Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 33 (2007) 286–295
Expert Systems with Applications
www.elsevier.com/locate/eswa

Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network

Rahime Ceylan, Yüksel Özbay *

Department of Electrical and Electronics Engineering, Engineering and Architecture Faculty, Selcuk University, 42075 Konya, Turkey

Abstract

Principal component analysis (PCA) and wavelet transform (WT) are two powerful techniques for feature extraction. In addition, fuzzy c-means clustering (FCM) is among considerable techniques for data reduction. In other words, the aim of using FCM is to decrease the number of segments by grouping similar segments in training data. In this paper, four different structures, FCM-NN, PCA-NN, FCM-PCA-NN and WT-NN, are formed by using these two techniques and fuzzy c-means clustering. In addition, FCM-PCA-NN is the new method proposed in this paper for classification of ECG. This paper presents a comparative study of the classification accuracy of ECG signals by using these four structures for computationally efficient early diagnosis. Neural network used in this study is a well-known neural network architecture named as multi-layered perceptron (MLP) with backpropagation training algorithm. The ECG signals taken from MIT-BIH ECG database, are used in training to classify 10 different arrhythmias. These are normal sinus rhythm, sinus bradycardia, ventricular tachycardia, sinus arrhythmia, atrial premature contraction, paced beat, right bundle branch block, left bundle branch block, atrial fibrillation and atrial flutter. Before testing, the proposed structures are trained by backpropagation algorithm. All of the structures are tested by using experimental ECG records of 92 patients (40 male and 52 female, average age is 39.75 ± 19.06). The test results suggest that FCM-PCA-NN structure can generalize better than PCA-NN and is faster than NN, FCM-NN and WT-NN.

Keywords: Principal component analysis; Wavelet transform; Fuzzy c-means clustering; ECG; Arrhythmia; Artificial neural network

1. Introduction

Electrocardiography deals with the electrical activity of the heart. Monitored by placing sensors at the limb extremities of the subject, electrocardiogram (ECG) is a record of the origin and the propagation of the electrical potential through cardiac muscles. It is considered a representative signal of cardiac physiology, useful in diagnosing cardiac disorders (Acharya, Bhat, Iyengar, Roo, & Dua, 2002; Osowski & Linh, 2001).

The state of cardiac heart is generally reflected in the shape of ECG waveform and heart rate. It may contain important pointers to the nature of diseases afflicting the heart. However, bio-signals being non-stationary signals, the reflection may occur at random in the time-scale (that is, the disease symptoms may not show up all the time, but would manifest at certain irregular intervals during the day). Therefore, for effective diagnostics, the study of ECG pattern and heart rate variability signal may have to be carried out over several hours. Thus, the volume of the data being enormous, the study is tedious and time consuming. Naturally, the possibility of the analyst missing (or misreading) vital information is high. Therefore, computer-based analysis and classification of diseases can be very helpful in diagnostics (Acharya et al., 2002). Several algorithms have been developed in the literature for detection and classification of ECG beats. Most of them use either time or frequency domain representation of the ECG waveforms, on the basis of which many specific features are defined, allowing the recognition between the beats

* Corresponding author. Tel.: +90 332 223 20 48; fax: +90 332 241 06 35.

E-mail address: yozbay@selcuk.edu.tr (Y. Özbay).

0957-4174/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2006.05.014

belonging to different classes. The most difficult problem faced by today's automatic ECG analysis is the large variation in the morphologies of ECG waveforms, not only of different patients or patient groups but also with in the same patient. The ECG waveforms may differ for the same patient to such extend that they are unlike to each other and at the same time alike for different types of beats. This is the main reason that the beat classifier, performing well on the training data generalizes poorly, when presented with different patients ECG waveforms (Osowski & Linh, 2001).

One of the methods of ECG beat recognition is neural network classification method (Engin & Demirağ, 2003; Foo, Stuart, Harvey, & Meyer-Baese, 2002; Haykin, 1994; Jang, Sun, & Mizutani, 1997; Karlık, Tokhi, & Alcı, 2003; Ozbay, 1999; Ozbay & Karlık, 2001; Pilla & Lopes, 1999). MLP (Multi-layer perceptron) (Castellano & Fanelli, 2000; Castellano & Fanelli, 2000; Dazzi et al., 2001; De, Basak, & Pal, 2002; Fan, Zhen, & Xie, 2003; Hall et al., 1992; Haseyama & Kitajima, 2001; Jang et al., 1997; Leski & Czogala, 1999; Liao, Celmins, & Hammell, 2003; Li, Mukaidono, & Turksen, 2002; Meesad & Yen, 2000; Pal, Bezdek, & Tsao, 1993; Pedrycz, 1998; Pektatlı, Ozbay, Ceylan, & Karlık, 2003; Ronen, Shabtai, & Guterman, 1998; Sebzalli & Wang, 2001; Seker, Evans, Aydın, & Yazgan, 2001; Shaout & Scharbonea, 2000; Stoeva & Nikov, 2000; Yang & Liu, 2003; Zhang & Kandel, 1998), which can be called "conventional BPNN (backpropagation neural networks)", has been shown to be able to recognize and classify ECG signals more accurately. For example Ozbay et al. and Foo et al. have used neural networks (Ozbay & Karlık, 2001; Ozbay, 1999; Foo et al., 2002). However, conventional BPNN suffers from slow convergence to local and global minima and from random settings of initial values of weights, which may make the neural networks have very poor mappings from inputs to outputs. So, researchers have started to use hybrid structure. For example Osowski and Linh (2001), Pilla and Lopes (1999) and Engin and Demirağ (2003) have used fuzzy hybrid neural network.

This paper proposes using fuzzy c-means clustering algorithm, principal component analysis and wavelet transform to make a neural network system more effective. The structure proposed in this paper, FCM-PCA-NN, is composed of three subnetworks: fuzzy classifier, layer of feature extraction with principal component analysis and neural network. The fuzzy self-organizing layer performs the preclassification task, layer of feature extraction performs elimination of inconsiderable features with principal component analysis and the following multilayer perceptron works as a final classifier. The fuzzy stage is responsible for the analysis of the distribution of data and grouping them into clusters with different membership values. After that, features of obtained clustered training patterns are extracted using principal component analysis. On the basis of these principal components, the MLP network classifies the applied input vector, representing the heart beat to the appropriate class. On the other hand, a number of segments in training patterns are reduced using fuzzy c-means clustering in fuzzy self-organizing layer and number of samples of clustered training patterns is decreased with principal component analysis, before inputs are presented to MLP. Therefore, training period of the neural network is decreased. This method proposed in this paper was used for electrocardiographic beat recognition and classification.

2. Material and methods

For recognition of the ECG waveform type, different solutions were presented in the literature, such as the MLP approach, self organizing map and the LVQ. We present the combination of the fuzzy self-organizing layer, principal component analysis and the MLP connected in cascade, named as the FCM-PCA-NN and then compare this technique with other techniques (PCA-NN, WT-NN) in the literature (Güler & Übeyli, 2005; Wang & Paliwal, 2003).

The self-organizing layer is responsible for the clustering of the input data. However, it is the fuzzy clustering, in which the input vector x is preclassified to all sets with different membership values. The penetration of the data space is better and the localization of the input vector x in the data space is more precise. The outputs of all self-organizing neurons (the cluster centers) form the input vector to principal component analysis. The output data vectors of principal component analysis form the input vector to the third subnetwork (MLP). MLP subnetwork is responsible for the final classification of the ECG beat.

2.1. The fuzzy c-means clustering

Structure identification of fuzzy systems is possible by constructing enough rules with appropriate input and output membership functions. The identified model can then be used to describe the behaviour of the target system as well as for prediction purpose. In this paper, we have trained the fuzzy layer by using the fuzzy c-means clustering algorithm.

The idea of fuzzy clustering is to divide the data into fuzzy partitions that overlap with one another. Therefore, the inclusion of data in a cluster is defined by a membership grade in [0, 1]. Formally, clustering an unlabeled data $X = \{x_1, x_2, \ldots, x_N\} \subset R^h$, where N represents the number of data vectors and h the dimension of each data vector, is the assignment of c-partition labels to the vectors in X. c-partition of X constitutes sets of $(c \cdot N)$ $\{u_{ik}\}$ membership values that can be conveniently arranged as a $(c \times N)$ matrix $U = [u_{ik}]$. The problem of fuzzy clustering is to find the optimum membership matrix U. The most widely used objective function for fuzzy clustering is the weighted within-groups sum of squared errors J_m, which is used to define the following constrained optimisation problem (Castellano & Fanelli, 2000; Castellano & Fanelli, 2000; Dazzi et al., 2001; De et al., 2002; Fan et al., 2003; Jang et al., 1997; Li et al., 2002; Liao et al., 2003; Meesad & Yen, 2000; Pektatlı et al., 2003; Seker et al., 2001; Zhang & Kandel, 1998)

$$\min\left\{J_m(U,V,X)=\sum_{k=1}^{N}\sum_{i=1}^{c}(u_{ik})^m\|x_k-v_i\|_A^2\right\} \tag{1}$$

where

$$U\in M_{fcn}=\left\{U\in\Re^{cN}\left|\begin{array}{l}0\leqslant u_{ik}\leqslant 1\forall ik\quad\text{and }\forall k,u_{ik}>0\exists i\\ 0<\sum\limits_{k=1}^{N}u_{ik}>\eta\forall i\quad\text{and }\sum\limits_{i=1}^{c}u_{ik}=1\forall k\end{array}\right.\right\}.$$

$V=\{v_1,v_2,\ldots,v_c\}$ is the vector of (unknown) cluster centers, and is $\|x\|_A=(x^TAx)^{1/2}$ an inner product norm. A is an $h\times h$ positive definite matrix, which specifies the shape of the clusters. The matrix A is commonly selected as the identity matrix, leading to Euclidean distance and, consequently, to spherical clusters.

Fuzzy partitions are carried out using the fuzzy c-means (FCM) algorithm through an iterative optimisation of (Fan et al., 2003; Li et al., 2002; Liao et al., 2003; Jang et al., 1997; Seker et al., 2001; Pektatlı et al., 2003; Zhang & Kandel, 1998) according to the following steps (Jang et al., 1997):

Step (1) Choose the number of clusters (c), weighting exponent (m), iteration limit ($iter$), termination criterion ($\varepsilon>0$), and norm for error $\|V_t-V_{t-1}\|$.

Step (2) Guess initial position of cluster centers:

$$V_0=\{v_{1,0},v_{2,0},\ldots,v_{c,0}\}\subset\Re^{ch}.$$

Step (3) Iterate for $t=1$ iter, calculate

$$u_{ik,t}=\left[\sum_{j=1}^{c}\left(\frac{\|X_k-V_{i,t-1}\|_A}{\|X_k-V_{j,t-1}\|_A}\right)^{2/m-1}\right]^{-1} \tag{2}$$

and

$$V_{i,t}=\frac{\sum_{k=1}^{N}(u_{ik,t})^m x_k}{\sum_{k=1}^{N}(u_{ik,t})^m} \tag{3}$$

IF error $=\|V_t-V_{t-1}\|\leqslant\varepsilon$, THEN stop, and put $(U_f,V_f)=(U_t,V_t)$ for NEXT t.

2.2. Principal component analysis (PCA)

PCA is a well-established technique for feature extraction and dimensionality reduction. It is based on the assumption that most information about classes is contained in the directions along which the variations are the largest. The most common derivation of PCA is in terms of a standardised linear projection which maximises the variance in the projected space (Özbay, Ceylan, & Karlik, 2006; Wang & Paliwal, 2003). For a given p-dimensional data set X, the m principal axes $T_1,T_2,\ldots,T_m$, where $1\leqslant m\leqslant p$, are orthonormal axes onto which the retained variance is maximum in the projected space. Generally, $T_1,T_2,\ldots,T_m$ can be given by the m leading eigenvectors of the sample covariance matrix $S=1/N\sum_{i=1}^{N}(x_i-\mu)^{\mathrm T}(x_i-\mu)$, where $x_i\in X$, μ is the sample mean and N is the number of samples, so that:

$$ST_i=\lambda_iT_i,\quad i\in 1,\ldots,m \tag{4}$$

where λ_i is the ith largest eigenvalue of S. The m principal components of a given observation vector $x\in X$ are given by

$$y=[y_1,y_2,\ldots,y_m]=\left[T_1^{\mathrm T}x,T_2^{\mathrm T},\ldots,T_m^{\mathrm T}\right]=T^{\mathrm T}x \tag{5}$$

The m principal components of x are decorrelated in the projected space. In multi-class problems, the variations of data are determined on a global basis, that is, the principal axes are derived from a global covariance matrix:

$$\widehat{S}=\frac{1}{N}\sum_{j=1}^{K}\sum_{i=1}^{N_j}(x_j-\hat{\mu})(x_j-\hat{\mu})^{\mathrm T} \tag{6}$$

where $\hat{\mu}$ is the global mean of all the samples, K is the number of classes, N_j is the number of samples in class j; $N=\sum_{j=1}^{K}N_j$ ve x_{ji} represents the ith observation from class j. The principal axes T_1, $T_2,\ldots,T_m$ are therefore the m leading eigenvectors of $\widehat{S}$:

$$\widehat{S}T_i=\hat{\lambda}_iT_i,\quad i\in 1,\ldots,m \tag{7}$$

where $\hat{\lambda}_i$ is the ith largest eigenvalue of $\widehat{S}$. An assumption made for feature extraction and dimensionality reduction by PCA is that most information of the observation vectors is contained in the subspace spanned by the first m principal axes, where $m<p$. Therefore, each original data vector can be represented by its principal component vector with dimensionality m (Wang & Paliwal, 2003).

2.3. Wavelet transform

The ECG signals are considered as representative signals of cardiac physiology, which are useful in diagnosing cardiac disorders. The most complete way to display this information is to perform spectral analysis. 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 (Güler & Übeyli, 2005). 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. This feature of using a smaller number of parameters to represent the ECG signal is particularly important for recognition and diagnostic purposes. The WT can be thought of as an extension of the classic Fourier transform, except that, instead of working on a single scale (time or frequency), it works on a multi-scale basis. This multi-scale feature of the WT allows the decomposition of a signal into a number of scales, each scale representing

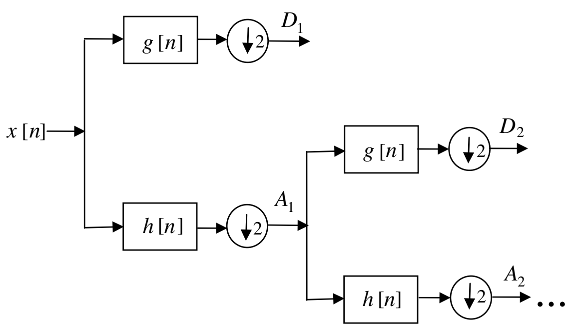

Fig. 1. Subband decomposition of discrete wavelet transform implementation; $g[n]$ is the high-pass filter, $h[n]$ is the low-pass filter.

a particular coarseness of the signal under study. The procedure of multiresolution decomposition of a signal $x[n]$ is schematically shown in Fig. 1. 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 down sampled outputs of first high- and low-pass filters provide the detail, $D1$ and the approximation, $A1$, respectively. The first approximation, $A1$ is further decomposed and this process is continued as shown in Fig. 1 (Güler & Übeyli, 2005).

Selection of appropriate wavelet and the number of decomposition levels is very important in analysis of signals using the WT. Usually, tests are performed with different types of wavelets and the one which gives maximum efficiency is selected for the particular application. The smoothing feature of the Daubechies wavelet of order 2 made it more suitable to detect changes of the ECG signals (Güler & Übeyli, 2005). Therefore, the wavelet coefficients were computed using the Daubechies wavelet of order 2 in the present study. The wavelet coefficients were computed using the MATLAB software package.

Selection of the ANN inputs is the most important component of designing the neural network based on pattern classification since even the best classifier will perform poorly if the inputs are not selected well. Input selection has two meanings: (1) which components of a pattern, or (2) which set of inputs best represent a given pattern. The computed discrete wavelet coefficients provide a compact representation that shows the energy distribution of the signal in time and frequency. Therefore, the computed detail and approximation wavelet coefficients of the ECG signals were used as the feature vectors representing the signals.

2.4. The multilayer perceptron

In this study, a three-layered feed-forward NN was used and trained with the error backpropagation. The input signals of NN were formed by cluster centers that generalized by fuzzy c-means clustering. Fig. 2 shows a general structure of the NN. The backpropagation training with generalized delta learning rule is an iterative gradient algorithm designed to minimize the root mean square error between the actual output of a multilayered feed-forward NN and a desired output. Each layer is fully connected to the previous layer, and has no other connection (Haykin, 1994; Karlık et al., 2003).

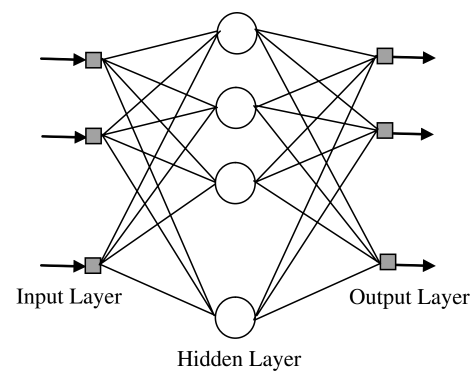

Fig. 2. The general structure of MLP NNs.

2.4.1. Backpropagation algorithm

1. *Initialisation*: Set all the weights and biases to small real random values.
2. *Presentation of input and desired outputs*: Present the input vector $x(1), x(2), \ldots, x(N)$ and corresponding desired response $d(1), d(2), \ldots, d(N)$, one pair at a time, where N is the number of training patterns.
3. *Calculation of actual outputs*: Use Eq. (8) to calculate the output signals $y_1, y_2, \ldots, y_{N_M}$

$$y_i = \varphi\left(\sum_{j=1}^{N_{M-1}} w_{ij}^{(M-1)} x_j^{(M-1)} + b_i^{(M-1)}\right), \quad i = 1, \ldots, N_{M-1} \tag{8}$$

4. *Adaptation of weights* (w_{ij}) *and biases* (b_i):

$$\Delta w_{ij}^{(l-1)}(n) = \mu \cdot x_j(n) \cdot \delta_i^{(l-1)}(n) \tag{9}$$

$$\Delta b_i^{(l-1)}(n) = \mu \cdot \delta_i^{(l-1)}(n) \tag{10}$$

where

$$\delta_i^{(l-1)}(n) = \begin{cases} \varphi'(\mathrm{net}_i^{(l-1)})[d_i - y_i(n)], & l = M \\ \varphi'\left(\mathrm{net}_i^{(l-1)}\right) \sum_k w_{ki} \cdot \delta_k^{(l)}(n), & 1 \leqslant l \leqslant M \end{cases} \tag{11}$$

in which $x_j(n)$ = output of node j at iteration n, l is layer, k is the number of output nodes of neural network, M is output layer, and ϕ is the activation function. The learning rate is represented by μ. It may be noted here that a large value of the learning rate may lead to faster convergence but may also result in oscillation. In order to achieve faster convergence with minimum oscillation, a momentum term may be added to the basic weight updating equation.

After completing the training procedure of the neural network, the weights of MLP are frozen and ready for use in the testing mode.

3. The results of numerical experiments

A significant number of data used for applications are naturally available in representations that are difficult to learn. Transforming the data into a more appropriate representation can facilitate the learning process. For instance, using a smaller number of parameters, which are often called features, to represent the signal under study is particularly important for recognition and diagnostic purposes. Given any set of features for data representation, it is therefore important to estimate the difficulty of learning the underlying concepts using that training data. The learning system should then seek to transform the representations into a space that is easier for learning purposes (Güler & Übeyli, 2005). The studies in the literature show that the degree of difficulty in training a neural network is inherent in the given set of training examples. By developing a technique for measuring this learning difficulty, they devise a feature construction methodology that transforms the training data and attempts to improve both the classification accuracy and computational times of artificial neural network (ANN) algorithms. The fundamental notion is to organize data by intelligent preprocessing, so that learning is facilitated (Güler & Übeyli, 2005; Piramathu, Ragavan, & Shaw, 1998). To implement this aim, in this paper, a clustering and feature extraction based approach is adopted for ECG signal, which belongs to 10 different arrhythmias used as input to the NN (see Fig. 5).

3.1. Structure and training data

Training data of ECG arrhythmias used in this study was taken from MIT-BIH ECG Arrhythmias Database. Selected types of arrhythmias were normal sinus rhythm (N), sinus bradycardia (Br), ventricular tachycardia (VT), sinus arrhythmia (SA), atrial premature contraction (APC), paced beat (P), right bundle branch block (R), left bundle branch block (L), atrial fibrillation (A.Fib) and atrial flutter (A.Fl). Training patterns had been sampled at 360 Hz, so we arranged them as 200 samples in the intervals of R–R for all arrhythmias, which are called as a segment. Training patterns were formed by mixing from the arrhythmias preprocessed with respect to the order above. The size of the training patterns was 106 segments*200 samples.

In this study, four different structures were formed for classification of ECG signals by using principal component analysis, fuzzy c-means clustering and wavelet transform. These structures were FCM-NN, PCA-NN, FCM-PCA-NN and WT-NN, respectively. Here, principal component analysis and wavelet transform were used to implement feature extraction, fuzzy c-means clustering were used to decrease number of data segments.

In first structure, FCM-NN (Özbay et al., 2006), training patterns were clustered using fuzzy c-means clustering algorithm before the NN training (see Fig. 3). So, a number of segments in each type of arrhythmia were reduced by using fuzzy c-means clustering and the new training set, whose size was 67 segments*200 samples, was obtained. Here, processes of reducing segments were made separately for each type of arrhythmia. The number of segments for each type of arrhythmias in new training set and original training set is given in Table 1. Finally, the obtained new training set was classified using the NN. The backpropagation algorithm is used for training. The optimum number of hidden nodes was determined via experimentation. The experimental results show that the optimum number of hidden nodes was 48 with the highest classification accuracy of 99.9% for 10,000 iterations. Here, nodes of input layer, hidden layer and output layer were used as 200, 48 and 10, respectively. Learning rate and momentum constant were chosen as 1.0 and 0.2 in training via experimentation.

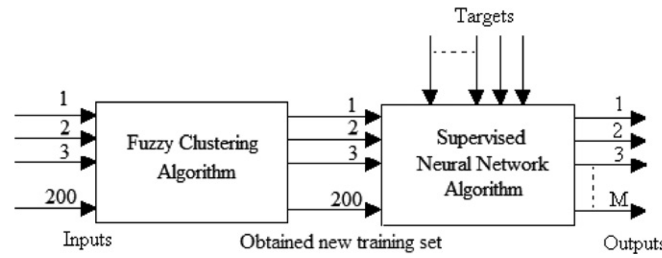

Fig. 3. FCM-NN architecture.

In PCA-NN structure, the number of samples of training patterns was reduced to make feature extraction with principal component analysis (see Fig. 4). It is shown in Fig. 7 that optimum number of principal component was found as 20. Obtained training patterns, whose size was 20 samples*106 segments, were presented to NN. So, the number of input nodes of NN was determined as 20. Furthermore, optimum neural network architecture was chosen as 20:30:10. Learning rate and momentum constant were chosen as 1.0 and 0.9 in training via experimentation.

In FCM-PCA-NN structure, the number of samples of training patterns, which were obtained to decrease the number of segments of training patterns with fuzzy c-means clustering, was reduced to make feature extraction

Table 1
The number of segments for each arrhythmia

Arrhythmia type	In original training set NN, PCA-NN, WT-NN	In new set obtained with FCM FCM-NN and FCM-PCA-NN
N	15	9
Br	15	9
VT	6	5
SA	15	9
APC	6	5
P	10	6
R	10	6
L	10	6
A. Fib	10	6
A.Fl	9	6
Total	106	67

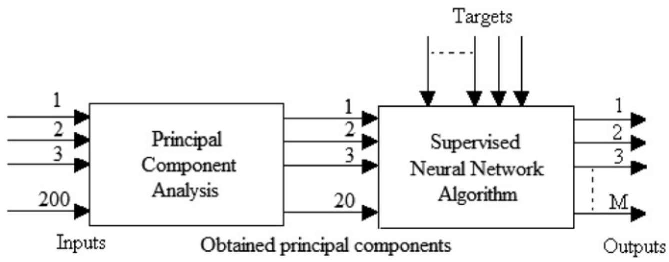

Fig. 4. PCA-NN architecture.

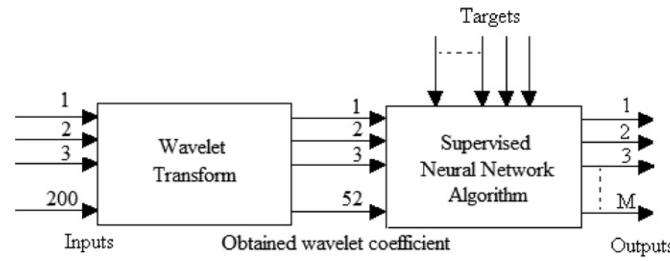

Fig. 6. WT-NN architecture.

with principal component analysis (see Fig. 5). It is shown in Fig. 7 that optimum number of principal component was found as 10. So, the number of input nodes of NN was determined as 10. However, optimum neural network architecture, learning rate and momentum constant were chosen as 10:30:10, 4.0 and 0.7 in training via experimentation, respectively.

In WT-NN structure, the number of samples of training patterns was reduced to make feature extraction with wavelet transform (see Fig. 6). Optimum number of input samples obtained with wavelet transform was found as 52. So, the number of input nodes of NN was determined as 52. In addition, optimum neural network architecture and learning rate, momentum constant were chosen as 52:70:10, 3.0 and 0.1 in training via experimentation, respectively.

3.2. Test results

Table 2 describes the test errors for each arrhythmia obtained with NN and other four structures. In Table 2, MCN and RMC are described as misclassification number and rate of misclassification, respectively. As noted, recognition rates vary between 98% and 99.9% with average recognition accuracy 99%. It follows from these results that the recognition rate obtained by PCA-NN, FCM-PCA-NN and WT-NN are better than results obtained by others.

Furthermore, NN was trained with backpropagation algorithm in four structures, separately. These structures and NN, trained by training set formed from MIT-BIH ECG Arrhythmias Database, were tested using the test ECG recorded from 92 patients (40 male and 52 female, average age is 39.75 ± 19.06). The ECG records were made in the Cardiology Department, Faculty of Medicine at Selcuk University in Konya, Turkey. Recording system given in the study of Ozbay and Karlık (2001), Özbay et al. (2006) is used for recording of ECG. These recorded patterns were normalised between 0 and 1 and were arranged as 200 samples in the intervals of R–R.

Table 3 shows the classification results obtained by NN and other four structures for test data recorded from patients. It can be seen in Table 3, noted recognition rates are approximately 99% for all architectures. Furthermore, if Table 3 is examined, it can be seen that FCM-PCA-NN is the best structure with high recognition rate in shorter time. Although test error is obtained as 0.4 in WT-NN, reached training time is 85.443 s. Detailed classification results for test data recorded from 92 patients can be seen in Tables 4–8. It is seen in Tables 4–8 that 945, 1093, 461, 1197 and 1828 segments are classified as normal sinus rhythm from 5342 segments in NN, FCM-NN, PCA-NN, FCM-PCA-NN and WT-NN, respectively.

3.3. Calculation of training and test errors

We found training errors given in tables and figures according to Eq. (12) (Özbay et al., 2006)

$$\text{Training error }(\%) = \left(\frac{\sum_{i=1}^{k}|t(i) - a(i)|}{m * n}\right) * 100 \qquad (12)$$

where $t(i)$ is desired outputs, $a(i)$ is outputs of neural network, k is the number of samples in training data (for example, in this study, for original training set, k is 21,200 samples, 106 segments∗200 samples), m is the number of segments in training data and n is the number of outputs of neural network.

We developed an algorithm for evaluation of test results. Desired values of node outputs of the output layer

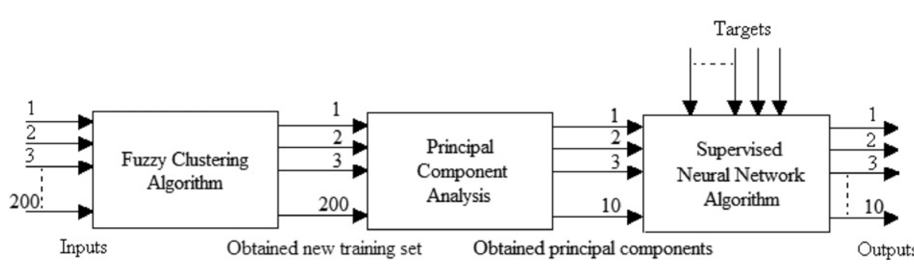

Fig. 5. FCM-PCA-NN architecture.

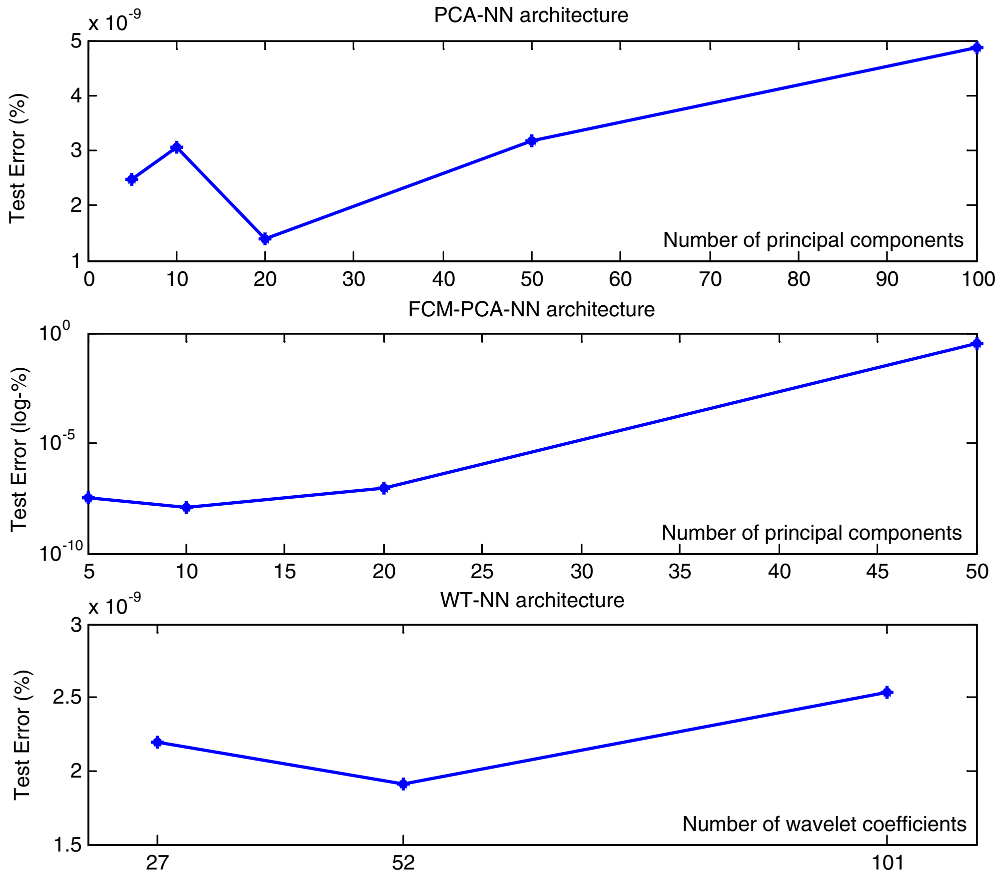

Fig. 7. Optimum number of principal components and wavelet coefficients for PCA-NN, FCM-PCA-NN and WT-NN architecture.

Table 2
Classification results for each arrhythmia in test

Arrhythmia type of test pattern	Number of segments	NN		FCM-NN		PCA-NN		FCM-PCA-NN		WT-NN	
		MCN	RMC (%)	MCN	RMC (%)	MCN	RMC (%)	MCN	RMC (%)	MCN	RMC (%)
N	15	0	0	0	0	0	0	0	0	0	0
Br	15	0	0	1	6.6	0	0	0	0	0	0
VT	6	0	0	0	0	0	0	0	0	0	0
SA	15	2	13.3	1	6.6	0	0	0	0	0	0
APC	6	1	16.6	1	16.6	0	0	0	0	0	0
P	10	0	0	0	0	0	0	0	0	0	0
R	10	0	0	0	0	0	0	0	0	0	0
L	10	0	0	0	0	0	0	0	0	0	0
A.Fib	10	0	0	0	0	0	0	0	0	0	0
A.Fl	9	0	0	0	0	0	0	0	0	0	0
Total	106	3	2.83	3	2.83	0	0	0	0	0	0
Average test error (%)		0.39		0.002		4.98×10^{-9}		5.05×10^{-9}		3.52×10^{-9}	

were logic-1 or logic-0 in the training pattern. Node outputs were changing between 0 and 1. If one of the node outputs of the output layer, $y(i) \geqslant 0.5$, we used $h(i) = |1 - y(i)|$ in the error calculation according to Eq. (13). Furthermore, $y(i) \geqslant 0.5$ and $y(i) >$ (other node outputs), we interpreted this as arrhythmia for the corresponding node. If $y(i) < 0.5$, we interpreted that there is no arrhythmia, and we used $h(i) = |0 - y(i)|$ according to Eq. (14). If all of node outputs were $y(i) < 0.5$, then an unknown state occurs. ANN did not classify this test pattern since similar pattern was not taught to the ANN beforehand. The unknown state is represented as "?" in Tables 4–8

$$y(i) \geqslant 0.5 \rightarrow h = \sum_{i=1}^{k} |1 - y(i)| \tag{13}$$

$$y(i) < 0.5 \rightarrow h = \sum_{i=1}^{k} |0 - y(i)| \tag{14}$$

Table 3
Classification results for 92 patients in test

No.	Architecture	Size of training set	Average training error (%)	Average test error (%)	Number of iteration	Training time (s)
1	NN	200*106	0.28	0.22	10,000	196.9
2	FCM-NN	200*67	9.92×10^{-10}	0.19	10,000	97.8
3	PCA-NN	20*106	7.07×10^{-9}	0.95	1094	7.571
4	FCM-PCA-NN	10*67	7.76×10^{-9}	0.91	2394	11.397
5	WT-NN	52*106	4.97×10^{-9}	0.4	6769	85.443

Table 4
The results from patient test data classified by NN

No.	M/F	Age	Samples	Number of segments	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9	Q_{10}	?	Error (%)
					N	Br	T	S	Apc	P	R	L	A.Fib	A.Flt		
1	M	23	11,200	56	25	0	0	0	0	0	0	0	3	0	28	0.0556
2	M	23	3600	18	0	0	0	0	0	0	0	0	0	0	18	4.82×10^{-5}
3	M	23	18,600	93	28	0	0	0	0	0	0	0	3	0	62	0.0555
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	F	16	13,800	69	0	0	0	0	0	0	0	0	0	0	69	8.3×10^{-4}
91	F	16	14,400	72	0	0	0	0	0	0	0	0	0	0	72	1.63×10^{-4}
92	F	16	17,400	87	0	0	0	0	0	0	80	0	0	0	7	0.6530
				5342	945	492	359	290	68	50	808	55	585	94	1596	0.22

Table 5
The results from patient test data classified by FCM-NN

No.	M/F	Age	Samples	Number of segments	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9	Q_{10}	?	Error (%)
					N	Br	T	S	Apc	P	R	L	A.Fib	A.Flt		
1	M	23	11,200	56	41	0	0	0	0	0	0	0	0	0	15	0.1535
2	M	23	3600	18	0	2	0	0	0	0	0	0	0	0	16	0.9822
3	M	23	18,600	93	67	0	0	0	0	0	0	0	0	0	26	0.3293
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	F	16	13,800	69	0	0	0	0	0	0	0	0	0	0	69	1.36×10^{-6}
91	F	16	14,400	72	0	0	0	0	0	0	0	0	0	0	72	0.0635
92	F	16	17,400	87	0	0	0	0	0	0	47	0	0	0	40	1.4100
				5342	1093	316	194	559	68	45	844	54	556	94	1519	0.19

Table 6
The results from patient test data classified by PCA-NN

No.	M/F	Age	Samples	Number of segments	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9	Q_{10}	?	Error (%)
					N	Br	T	S	Apc	P	R	L	A.Fib	A.Flt		
1	M	23	11,200	56	0	0	0	0	0	0	1	0	9	26	20	1.356
2	M	23	3600	18	14	1	0	1	0	0	0	0	0	0	2	0.726
3	M	23	18,600	93	54	0	0	5	0	0	1	0	0	0	33	0.943
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	F	16	13,800	69	0	0	0	0	0	0	0	0	0	0	69	0.194
91	F	16	14,400	72	0	0	0	0	0	0	0	0	0	0	72	2.72
92	F	16	17,400	87	0	0	0	0	0	87	0	0	0	0	0	0.687
				5342	461	157	441	553	30	349	174	74	583	233	2287	0.95

where k is a number of samples in test data. Then, average test errors given in tables and figures are found according to Eq. (15) by using calculated errors (h) (Özbay et al., 2006).

Table 7
The results from patient test data classified by FCM-PCA-NN

No.	M/F	Age	Samples	Number of segments	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9	Q_{10}	?	Error (%)
					N	Br	T	S	Apc	P	R	L	A.Fib	A.Flt		
1	M	23	11,200	56	0	0	0	4	0	0	0	1	0	0	51	0.732
2	M	23	3600	18	17	0	0	0	0	0	0	0	0	0	1	5.69×10^{-6}
3	M	23	18,600	93	10	0	0	3	0	0	0	1	1	0	78	0.587
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	F	16	13,800	69	69	0	0	0	0	0	0	0	0	0	0	8.84×10^{-4}
91	F	16	14,400	72	72	0	0	0	0	0	0	0	0	0	0	0.002
92	F	16	17,400	87	0	0	0	0	0	0	0	0	0	0	87	0.007
				5342	1197	126	45	189	69	124	30	161	571	141	2689	0.91

Table 8
The results from patient test data classified by WT-NN

No.	M/F	Age	Samples	Number of segments	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9	Q_{10}	?	Error (%)
					N	Br	T	S	Apc	P	R	L	A.Fib	A.Flt		
1	M	23	11,200	56	51	0	0	0	0	0	0	0	0	0	5	0.18
2	M	23	3600	18	17	0	0	0	0	0	0	0	0	0	1	0.016
3	M	23	18,600	93	88	0	0	0	0	0	0	0	0	0	5	0.061
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	F	16	13,800	69	69	0	0	0	0	0	0	0	0	0	0	0.126
91	F	16	14,400	72	72	0	0	0	0	0	0	0	0	0	0	4.59×10^{-4}
92	F	16	17,400	87	0	0	0	0	0	0	87	0	0	0	0	0.023
				5342	1828	609	113	401	79	58	1125	55	402	118	554	0.40

$$\text{Test Error }(\%) = \left(\frac{\sum_{j=1}^{n} h(j)}{m * n} \right) * 100 \tag{15}$$

where m is the number of segments in test data and n is the number of outputs of neural network.

4. Conclusion

In this paper, the FCM-PCA-NN has been developed and presented to classify electrocardiography signals. A comparative assessment of the performance of FCM-PCA-NN with NN, FCM-NN, PCA-NN and WT-NN show that more reliable results are obtained with the FCM-PCA-NN in shorter time for the classification of ECG signals. MLP NNs are still able to generalize with good recognition accuracy. However, they take longer time to train. The aim in developing FCM-PCA-NN was to achieve more optimum results with relatively few signal features. It has been demonstrated that the training time of the FCM-PCA-NN was 5.8% of the time required by the MLP NN. So, benefit is obtained as 93.2% in training time with %99 accuracy rate. We hope that the performance of the method will be better, if the number of the beats is increased for the training.

This technique is obtained by incorporating the techniques of fuzzy c-means clustering method, backpropagation learning and principal component analysis and combining their advantages. So, it can be said that the structure, which is more beneficial structure than conventional BPNN to recognize and classify ECG signals, is obtained.

Acknowledgement

This work is supported by the Coordinatorship of Selcuk University's Scientific Research Projects.

References

Acharya, R., Bhat, P. S., Iyengar, S. S., Roo, A., & Dua, S. (2002). Classification of heart rate data using artificial neural network and fuzzy equivalence relation. *The Journal of the Pattern Recognition Society*.

Castellano, G., & Fanelli, A. M. (2000). A self-organizing neural fuzzy inference network. In *Proceedings of IEEE International Joint Conference on Neural Networks* (Vol. 5), 14–19 Italy.

Castellano, G., & Fanelli, A. M. (2000). Fuzzy inference and rule extraction using a neural network. *Neural Network World Journal, 3*, 361–371.

Dazzi, D., Taddei, F., Gavarini, A., Uggeri, E., Negra, R., & Pezzarossa, A. (2001). The control of blood glucose in the critical diabetic patient: a neuro-fuzzy method. *Journal of Diabetes and Its Complications, 15*, 80–87.

De, R. K., Basak, J., & Pal, S. K. (2002). Unsupervised feature extraction using neuro-fuzzy approach. *Fuzzy Sets and Systems, 126*, 277–291.

Engin, M., & Demirağ, S. (2003). Fuzzy-hybrid neural network based ECG beat recognition using three different types of feature set. *Cardiovascular Engineering: An International Journal, 3*(2), 1–80.

Fan, J., Zhen, W., & Xie, W. (2003). Supervised fuzzy c-means clustering algorithm. *Pattern Recognition Letters, 24*, 1607–1612.

Foo, S. Y., Stuart, G., Harvey, B., & Meyer-Baese, A. (2002). Neural network-based ECG pattern recognition. *Engineering Applications of Artificial Intelligence, 15*, 253–260.

Güler, İ., & Übeyli, E. D. (2005). ECG beat classifier designed by combined neural network model. *Pattern Recognition, 38*, 199–208.

Hall, L. O., Bensaid, A. M., Clarke, L. P., Velthuizen, R. P., Silbiger, M. S., & Bezdek, J. C. (1992). A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images on the brain. *IEEE Transaction on Neural Networks, 3*(5), 672–682.

Haseyama, M., & Kitajima, H. (2001). An ARMA order selection method with fuzzy reasoning. *Signal Processing, 81*, 1331–1335.

Haykin, S. (1994). *Neural networks: a comprehensive foundation*. New York: Macmillan.

Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). *Neuro-fuzzy and soft computing*. USA: Prentice-Hall.

Karlık, B., Tokhi, M. O., & Alcı, M. (2003). A fuzzy clustering neural network architecture for multifunction upper-limb prosthesis. *IEEE Transaction on Biomedical Engineering, 50*(11), 1255–1261.

Leski, J., & Czogala, E. (1999). A new artificial neural network based fuzzy inference system with moving consequents in if-then rules and selected applications. *Fuzzy Sets and Systems, 108*, 289–297.

Liao, T. W., Celmins, A. K., & Hammell, R. J. II, (2003). A fuzzy c-means variant for the generation of fuzzy term sets. *Fuzzy Sets and Systems, 135*, 241–257.

Li, R., Mukaidono, M., & Turksen, I. B. (2002). A fuzzy neural network for pattern classification and feature selection. *Fuzzy Sets and Systems, 130*, 101–108.

Meesad, P., & Yen, G. G. (2000). Pattern classification by a neuro fuzzy network application to vibration monitoring. *ISA Transactions, 39*, 293–308.

Osowski, S., & Linh, T. H. (2001). ECG beat recognition using fuzzy hybrid neural network. *IEEE Transaction on Biomedical Engineering, 48*(11), 1265–1271.

Ozbay, Y. (1999). Fast recognition of ECG arrhythmias, PhD Thesis, Institute of Natural and Applied Science, Selcuk University.

Ozbay, Y., & Karlık, B. (2001). A recognition of ECG arrhythmias using artificial neural network, *Proceedings of the 23rd annual conference–IEEE/EMBS*, Istanbul, Turkey.

Özbay, Y., Ceylan, R., & Karlik, B. (2006). A fuzzy clustering neural network architecture for classification of ECG arrhythmias. *Computers in Biology and Medicine, 36*, 376–388.

Pal, N. R., Bezdek, J. C., & Tsao, E. C. (1993). Generalized clustering networks and Kohonen's self-organizing scheme. *IEEE Transaction on Neural Networks, 4*(4), 549–557.

Pedrycz, W. (1998). Conditional fuzzy clustering in the design of radial basis function neural networks. *IEEE Transaction on Neural Networks, 9*(4), 601–612.

Pektatlı, R., Ozbay, Y., Ceylan, M., & Karlık, B. (2003). Classification of ECG signals using fuzzy clustering neural networks (FCNN). In *Proceeding of International XII. TAINN'03* (Vol. 1, No. 1, pp. 105–108). Çanakkale, Turkey.

Pilla, V., & Lopes, H. S. (1999). Evolutionary training of a neuro-fuzzy network for detection of P wave of the ECG, In *Proceedings of the third international conference on computational intelligence and multimedia applications* (pp. 102–106). New Delhi, India.

Piramathu, S., Ragavan, H., & Shaw, M. J. (1998). Using feature construction to improve the performance of neural networks. *Management Science, 44*(3), 416–430.

Ronen, M., Shabtai, Y., & Guterman, H. (1998). Rapid process modelling-model building methodology combining unsupervised fuzzy clustering and supervised neural networks. *Computer Engineering, 22*, 1005–1008.

Sebzalli, Y. M., & Wang, X. Z. (2001). Knowledge discovery from process operational data using PCA and fuzzy clustering. *Engineering Applications of Artificial Intelligence, 14*, 607–616.

Seker, H., Evans, D. H., Aydın, N., & Yazgan, E. (2001). Compensatory fuzzy neural network- based intelligent detection of abnormal neonatal cerebral doppler ultrasound waveforms. *IEEE Transaction on Information Technology in Biomedicine, 5*(3), 187–194.

Shaout, A., & Scharbonea, J. (2000). Fuzzy logic based modification system for the learning rate in back-propagation. *Computers and Electrical Engineering, 26*, 125–139.

Stoeva, S., & Nikov, A. (2000). A fuzzy back-propagation algorithm. *Fuzzy Sets and Systems, 112*, 27–39.

Wang, X., & Paliwal, K. K. (2003). Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition, Pergamon. *The Journal of the Pattern Recognition Society, 36*, 2429–2439.

Yang, M., & Liu, H. (2003). Fuzzy least-squares algorithms for interactive fuzzy linear regression models. *Fuzzy Sets and Systems, 135*, 305–316.

Zhang, Y., & Kandel, A. (1998). Compensatory neuro-fuzzy systems with fast learning algorithms. *IEEE Transaction on Neural Networks, 9*(1), 83–105.