the hyperspheres. A drawback of the algorithm is that the order in which the vectors are selected in the training set affects how the centers of the HSs and their radii are chosen. There is no differentiation as to whether the centers of the hyperspheres are nearer or far away from the class boundaries. This results in many HSs being created some of them being redundant for classification.

It is observed that new neural network structures have been investigated to increase the classification performances of the ECG beats [7]. In this study, a novel neural network structure is developed to produce a solution to the problems mentioned above.

3.1. Partitioning of the feature space

Each node in the first layer of MLP forms a hyperplane (HP) in the feature space. The HPs intersect each other and produce regions which are called cells [33]. Fig. 3a shows the cells which are formed by the intersections of three HPs (nodes) of a single neural network in two dimensions. In Fig. 3a, three classes are represented by seven cells in the sample space. ce_i represents the ith cell, f is a hard limiter that generates binary values. Y_i is the binary output of the HP_i. The output of the first layer is a binary vector, to be called a codeword. This codeword specifies on which side of each HP the points in that cell lie. Each codeword represents a cell in the feature space.

InS network has three layers. Each node in the first layer forms a hypersphere (HS) in the feature space. The HSs form regions which are called cells. Fig. 3b shows the cells which are formed by the intersections of three HSs (nodes) of a single neural network. In RCE and GARCE [14] networks, each node represents a single closed region and HS intersection is not allowed. Therefore, a single node can form only one cell. In InS network, HSs intersect with each other. By the intersection of the HSs, more cells are generated compared to the RCE and GARCE networks.

3.2. Structure of the novel hybrid neural network

Fig. 4 shows the structure of the InS network. The weights of a node in the first layer represent a single hypersphere in the input space of the network. Mathematical expressions of the weights of the nodes in the first layer are as follows:

$$D_j = \sum_{i=1}^{N} (x_i - w_{ji})^2 \qquad Y_j = \begin{bmatrix} 1 & D_j \le r_j^2 \\ 0 & \text{otherwise} \end{bmatrix} \tag{4}$$

where x_i represents the elements of the input vector, w_j is the weight vector which represents the center of the jth hypersphere (HS_j), Y_j is the output of the jth hypersphere in the first layer, r_j is the radius of the HS_j, N is the dimension of the feature space.

In the study, the number and the weights of the nodes in the first layer are adjusted by the genetic algorithms. In the second layer, the weights of the nodes are formed by the codewords found during the training process. The block at the output of gorithms of the classical networks direct the nodes in the feature space. Fig. 2 shows the two-dimensional sample space. In this space, there are three classes: white, light gray and dark gray (black colored region does not belong to any class). Training set is formed by selecting 100 vectors of the training set. Codewords formed by the cells that contain at least one vector are used to form the weights of the nodes in the second layer. During the training, more than one node may be generated for any class. The third layer is used to combine these nodes as a single class node.

3.3. Genetic algorithms for the training of the InS network

In the literature, it is observed that each neural network is trained by its own special algorithm. First and second layers of the InS network consist of nodes which are represented by different mathematical expressions (they have different structures). In the study, GAs are used to determine the weights of the nodes in the first layer. The nodes in the second layer are formed by the codewords corresponding to the cells, which con-

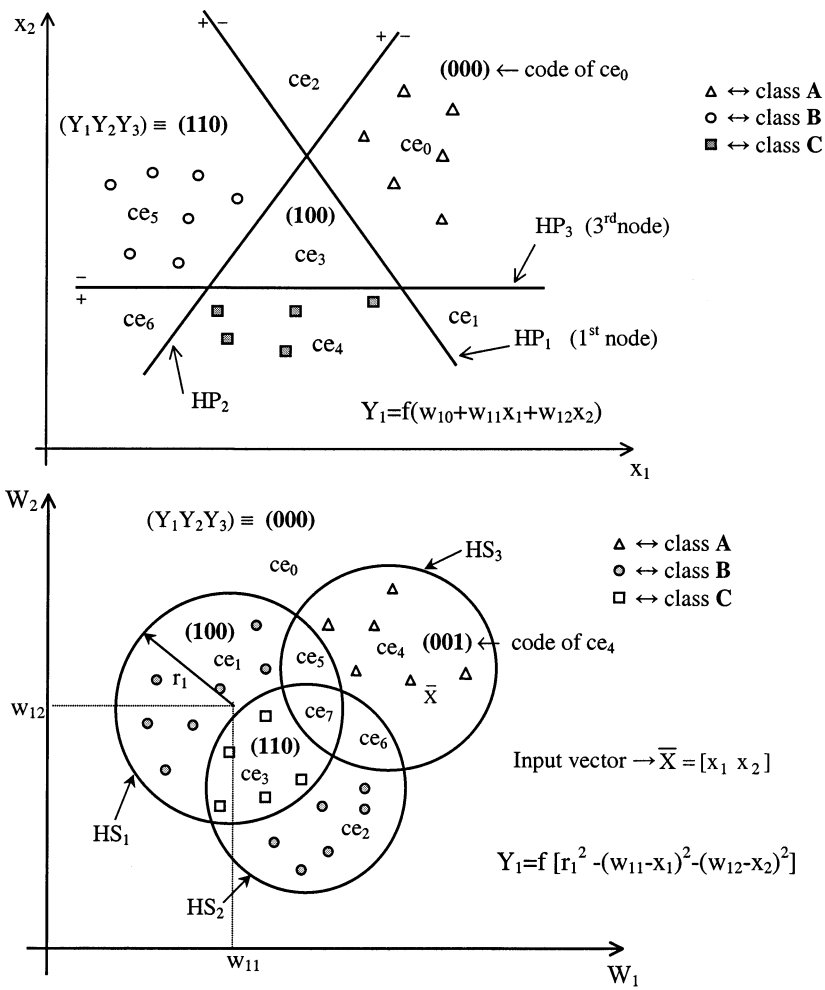

Fig. 3. a: The cells formed by intersecting three HPs (nodes) in two dimensions. three classes are represented by seven cells (ce_i). A binary codeword is attached to each cell in the feature space. b: The cells formed by intersecting three HSs in two dimensions. Three classes are represented by eight cells (ce_i).

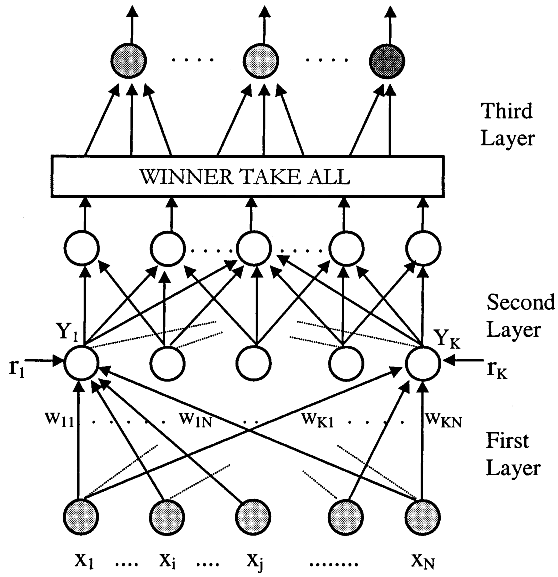

Fig. 4. The structure of the InS network. N is the dimension of the feature space.

sist of at least one vector of the training set. The nodes in the third layer are used for logical OR operation.

In genetic operations [34,35], one generation consists of reproduction, crossover and mutation operations. Reproduction operation depends on a fitness value. The fitness value determines the copy number of the strings in the genetic pool. The fitness function (FF) is used to optimize the location and the radius of the HS so that the cells, which are formed by the intersections of the HSs contain the highest number of vectors of the same class.

The fitness function is determined as follows:

$$T = \bigcup_{i=0}^{M-1} \{\mathrm{ce}_i\}$$

$$\{\mathrm{ce}_{ik}\} = \bigcup_{k=1}^{C} \{\mathrm{ce}_{ik}\}$$

$$= \{\mathrm{ce}_{i1}\} \cup \{\mathrm{ce}_{i2}\} \cup \ldots\ldots \cup \{\mathrm{ce}_{iC}\}$$

$$\mathrm{MAX}_i = \#\{\mathrm{ce}_{ij}\}$$

$$= \mathrm{Max}(\#\{\mathrm{ce}_{i1}\}, \#\{\mathrm{ce}_{i2}\}, \ldots, \#\{\mathrm{ce}_{iC}\})$$

$$FF = \sum_{i=0}^{M-1} \frac{\mathrm{MAX}_i}{\#\{T\}} \tag{5}$$

where $\{T\}$ is the training set, ce_i is the ith cell, $\{\mathrm{ce}_i\}$ is the set of vectors in the ith cell, $\{\mathrm{ce}_{ij}\}$ is the set of vectors of the jth class in the ith cell, M is the number of cells formed in the feature space, C is the number of classes.

Genetic pool consists of 16 strings, each of which represents a single node. Each weight of a node is represented by eight bits in the genetic pool. Candidate network is formed by the winner network that has the constant nodes found in previous generations and a single node that is being searched in the genetic pool. Since the genetic pool contains 16 nodes, there are 16 different candidate networks (CN), shown in Fig. 5.

In a single generation, after the fitness value of each CN is computed, copies of strings according

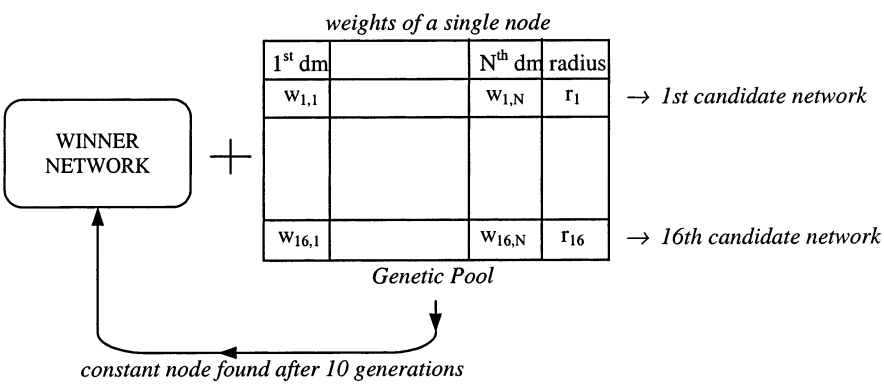

Fig. 5. Candidate networks formed by the constant nodes in winner network and a single node in the genetic pool. Dimension is abbreviated as dm.

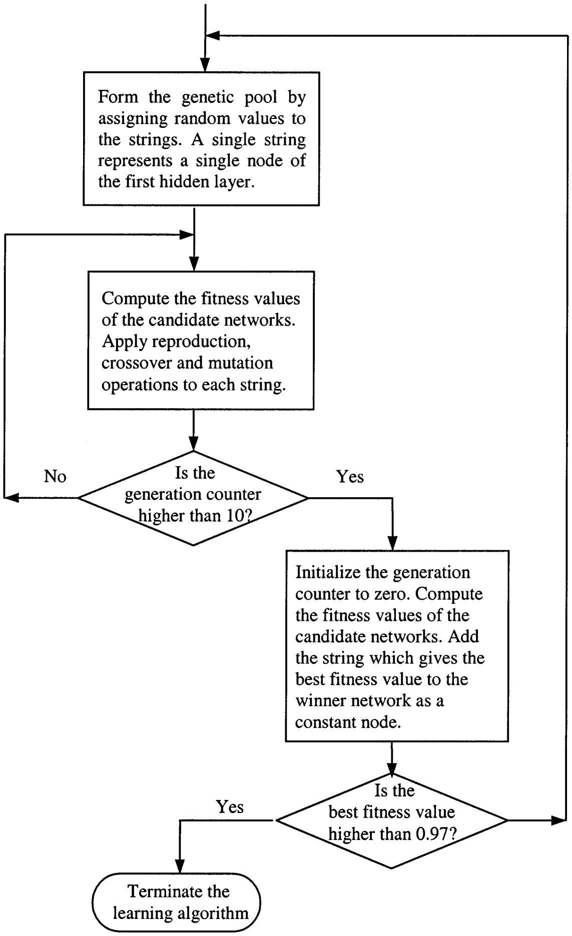

Fig. 6. Learning algorithm of the InS network.

to the fitness value are produced in the pool. And then, crossover and mutation operations are applied to each string. After ten generations, the string in the genetic pool, which gives the best fitness value, is added to the winner network as a constant node.

If the maximum fitness value is above a termination value, training algorithm is terminated. In the study, this value is chosen as 97%. The winner network is determined by the CN that gives the maximum fitness value.

Otherwise, the genetic pool is prepared to search for a new node. The w_{ji} and r_j parameters of the nodes are randomly chosen. The search is continued by using the CNs with increased number of nodes until the termination criterion is satisfied. Fig. 6 shows the training algorithm of the InS network.

4. Computer simulations and conclusion

In the study, topologies of the RCE and InS networks are determined after ten different trials with the same training set. For each network the topology, which represents the class distribution best in the feature space, is used through Sections 4.1 and 4.2. The classification results of MLP for different node configurations and learning rates are shown in Tables 1–3. Three thousand iterations are used to get the results in these tables.

The numbers of nodes in RCE and InS networks are automatically determined during the training. However, the number of nodes of MLP presented in Tables 1–3 are estimated before the training. And trials took too long time to determine the number of nodes in each hidden layer.

Mutation rate and depth of the pool affect the convergence time to the optimum solution. Mutation is used to prevent the solution to be caught by local minima. As random walk may occur around the optimum solution, mutation rate must be kept at a low value. Depth of the pool must also be kept at a low value. Otherwise, the computational time for one generation will increase. In the paper, mutation rate and depth of the pool are experimentally determined as 0.02 and 16, respectively.

All the simulations are performed on Pentium III–450 MHz PC using MATLAB 5.1 and Visual C.

4.1. Classification of the sample space

2D-sample space is used to demonstrate the classification performances of the networks. It is not related to the classification of the ECG signals. Fig. 7(a) shows the classification result of RCE network obtained by using 29 nodes. Black color shows vectors (points) that belong to more than one class. Fig. 7(b) shows the classification result of InS network obtained by using six nodes. Fig. 7(c) shows the classification result of MLP obtained by using 70 nodes in the first hidden

Table 1
Nine different topologies of MLP for the classification of 2D sample space

Structure[a] *I*	*F*	*S*	*O*	Training time (min:s)	Learning rate	Error
2	70	80	3	14:30	0.01	27.2/300
2	70	80	3	14:24	0.005	19.24/300
2	70	80	3	14:10	0.002	1.1/300
2	60	70	3	11:04	0.01	18.48/300
2	60	70	3	10:25	0.005	14.49/300
2	60	70	3	10:25	0.002	7.17/300
2	40	50	3	7:00	0.01	35.86/300
2	40	50	3	6:40	0.005	24.04/300
2	40	50	3	6:38	0.002	18.78/300

[a] *I*, *F*, *S*, *O* letters stand for the number of nodes in input, first hidden, second hidden and output layers.

Table 2
Nine different topologies of MLP trained by wavelet coefficients

Structure[a] *I*	*F*	*S*	*O*	Training time (min:s)	Learning rate	Error
8	40	50	10	47:25	0.01	457/1500
8	40	50	10	47:15	0.005	150/1500
8	40	50	10	48:27	0.001	92.53/1500
8	30	40	10	33:10	0.01	951/1500
8	30	40	10	32:29	0.005	150/1500
8	30	40	10	33:05	0.001	80/1500
8	20	30	10	22:16	0.01	1360.7/1500
8	20	30	10	23:37	0.005	318/1500
8	20	30	10	23:26	0.001	120/1500

[a] *I*, *F*, *S*, *O* letters stand for the number of nodes in input, first hidden, second hidden and output layers.

Table 3
Nine different topologies of MLP trained by Fourier coefficients

Structure[a] *I*	*F*	*S*	*O*	Training time (min:s)	Learning rate	Error
8	40	50	10	48:10	0.01	1440/1500
8	40	50	10	47:00	0.005	701/1500
8	40	50	10	49:00	0.001	161/1500
8	30	40	10	33:51	0.01	1428/1500
8	30	40	10	36:36	0.005	412/1500
8	30	40	10	34:00	0.001	149/1500
8	20	30	10	22:43	0.01	1468/1500
8	20	30	10	23:00	0.005	322/1500
8	20	30	10	23:05	0.001	179/1500

[a] *I*, *F*, *S*, *O* letters stand for the number of nodes in input, first hidden, second hidden and output layers.

Table 4
Sensitivity, PP, TCA, learning time (LT) and number of nodes (NN) of MLP, RCE and InS networks

ECG beats	RCE				InS				MLP			
	DWT		DFT		DWT		DFT		DWT		DFT	
	Sens.	PP	Sens.	PP	Sens.	PP	Sens.	PP	Sens.	PP	Sens.	PP
Beat N	0.932	0.860	0.863	0.731	1.0	0.967	0.773	0.743	1.0	0.926	0.893	0.937
beat L	0.391	0.722	0.608	0.38	0.946	0.91	0.78	0.704	0.48	0.96	0.813	0.772
beat R	0.784	0.739	0.47	0.736	0.986	0.942	0.82	0.702	0.746	0.991	0.886	0.734
beat P	0.945	0.478	0.703	0.672	0.946	0.891	0.66	0.611	0.9	0.937	0.74	0.822
beat p	0.917	1.0	0.261	0.853	0.986	0.913	0.98	0.88	1.0	0.887	0.973	0.948
beat a	0.63	0.963	0.85	0.922	1.0	0.98	0.946	0.835	0.993	0.788	0.96	0.913
beat E	0.517	0.983	0.333	0.913	0.946	0.888	0.82	0.848	0.953	0.836	0.793	0.815
beat V	0.865	0.545	0.776	0.606	0.913	0.935	0.666	0.757	0.986	0.813	0.92	0.762
beat F	0.7	0.656	0.669	0.35	0.866	0.991	0.46	0.76	0.753	0.849	0.493	0.891
beat f	0.438	0.735	0.806	0.937	0.983	0.901	0.913	0.878	0.946	0.876	1.0	0.949
NN	266		317		20		24		8-30-40-10		8-30-40-10	
TCA (%)	60		50.1		95.7		78.2		87.6		84.7	
LT	15 s		20 s		13 min		15 min		33 min		34 min	

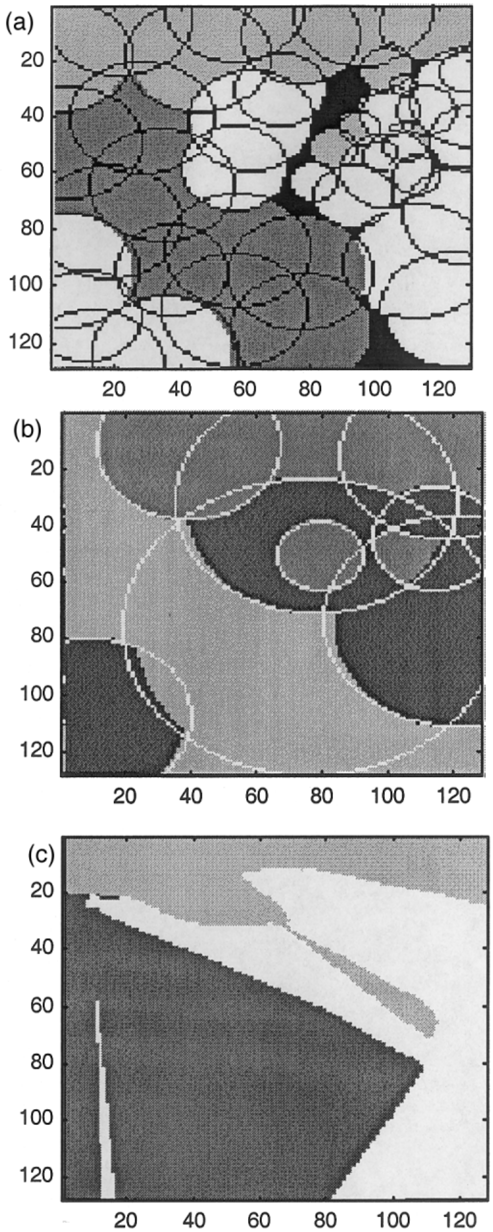

Fig. 7. Classification results of (a) RCE, (b) InS and (c) MLP for the 2D sample feature space. Circles in (a) and (b) are the geometric representations of the nodes of RCE and InS in 2D.

layer, 80 nodes in the second hidden layer and three nodes in the output layer.

Table 1 shows the training results of the MLP with different topologies for 2D sample space. In two dimensions, computer simulations show that InS gives better classification accuracy with less number of nodes compared to the RCE and MLP networks

4.2. Classification of the ECG signals

The waveforms of ten different ECG beats classified in the study are shown in Fig. 8a–j: normal beat (N), left bundle branch block beat (L), right bundle branch block beat (R), premature ventricular contraction (V), paced beat (P), non-conducted P wave (p), aberrated atrial premature beat (a), fusion of ventricular and normal beat (F), fusion of paced and normal beat (f), and ventricular escape beat (E). Test and training sets are separately formed by choosing 1500 vectors (150 vectors from each class) with eight dimensions. Training and test sets are formed by data obtained from records #100, 106, 107, 109, 111, 118, 124, 201, 202, 207, 208, 210, 212, 213, 214, 217, 219, 220, 221, 231 of MIT-BIH database and a real-time ECG measurement system. In order to enhance the generalization capability of the neural network, the training and the test sets are formed by data obtained from different patients. It is observed that for some beat types, there are waveform variations among the vectors belonging to the same class, and some beat types do not even contain P waves or ST segments.

Fig. 9 shows the relation between the divergence value and dimension of the feature vectors formed by DWT and DFT. There is not a significant increase in the divergence value above eight dimensions. Therefore, in the two feature extraction methods, classification performance will slightly increase by including new features to the eight dimensional vectors. Increasing the dimension will cause the computational load and the number of nodes used in the network to increase.

Tables 2 and 3 show the training results of MLP with different topologies for ECG signals. The numbers of nodes of RCE and InS networks are automatically determined by their own training algorithms.

Three statistics: Sensitivity, positive predictivity (PP) and total classification accuracy (TCA) are used to compare the results in Tables 4 and 5. The respective definitions are as follows:

$$\text{Sensitivity} = \#\{\mathrm{TP}_i\}/(\#\{\mathrm{TP}_i\} + \#\{\mathrm{FN}_i\})$$

$$\text{Positive predictivity} = \#\{\mathrm{TP}_i\}/(\#\{\mathrm{TP}_i\} + \#\{\mathrm{FP}_i\})$$

$$\mathrm{TCA} = \sum_{j=1}^{10} \frac{\#\{\mathrm{TP}_j\}}{\#\{\mathrm{Tr}\}} \tag{6}$$

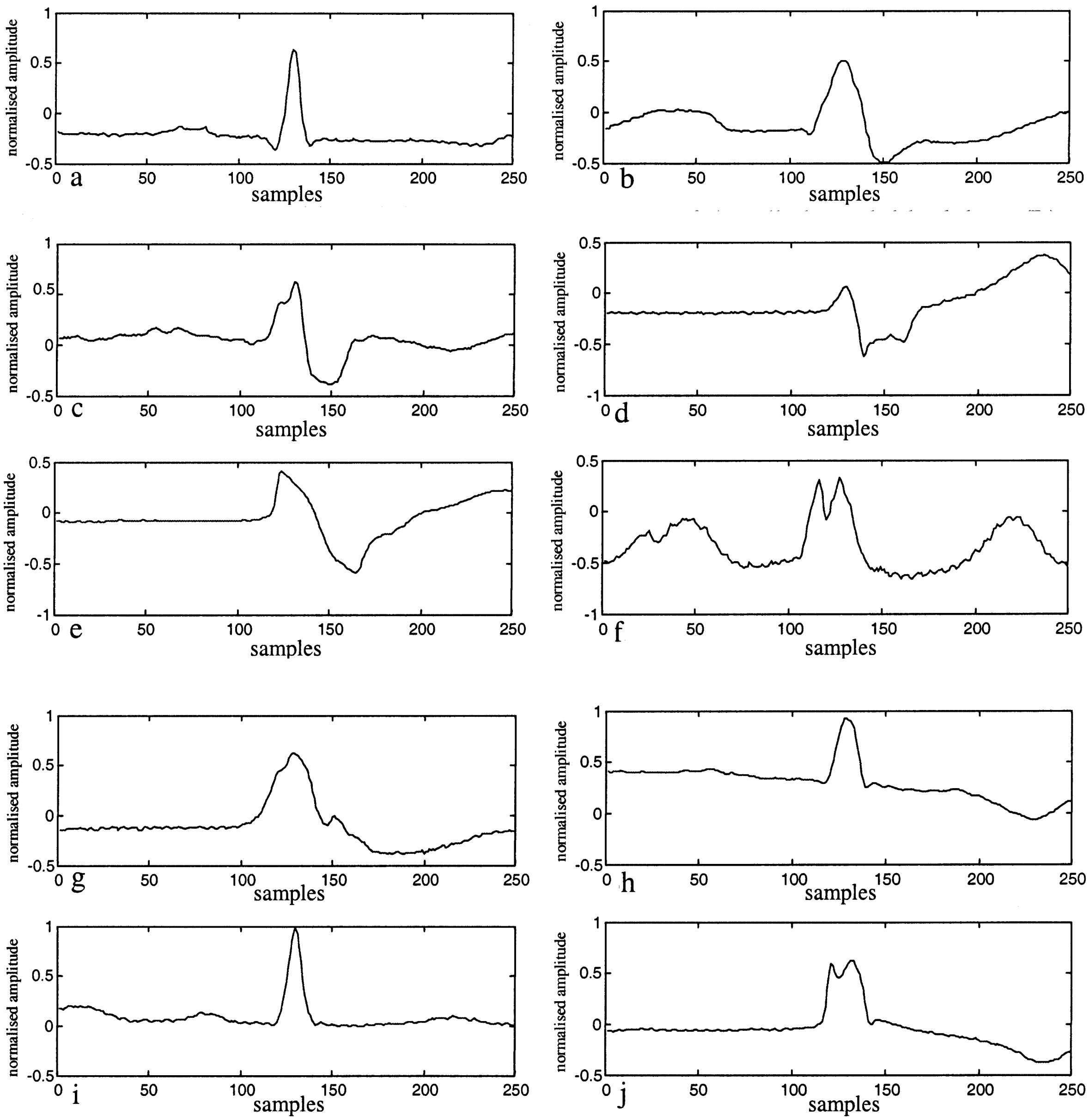

Fig. 8. a: Normal beat (N); b: Left bundle branch block beat (L); c: Right bundle branch block beat (R); d: Premature ventricular contraction (V); e: Paced beat (P); f: Aberrated atrial premature beat (a); g: Ventricular escape beat (E); h: Non-conducted P wave (p); i: Fusion of ventricular and normal beat (F); and j: Fusion of paced and normal beat (f).

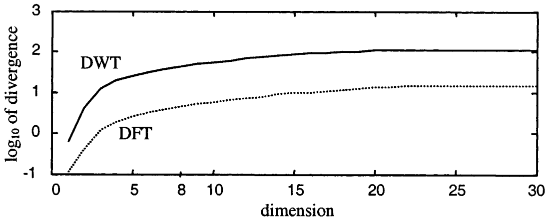

Fig. 9. The relation between the divergence value and dimension.

where $\#\{TP_i\}$ is the number of true positives which are the right classifications of the ith class; $\#\{FP_i\}$ and $\#\{FN_i\}$ are the number of discrepancies of the ith class, i.e. errors in computer findings; $\#\{Tr\}$ is the number of all beats in the training set [3].

Sensitivity is the fraction of i-type beats that are correctly classified among all i-type beats labelled by the experts. PP is the fraction of real i-type beats among all beats classified as i-type by computer.

The numbers of nodes generated, learning times, and classification performances of MLP, RCE and InS networks are shown in Table 4. From the table, it is observed that the sensitivity of the beat-L classified by the RCE is 0.391. As seen from Fig. 7(a), the learning algorithm of the RCE does not locate hyperspheres to optimum positions, leading to the generation of non-decision regions (black color). For this reason, low performances are achieved by the RCE in the classification of the sample space and ECG beats. Computer simulations show that InS gives better classification accuracy with less number of nodes and better classification results are produced by using the feature extraction method based on wavelet transform.

In order to increase the number of vectors in the test set and to analyse the effects of the noise on the classification performance, a new test set is formed by 15 000 vectors (1500 vectors from each class) with eight dimensions. Table 5 shows the classification performance of InS network for noisy ECG data. The vectors of the new test set are formed by adding white noise to the original ECG signal. SNR ($10\log(\sigma_{signal}/\sigma_{noise})$) rates of each beat type are shown in Table 5. It is observed that classification performance for the N, P and F beats decreases in the existence of noise and the wavelet transform gives higher classification performance compared to the Fourier transform.

4.3. Conclusion

It is important to increase the number of categories (beat classes) to enhance the analysis of ECG

Table 5
Sensitivity, PP and TCA of the InS network for two sets of ECG data with noise variances of 8.3 and 33.3

ECG beats	InS network									
	DWT		DFT		SNR $\sigma_{n1} = 8.3$	DWT		DFT		SNR $\sigma_{n2} = 33.3$
	Sens.	PP	Sens.	PP		Sens.	PP	Sens.	PP	
Beat N	0.722	0.931	0.761	0.646	12.8	0.646	0.904	0.753	0.494	6.8
Beat L	0.899	0.81	0.7	0.495	16.1	0.843	0.705	0.493	0.312	10.1
Beat R	0.88	0.892	0.622	0.668	14.7	0.858	0.835	0.438	0.591	8.7
Beat P	0.808	0.864	0.616	0.502	17.2	0.766	0.816	0.533	0.408	11.1
Beat p	0.976	0.902	0.924	0.729	16.4	0.97	0.829	0.838	0.628	10.4
Beat a	0.972	0.949	0.576	0.817	17.8	0.896	0.918	0.19	0.637	11.8
Beat E	0.873	0.897	0.607	0.809	16.1	0.848	0.833	0.372	0.762	10.1
Beat V	0.897	0.675	0.705	0.625	15.4	0.873	0.659	0.708	0.409	9.3
Beat F	0.81	0.929	0.272	0.546	17.8	0.485	0.866	0.173	0.34	11.9
Beat f	0.958	0.895	0.756	0.834	15.1	0.945	0.879	0.406	0.826	9.1
TCA (%)	88		65.4			81.3		49		

signals. However, as the number of categories increases, performance of class separation decreases. In this study, a new neural network structure and a new feature extraction method based on wavelet transform are developed to increase both the number of categories and the performance of class separation.

All ECG beats are obtained from standard bipolar derivation-II. The use of single derivation decreases the complexity of the system. Decision making is performed in three stages comprising normalization process, feature extraction method, and artificial neural network as the classifier. Each stage is thoroughly investigated to enhance the performance of the system. Since three processes are realized for each derivation, computational time and the complexity of decision making will increase with the use of extra derivations.

Decision making process is affected by the peak-to-peak magnitudes, the offset of the signal, and R-peak position in the windowed ECG. Age and sex of the patient, electrode contact rezistance and the settings of the measurement system will affect the ECG signal. These effects are eliminated by the normalization process. Normalization decreases the dependence of the classification task on the training set. If a normalization process is not performed, decision making will higly depend on the training set. As a consequence, ECG beats that do not exist in the training set may not be correctly classified.

The performance results presented in Table 5 are obtained without a filtering process. In case of classification of the ECG beats with high noise variance, filtering is required prior to normalization process.

In the study, divergence analysis enables us to order the features in the vectors according to their importance in class separation, and gives information about the distribution of the class vectors. As the class vectors scatter, the divergence value decreases. In this study, the dimension of the feature vectors is decreased to a value (8) lower than the dimension value (15) in the previous study [14] because of the high generalization ability of the new InS network. In order to analyse the effects of decreasing the dimension of the feature vectors, training and test sets are grown from 1000 to 1500 and 15 000. These sets are enhanced by data obtained from different patients. Simulations showed that high classification performances are achieved by the new feature vectors with eight dimensions even in existence of noise, and reduction of the dimension to a value lower than eight causes the number of nodes to increase and the classification performance of the InS network to decrease.

In the literature, it is observed that each neural network was trained by its own special algorithm. In the study, genetic algorithms are used to determine the weights and the number of the nodes in the first layer of the network. GAs gave satisfactory results in the training of the InS network. By using the InS network, a classification performance of 96% is achieved for ten different categories of ECG beats.

References

[1] G. Bortolan, S. Fusaro, Feature reduction and RBF in classifiers based on ANN, 18th Annual International Conference of the IEEE-EMBS (1996) paper no. 819.

[2] P. Chazal, B.G. Celler, Selecting a neural network structure for ECG diagnosis, 20th Annual International Conference of the IEEE-EMBS (1998) 1422–1425.

[3] Y.H. Hu, S. Palreddy, W.J. Tompkins, A patient-adaptable ECG beat classifier using a mixture of experts approach, IEEE Trans. Biomed. Eng. 44 (9) (1997) 891–900.

[4] S. Palreddy, W.J. Tompkins, Y.H. Hu, Customization of ECG beat classifiers developed using SOM and LVQ, 17th Annual International Conference of the IEEE-EMBS, (1995) paper no. 778.

[5] L. Reinhardt, R. Vesanto, J. Montonen, T. Fetsch, M. Makijarvi, G. Sierra, G. Breithardt, Application of learning vector quantization for localization of myocardial infarction, 18th Annual International Conference of the IEEE-EMBS, (1996) paper no. 211.

[6] M.R. Risk, J.F. Sobh, J.P. Saul, Beat detection and classification of ECG using self organizing maps, 19th Annual International Conference of the IEEE-EMBS (1997) 89–91.

[7] R. Silipo, C. Marchesi, Artificial neural networks for automatic ECG analysis, IEEE Trans. Signal Processing 46 (5) (1998) 1417–1425.

[8] R. Silipo, M. Gori, A. Taddei, M. Varanini, C. Marchesi, Classification of arrhythmic events in ambulatory electrocardiogram, using artificial neural networks, Comput. Biomed. Res. 28 (4) (1995) 305–318.

[9] R. Silipo, G. Bortolan, C. Marchesi, Supervised and unsupervised learning for diagnostic ECG classification, 18th Annual International Conference of the IEEE-EMBS (1996) paper no 1054.
[10] K. Yonghong, J. Jie, H. Yecho, L. Zhicheng, Paced ECG analysis based on neural networks, 20th Annual International Conference of the IEEE-EMBS (1998) 206–209.
[11] M.H. Fredric, H. Soowhan, Classification of cardial arrhythmias, IEEE Trans. Biomed. Eng. 43 (4) (1996) 425–430.
[12] J.S. Paul, M.R.S. Reddy, V.J. Kumar, Automatic detection of PVC's using autoregressive models, 19th Annual International Conference of the IEEE-EMBS (1997) 68–71.
[13] B.G.Celler, P. Chazal, Low computational cost classifiers for ECG diagnosis using neural networks, 20th Annual International Conference of the IEEE-EMBS (1998) 1337–1340.
[14] Z. Dokur, T. Ölmez, E. Yazgan, Comparison of discrete wavelet and Fourier transforms for ECG beat classification, Electron. Lett. 35 (18) (1999) 1502–1504.
[15] I. Jouny, P. Hamilton, M. Kanapathipillai, Adaptive wavelet representation and classification of ECG signals, 16th Annual International Conference of the IEEE-EMBS, (1994) paper no. 48.
[16] J. Yao, Q. Gan, X. Zhang, J. Li, Pruning algorithm in wavelet neural network for ECG signal classification, 20th Annual International Conference of the IEEE-EMBS (1998) 1482–1485.
[17] Z. Dokur, T. Ölmez, E. Yazgan, O.K. Ersoy, Detection of ECG waveforms by neural networks, Med. Eng. Phys. 19 (8) (1997) 738–741.
[18] Z. Dokur, T. Ölmez, M. Korürek, E. Yazgan, Detection of ECG waveforms by using artificial neural networks, 18th Annual International Conference of the IEEE-EMBS, Amsterdam (1996) paper no.1038.
[19] Z. Dokur, T. Ölmez, M. Korürek, E. Yazgan, Detection of ECG waveforms by using fuzzy classifier, (in Turkish), 3rd Annual Conference of Signal Processing and App.-SIU'95, B: Signal Processing, Cappadocia, Turkey (1995) 208–213.
[20] K. Minami, H. Nakajima, T. Toyosima, Real-time discrimination of ventricular tachyarrhythmia with Fourier–Transform neural network, IEEE Trans. Biomed. Eng. 46 (2) (1999) 179–185.
[21] T. Ölmez, Z. Dokur, E. Yazgan, Classification of ECG waveforms using a novel neural network, 20th Annu. Int. Conf. IEEE-EMBS Hong Kong 20 (3) (1998) 1616–1619.
[22] T. Ölmez, Z. Dokur, E. Yazgan, Classification of ECG waveforms by using genetic algorithms, 19th Annual International Conference of the IEEE-EMBS, Chicago, USA (1997) 92–94.
[23] A. Cohen, Biomedical Signal Processing, vol. II, CRC Press, Boca Raton, Florida, 1986, pp. 75–79.
[24] A. Elias, L. Leija, C. Alvarado, P. Hernández, A. Gutiérrez, Personal computer system for ECG recognition in myocardial infarction diagnosis based on an artificial neural network, 19th Annual International Conference of the IEEE-EMBS (1997) 1095–1096.
[25] S.M. Szilágyi, Event recognition, seperation and classification from ECG recordings, 20th Annual Int. Conference of the IEEE-EMBS (1998) 236–239.
[26] O. Wieben, W.J. Tompkins, V.X. Afonso, Classification of PVCs with a fuzzy logic system, 19th Annual International Conference of the IEEE-EMBS (1997) 65–67.
[27] M. Yang, W. Hu, L. Shyu, ECG events detection and classification using wavelet and neural networks, 19th Annual International Conference of the IEEE-EMBS (1997) 280–281.
[28] MIT-BIH Arrhythmia Database. Available from Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 20A-113, Cambridge, MA 02139, USA, 1992.
[29] I. Daubechies, Ten Lectures on Wavelets, Capital City Press, Vermont, 1994.
[30] R.P. Lippmann, An introduction to computing with neural nets, IEEE ASSP Magazine April (1987) 4–22.
[31] 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.
[32] D.L. Reilly, L.N. Cooper, C. Elboum, Neural model for category learning, Biol. Cybernet. 45 (1982) 35–41.
[33] J. Makhoul, A.E. Jaraudi, R. Schwartz, Formation of disconnected regions with a single hidden layer, IJCNN Int. Joint Conf. Neural Networks 1 (1989) 455–460.
[34] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Longman, Inc., Massachusetts, 1989.
[35] J.H. Holland, Genetic algorithms, Scientific American July (1992) 44–50.