A Trainable Neural Network Ensemble for ECG Beat Classification

Atena Sajedin, Shokoufeh Zakernejad, Soheil Faridi, Mehrdad Javadi and Reza Ebrahimpour

Abstract—This paper illustrates the use of a combined neural network model for classification of electrocardiogram (ECG) beats. We present a trainable neural network ensemble approach to develop customized electrocardiogram beat classifier in an effort to further improve the performance of ECG processing and to offer individualized health care.

We process a three stage technique for detection of premature ventricular contraction (PVC) from normal beats and other heart diseases. This method includes a denoising, a feature extraction and a classification. At first we investigate the application of stationary wavelet transform (SWT) for noise reduction of the electrocardiogram (ECG) signals. Then feature extraction module extracts 10 ECG morphological features and one timing interval feature. Then a number of multilayer perceptrons (MLPs) neural networks with different topologies are designed.

The performance of the different combination methods as well as the efficiency of the whole system is presented. Among them, Stacked Generalization as a proposed trainable combined neural network model possesses the highest recognition rate of around 95%. Therefore, this network proves to be a suitable candidate in ECG signal diagnosis systems. ECG samples attributing to the different ECG beat types were extracted from the MIT-BIH arrhythmia database for the study.

Keywords—ECG beat Classification; Combining Classifiers; Premature Ventricular Contraction (PVC); Multi Layer Perceptrons; Wavelet Transform

I. INTRODUCTION

CORRECT and computationally efficient means of classifying electrocardiography (ECG) arrhythmias has been the subject of considerable research effort in recent years.

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 [1, 2]. Automatic arrhythmia detection and classification of arrhythmias are important in clinical cardiology, especially when performed in real time. This is achieved through the analysis of ECG beat and its extracted features [2].

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 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 within the same patients. The ECG waveforms may differ for the same patient to such an extent that they are dissimilar to each other and at the same time they are similar for different types of beats. This is main reason that the beat classifier, performing well on the training data, generalizes poorly when presented with different patients' ECG waveforms [2].Researchers have demonstrated the features extractions using digital filtering [3], Fourier Transform [4], Wavelet Transform [5,6], Principal Component Analysis (PCA) [7,8] and Independent Component Analysis (ICA) [9,10] which were using as an input pattern for classification task to the classifier.

One of the common approaches in ECG beat classification is Artificial Neural Networks (ANNs) that have shown accurate performance in different classification tasks. Among ANNs, Multi-Layer Perceptrons (MLPs) [1-10] and Radial Basis Function (RBF) [11-13] networks have been of considerable interest. Combining neural classifiers is one approach to improving the classification performance, particularly for complex problems. The idea behind combinations is the so-called divide-and-conquer principle, according to which a complex computational task is solved by dividing it into a number of computationally simple tasks and then combining the solutions of those tasks [14, 15].

There are two main strategies for combining classifiers: fusion and selection. In classifier fusion, it is supposed that each combining member is trained on the whole problem space, whereas in classifier selection, each member is assigned a particular part of the problem space to learn. Thus, in the former strategy, the final decision is made by taking into account the decisions of all members, while in the latter strategy; the final decision is made by aggregating the decisions of a few experts [16].

However conventional BPNN (Back Propagation Neural Networks) 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 output. So, researchers have started to use hybrid structure. For example Pilla and Lopes [17], Osowski and Linh [2] and Engin and Demirag [18] have used fuzzy hybrid neural network and Guler and Übeyli [19-21], have used combined neural networks, Mixture of Experts (MME) network structure and error correcting output codes method.

We used different combination methods and comparison of them shows that the stacked generalization method has yielded the best results. The general framework using an ensemble of models consists of two levels and is often referred to as stacked generalization. In the first level, various learning methods are used to learn different models from the original data set. The recognition of the models from the first level along with the corresponding target class of the original input data are then used as inputs to learn a second level model [22]. We used five different combination methods: Max, Min, Average, Product and Stacked Generalization for ECG signal classification.Experimental results indicate that our proposed trainable combining method performs better than other combining methods. The outline of this study is as follows. In Section 2, 3 we explain and spectral analysis of signals using DWT in order to extract features characterizing the behavior of the signal under study. In Section 4, 5 we present description of neural network models including Multi-Layer Perceptrons neural network and combined neural network topology used in this study. In Section 6, we present the results of our experiments involving the application of combined neural network model to the ECG signals. Finally, in Section 7 we conclude the study.

II. SIGNAL PREPROCESSING

The objectives of preprocessing stage are the omission of high-frequency noise and the enhancement of signal quality to obtain appropriate features. Furthermore, we should remove equipment and environment influences on recorded signals. Hence it is, Noise reduction is one of the main operations in the course of preprocessing. Conventionally, ECG signal is measured on static conditions since various types of noise including muscle artifacts and electrode moving artifacts are coupled in dynamic environment. To remove such noises an advanced signal processing method, such as discrete wavelet transform (DWT) denoising technique [23] should be used. This method has been emerged over recent years as a powerful time–frequency analysis and signal coding tool favored for the interrogation of complex signals,. However,

DWT is not a time-invariant transform. To solve this problem, we used the stationary wavelet transform (SWT) which is also known as the un decimated wavelet transform (UWT) or translation-invariant wavelet transform. SWT uses the average of several de noised signals that are obtained from the coefficients of e-decimated DWT [24].

Suppose the signal $S \in L^2$ The SWT is given by:

$$w_{v(\tau)} = \frac{1}{\sqrt{v}} \int_{-\infty}^{\infty} s(t) \Psi^* \left(\frac{t - \tau}{v} \right) dt \tag{1}$$

Where $v = 2^k$, $k \in Z$, $\tau \in R, \psi^*$ is the complex conjugate of the mother wavelet. Fig. 1 shows the block diagram of SWT.

The fig. 1 shows a decomposition of three levels, the blocks of $H(z)$ and $H_{r(z)}$ are the decomposition and reconstruction high pass filters and the blocks of $G(z)$ and $G_{r(z)}$ are low pass filters. $d(.,.)$ denotes the decomposition coefficients and $a(.,.)$ denotes the approximation coefficients.There is no absolute way to choose a certain wavelet. The choice of the wavelet function depends on the application. The Daubechies algorithm is conceptually more complex and has a slightly higher computational overhead. . But, the Daubechies algorithm picks up detail that is missed by the other wavelet algorithm [25]. Even if a signal is not well represented by one member of the Daubechies family, it may still be efficiently represented by another. Selecting a wavelet function which closely matches the signal to be processed is of utmost importance in wavelet applications [26].

Daubechies wavelet family are similar in shape to QRS complex and their energy spectrum are concentrated around low frequencies. Here, For denoising, we have used the Daubechies wavelet functions (db1) with decomposition level of five based on our extensive experiments. In this study, we have used the Savitsky–Golay filtering method For smoothing of the ECG signals [24]. The filter coefficients are achieved by the un-weighted linear least-squares fitting method using a polynomial function.
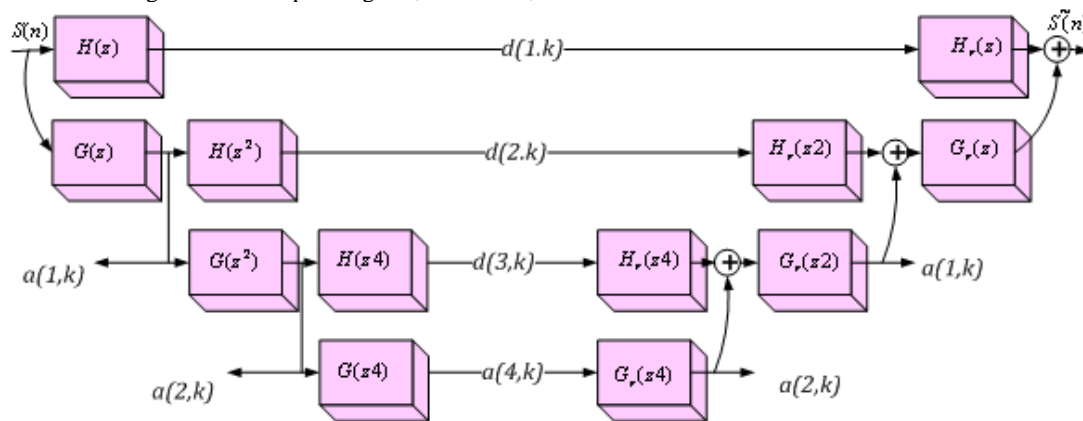

Fig. 1 Block diagram of SWT. H(z) and Hr(z) are the decomposition and reconstruction high pass filters. G(z) and Gr(z) are low pass filters. Term d (.,.) denotes the decomposition coefficients and a (., .) denotes the approximation coefficients.

Because of this, the Savitsky–Golay filter is also called a digital smoothing polynomial filter or a least-squares smoothing filter. A higher degree of polynomial makes it possible to achieve high level smoothing without attenuation of the data features. The Savitsky–Golay filtering method is often used for frequency information or spectroscopic (peak) information. For the first type, it conserves the high-frequency components of the signal and for the second type it conserves higher moments of the peak.

III. FEATURE EXTRACTION

Feature extraction is one of the most important processes for classification tasks.In this study, we have used a balanced combination of morphological and timing features. These features describe the basic shape of the signals and position of waves within a given window of beats. The extracted parameters that describe the basic shape of the beats are: amplitude of P-peak (amP), amplitude of Q-valley (ampQ), amplitude of R-peak (ampR),amplitude of S-valley (ampS) and amplitude of T-peak (ampT). Features that describe wave position within a beat window are: position of P-peak (posP), position of Q-valley (posQ), position of R-peak (posR), position of S-valley (posS) and position of T-peak (posT). The time duration between PVC beats contains useful information about their types. So we define a feature called rat RR. This feature is defined as the time ratio between the last beat to the next one. To extract this feature we propose a two-steps method. The first step involves the cutting of the normal and PVC and other beats by making use of the annotation files which exist in MIT–BIH arrhythmia database.

The second step involves identification of peaks and valleys in every normal or abnormal beat and obtaining their respective amplitudes and positions. In order to break to normal and abnormal beats, we process annotated file records from MIT–BIH database. For example to extract normal beats, the algorithm examines the annotation file which contains the sample number of the normal beat. Then it creates a matrix with rows equal to the number of normal beats. An R-wave detector is required to initialize our computer-aided ECG classification process. Next, the algorithm saves 40 samples surrounding the target normal beat from all the recorded samples. The sample beat itself is also saved in the final matrix. We extracted the abnormal beats in the same manner too. After classification of normal and abnormal beats, peaks and valleys are detected. For this purpose, we implemented the Al-Alaoui algorithm [27]. The peak and valley detector correctly detects the P, Q, R, S, and T waves. Sometimes, especially in the case of arrhythmia, it is possible for the algorithm to recognize extra peaks or valleys. Since the P and the T waves exist at the beginning and the end of each window, respectively, in such a case the first peak is set as the P and the last peak as the T wave; other peaks are hence rejected. Similarly, the algorithm marks the nearest valley at the left of center of the beat as the Q, and the nearest valley to the right of center of the beat as the S wave. We extracted ten ECG morphological, as well as one timing interval features.

An RR-interval ratio (IR) reflecting the deviation from a constant beat rate was defined as:

$$IR_i = \frac{T_i - T_{i-1}}{T_{i+1} - T_i} \tag{2}$$

Where T_i represents the time at which the R-wave for beat I occurs. The local RR-interval ratio information was chosen as opposed to other parameters such as average RR-interval or time series because it provides a convenient differentiator between normal beats and PVC beats, and is normalized by definition (at constant rate).

IV. DESCRIPTION OF NEURAL NETWORK MODELS

The artificial neural network type used for this research was the standard feed-forward multi-layer Perceptrons. As a starting point, we define the net input to a processing element q for the m th ECG input pattern as:

$$vq(m) = \sum_{j=0}^{j-1} wqjtj \ (m) + bq \tag{3}$$

where tj(m), j = 0,1,. . . ,J _ 1 are the values of the wavelet coefficients for the mth input pattern, wqj is the neural network weight of the connection between coefficient j and processing element q and bq is an ANN bias element. The output of the processing element is the result of passing the scalar value mq(m) through its activation function.

$$\varphi q(.) : yq(m) = \varphi q(vq(m)) \tag{4}$$

In this work we have used the logistic sigmoid function, which has a range of values between zero and one:

$$\varphi q(vq(m)) = \frac{1}{1 + e^{-vq(m)}} \tag{5}$$

The output of the feed-forward MLP is of the form

$$y = \varphi_2(\varphi \ W_{21}(\varphi \ W_{1x} + b_1) + b_2) \tag{6}$$

where φ_1 is the activation function of the hidden layer, which is assumed to apply element-wise to its vector argument, b_2 is the scalar bias in the output layer, W_1 and W_2 are weight matrices and b_1 is a bias vector. We have used a back-propagation training algorithm that employs a gradient search to minimise the mean-square-error (MSE) cost function:

$$MSE = \frac{1}{M} \sum_{m=1}^{M} [y(m) - d(m)]^2 \tag{7}$$

where $d(m)$ is the desired output for the mth training pattern, $y(m)$ is the actual MLP output and M is the total number of training samples. Very often, the ANN architectures used are based on the principle that the neural network should be larger than require d rather than not sufficiently complex. As a consequence of this design choice, there existed the possibility of data over-fitting, which is undesirable since such practice can potentially compromise the generalization ability of the network. To minimize the risk of data over-fitting a cross validation method can be used, where the available data is

divided into three distinct data sets: the training set, the validation set and the testing set. The ANN was then trained as usual with the training data set but every so often its generalization ability was cross-validated with the validation set. As soon as the classification accuracy of the validation set deteriorated, the training stopped. This is commonly referred to in the NN literature as the ''early stopping method of training'' [28].

V. NEURAL NETWORK COMBINING METHODOLOGY

From a computational viewpoint, according to the principle of divide-and-conquer, a complex computational task is solved by dividing it into a number of computationally simple tasks and then combining the solutions of those tasks. In supervised learning, computational simplicity is achieved by distributing the learning task among a number of experts, which in turn divides the input space into a set of subspaces [28].

There are generally two types of combining strategies: expert selection and expert fusion [29]. The selection paradigm is based on the assumption that each of the base experts is specialized in a particular local area of the problem space. There can be one specific expert nominated to make the decision in each subspace, as was done by Rastrigin and Erenstein [30], or in some cases one can devote more than one local expert to a local area, as was done by Jacobs, Jordan, Nowlan, and Hinton [31]and Alpaydin and Jordan [32]. Expert fusion assumes that all experts are trained over the whole problem space, and are therefore considered as competitive rather than complementary [33][34].

As the input signal is involved in the combining procedure, combining neural networks as experts may be classified into two major categories:

I. Static structures: In this class of combining methods of neural networks, the responses of several predictors (neural networks) are combined by means of a mechanism which does not involve the input signal; hence the designation "static".

II. Dynamic structures: In the second class of combining methods, the input signal is directly involved in actuating the mechanism that integrates the outputs of the individual experts into an overall output; hence the designation "dynamic" [28]. The combination methodologies from the combiner viewpoint are divided into two categories: non-trainable and trainable:

5.1. Non-trainable (Rule-Based) combiners
Simple algebraic combiners are, in general, non-trainable combiners of continuous outputs. The total support for each class is obtained as a simple function of the supports received by individual classifiers. Following the same notation in [34], we represent the total support received by class ωj, the j th column of the decision profile $DP(X)$, as

$$\mu_{J(X)} = \Im[d_{1,j}(X),....,d_{T,j}(X)] \tag{8}$$

where is the combination function, such as one of those listed below.

5.1.1. Mean Rule: The support for ωj, is obtained as the average of all classifiers' j th outputs, that is, the function $\Im(.)$ is the averaging function. The mean rule is equivalent to the sum rule (within a normalization factor of $\dfrac{1}{T}$), which also appears often in the literature.

In either case, the ensemble decision is taken as the class ωj, for which the total support is $\mu_{j(X)}$ largest.

$$\mu_j(x) = \frac{1}{T}\sum_{t=1}^{T} d_{t,j}(X) \tag{9}$$

5.1.2. Minimum/Maximum/Median Rule: As the names imply ,these functions simply take the minimum, maximum or the median among the classifiers' individual outputs.

$$\mu_j(x) = \min_{t=1...T}\{d_{t,j}(X)\} \tag{10}$$

$$\mu_j(x) = \max_{t=1...T}\{d_{t,j}(X)\} \tag{11}$$

$$\mu_j(x) = \underset{t=1...T}{median}\{d_{t,j}(X)\} \tag{12}$$

In any of these cases, the ensemble decision is again chosen as the class for which total support is largest. The minimum rule is the most conservative combination rule, as it chooses the class for which the minimum support among the classifiers is largest. Also worth mentioning is that the trimmed mean at limit 50% is equivalent to the median rule.

5.1.3. Product Rule: In product rule, supports provided by the classifiers are multiplied. This rule is very sensitive to the most pessimistic classifiers: a low support (close to 0) for a class from any of the classifiers can effectively remove any chance of that class being selected. However, if individual posterior probabilities are estimated correctly at the classifier outputs, then this rule provides the best estimate of the overall posterior probability of the class selected by the ensemble.

$$\mu_j(x) = \frac{1}{T}\prod_{t=1}^{T} d_{t,j}(X). \tag{13}$$

5.2. Trainable combiners
In trainable combiners, a learning process makes the combiner learn to map the base classifiers' outputs to the target space.

5.2.1. Stacked Generalization Method
The general framework of this method (see Figure 2) consists of two levels. The first level, is formed by base classifiers which are trained using the input data and the target output. The output of first level is then used as the input of second level. As is shown in Figure 2*b, a set of K "first level" networks from N_1^0 to N_k^0 are arranged as the first layer, and

their outputs are combined using a "first level" network N^1. First, the first level networks are trained, using the input data and the target outputs. Then the outputs of the first layer with the corresponding target class are used to train the second level network [36].

VI. EXPERIMENTAL RESULT

6.1. MIT–BIH arrhythmia database

The MIT–BIH arrhythmia database [37] was used as the data source in this study. The database contains 48 recordings each of which has a duration of 30 minutes and includes two leads;

the modified limb lead II and one of the modified leads V1, V2, V4 or V5. The sampling frequency is 360 Hz, the data are band pass filtered at 0.1–100 Hz and the resolution is 200 samples per mV. twenty-three of the recordings are intended to serve as a representative sample of routine clinical recordings and 25 recordings contain complex ventricular, junctional and supra ventricular arrhythmias.
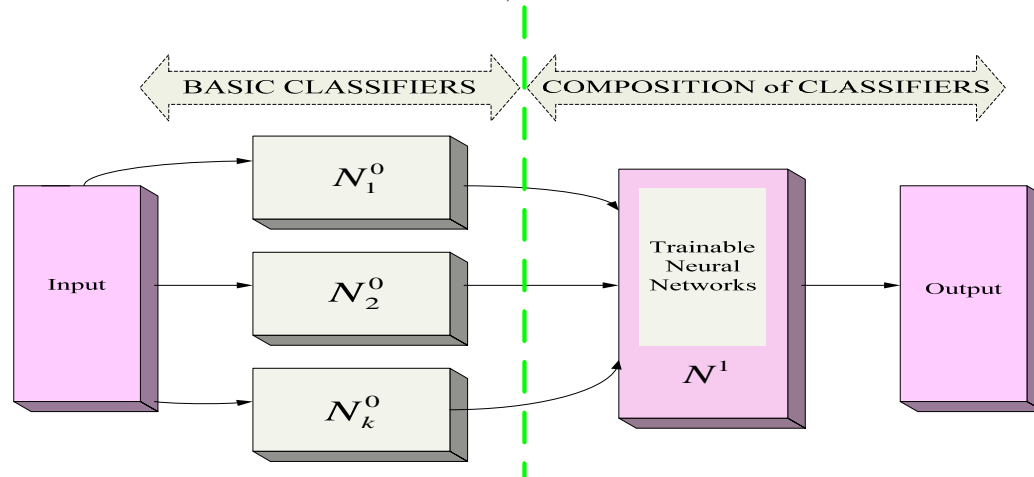

Fig. 2 Block diagram of combined neural network; Stacked generalization method

There are over 109,000 labeled ventricular beats from 15 different heartbeat types. There is a large difference in the number of examples in each heart beat type. The largest class is ''Normal beat'' with about 75,000 examples and the smallest class is ''Supra ventricular premature beat'' (SP) with just two examples. The database is indexed both in timing information and beat classification. For more details about MIT–BIH arrhythmia database see [38].We used a total of seven records marked as: 100, 101, 102, 104, 105, 106, and 107 in the database. We extracted a total of 15,566 beats: 8390 normal beats, 627 abnormal PVC arrhythmia beats, and 6549 other arrhythmic beats. We used the database index files from database to locate beats in ECG signals.

6.2 Application of Stacked Generalization to ECG signals

We applied five combination methods and observed significant performance enhancement using proposed method.

6.2.1. Basic classifier structure

We have used MLPs classifiers with Back Propagation training algorithm. Each classifier has 11 nodes in the input

layer equal to the number of extracted features. Also, they have one hidden layer and the number of its neurons determined via trial and error method; in this manner we have increased the number of hidden neurons from 5 to 80 in order to obtain the recognition rates and finally find the optimal number of hidden neurons. The average recognition rate for MLPs network, with one hidden layer is obtained 94% for ten times run and maximum recognition rate is achieved 94.4%. An important issue in combining classifiers is the diversity of the classifiers in learning the input data. When the base classifiers of a combining structure are diverse (i.e., they learn different areas of the input space), they become specialized in specific areas of the input space, and consequently have fewer errors in those areas. Thus, combining the outputs of classifiers that are perfectly diverse, improves the overall performance. For diversifying the basic classifiers, different training parameters and classifiers with different structures have used. In table 1, properties of three basic diverse classifiers were shown.

TABLE I RECOGNITION RATES AND OTHER PARAMETERS OF THE BASE CLASSIFIER

Base Classifiers	Number of Neurons in Hidden Layer	Number of Epochs	Initial weights Range	Recognition rates (%)
classifier 1	17	1700	[-2 2]	93.71
classifier 2	22	2700	[-3 3]	93.77
classifier 3	25	2000	[-4 4]	93.78

6.2.2. Application of different combining methodologies and proposed method to ECG signals

In all experiments, three base classifiers are first created. In test phase, for Max, Min, Average and Product methods, their outputs are combined with corresponding rules. In Stacked Generalization scheme, the outputs are used as inputs to second level classifier to learn the mapping between the base classifiers' outputs and the target of test sample. As we have three classifiers in the first level to solve a three class problem, the input of the combiner (second-level classifier) in conventional Stack Generalization, will have nine elements.

In Table 2, the recognition rates of different combining methods are listed. To find the best topology of the combiner, we employed the same strategy as for the base classifiers by increasing the number of hidden neurons from 5 to 60 and investigating the best recognition rate on our validation set. The optimal numbers of hidden neurons were found to be 30 for Stack Generalization.

Table 2, shows the recognition rates of the best topologies of the method, as well as the other four methods. In all experiments, the classification performance is measured using four-fold cross validation technique. Note that the mentioned recognition rates are the average from 10 times run.

TABLE II RECOGNITION RATES FOR DIFFERENT COMBINING METHODS AND STACKED GENERALIZATION

Combining Method	Max	Min	Average	Product	Stacked Generalization
Recognition Rate (%)	93.3	94.09	94.05	93.63	95

The structure of Stacked Generalization at first level consists of previously used MLPs architecture as base classifiers. The optimum structure of combiner at second level is extracted through a try and error trial for Stacked Generalization (see Table 3).

At second level in combiner network Stacked Generalization uses an MLP with the structure of 9:25:3 with sigmoid tangent transfer function in hidden layer and sigmoid logarithmic activation function in output layer.

TABLE III TESTING DIFFERENT TOPOLOGIES IN SECOND LEVEL NETWORK FOR A BEST STACKED AND STACK COMBINER

Number of hidden Neurons (Stacked method)	6	8	10	12	14	16	18	20
Average Classification rate (%)	72.5	73.6	72.9	75.97	74.6	73.4	72.2	73.3

VII. CONCLUSION

This paper presented the use of combined neural network methods to improve an accuracy of ensemble neural networks for classification of the ECG signals. We have used five different combination methods; Max, Min, Average, Product and Proposed Stacked generalization for ECG signals classification.

The combined neural network used for classification of the ECG signals was trained, cross validated and tested with the extracted features using discrete wavelet transform of the ECG signals. Considering the experimental results, the best method in our work is stacked generalization method with highest recognition rate of 95%.

REFERENCES

[1] R. Acharya, P.S. Bhat, S.S. Iyengar, A. Roo and S. Dua, Classification of heart rate data using artificial neural network and fuzzy equivalence relation, Pattern Recognition, vol. 36, no. 1, pp. 61-68, 2003.

[2] S. Osowski and T.H. Linh, ECG beat recognition using fuzzy hybrid neural network, IEEE Trans. Biomed. Eng., vol. 48, no. 11, pp. 1265–1271, 2001.

[3] O. Pichler, A. Teuner and B.J. Hosticka, A comparison of texture feature extraction using adaptive Gabor filtering, pyramidal and tree structured wavelet transforms, Pattern Recognition, vol. 29, no. 5, pp. 733-742, 1996.

[4] Z. Dokur, T. Ölmez and E. Yazgan, Comparison of discrete wavelet and Fourier transforms for ECG beat classification, Electron. Lett., vol. 35, no. 18, 1999.

[5] K. Sternickel, Automatic pattern recognition in ECG timeseries, Comput. Meth. Prog. Biomed. 68 109–115,2002.

[6] Z. Dokur and T. Olmez, ECG beat classification by a novel hybrid neural network, Comput. Meth. Prog. Biomed., vol. 66, pp. 167–181, 2001.

[7] X. Wang and K. Paliwal, Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition, Pattern Recognition, vol. 36, pp. 2429–2439, 2003.

[8] Y. Ozbay, R. Ceylan and B. Karlik, A fuzzy clustering neural network architecture for classification of ECG arrhythmias, Computers in Biology and Medicine, vol. 36, pp. 376–388, 2006.

[9] A. Hyvarinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Transactions on Neural Networks, vol. 10, no. 3, pp. 626–634, 1999.

[10] Z. Wang, Z. He and J. Z. Chen, Blind EGG separation using ICA neural networks, In Proceedings of the 19th annual international conference of the IEEE-EMBS, Chicago, IL. USA, vol. 3, pp. 1351–1354, 1997.

[11] Y. Ozbay and B. Karlik, A recognition of ECG arrhythmias using artificial neural network, Proceedings of the 23rd Annual Conference, IEEE/EMBS, Istanbul, Turkey, 2001.

[12] Y. Ozbay, Fast recognition of ECG arrhythmias, Ph.D. Thesis, Institute of Natural and Applied Science, Selcuk University, 1999.

[13] S.Y. Foo, G. Stuar and B. Harvey, A. Meyer-Baese, Neural network-based ECG pattern recognition, Eng. Appl. Artif. Intell., vol. 15, pp. 253–260, 2002.

[14] R. Ebrahimpour, E. Kabir, H. Esteky and M. R. Yousefi, A mixture of multilayer perceptron experts network for modeling face/nonface recognition in cortical face processing regions, Intelligent Automation and Soft Computing, vol. 14, no. 2, pp. 145–156, 2008.

[15] R. Ebrahimpour, E.Kabir, H. Esteky, M. R. Yousefi, View-independent face recognition with mixture of experts, Neurocomputing, vol. 71, no. 4-6, pp. 1103–1107, 2008.

[16] L.Kuncheva, Combining pattern classifiers: Methods and algorithms. New York: Wiley, 2004.

[17] V. Pilla and H.S. Lopes, Evolutionary training of a neuro-fuzzy network for detection of P wave of the ECG, Proceedings of the Third International Conference on Computational Intelligence and Multimedia Applications, New Delhi, India, pp. 102–106, 1999.

[18] M. Engin and S. Demira˘g, Fuzzy-hybrid neural network based ECG beat recognition using three different types of feature set, Cardiovasc. Eng. Int. J., vol. 3, no. 2, pp. 71–80, 2003.

[19] I. Guler, and E. D. Ubeyli, A mixture of experts network structure for modelling Doppler ultrasound blood flow signals, Computers in Biology and Medicine, vol. 35, no. 7, pp. 565–582, 2005.

[20] I. Guler, and E. D. Ubeyli, ECG beat classifier designed by combined neural network model, Pattern Recognition, vol. 38, no. 2, pp. 199–208, 2005.

[21] E.D. Ubeyli, Support vector machines for detection of electrocardiographic changes in partial epileptic patients, Engineering Applications of Artificial Intelligence, vol. 21, pp. 1196–1203, 2008.

[22] D.H. Wolpert, Stacked generalization, Neural Networks, vol. 5, pp. 241–259, 1992.

[23] D. Donoho and I. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, Journal of the American Statistical Association, vol. 90, no. 432, pp. 1200–1223, 1995.

[24] D.Donoho, De-noising by soft-thresholding, in: IEEE Transactions on Information Theory, vol. 41, pp. 613–627, 1995.

[25] Daubechies, I., Ten Lectures on Wavelets, SIAM, Philadelphia, 1992.

[26] Amara Grap, An Introduction to Wavelets, IEEE Comp.Sc. And Eng., Vol. 2, No. 2, 1995.

[27] M.A. Al-Alaoui, A unified analog and digital design to peak and valley detector window peak and valley detectors and zero crossing detectors, IEEE Trans. Instrum. Meas., vol. 35, pp. 304–307, 1986.

[28] S. Haykin, Neural networks: A comprehensive foundation, USA: Prentice Hall, 1999.

[29] K. Woods, W.P. Kegelmeyer and K. Bowyer, Combination of multiple classifiers using local accuracy estimates, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, pp. 405–410, 1997.

[30] L. A. Rastrigin and R. H. Erenstein, Method of collective recognition. Moscow: Energoizdat (in Russian), 1982.

[31] R. A. Jacobs, M. I. Jordan, S. E. Nowlan and G. E. Hinton, Adaptive mixture of experts, Neural Computing, vol. 3, pp. 79–87, 1991.

[32] E. Alpaydin and M. I. Jordan, Local linear perceptrons for classification, IEEE Transactions on Neural Networks, vol. 7, no. 3, pp. 788–792, 1996.

[33] L. Xu, A. Krzyzak and C. Suen, Methods of combining multiple classifiers and their application to handwriting recognition, IEEE Transactions on Systems, Man, and Cybernetics, vol. 22, pp. 418–435, 1992.

[34] K. C. Ng, and B. Abramson, Consensus diagnosis: A simulation study, IEEE Transactions on Systems, Man, and Cybernetics, vol. 22, pp. 916–928, 1992.

[35] R. Pektatli,Y. Ozbay,M. Ceylan and B. Karlik, Classification of ECG signals using fuzzy clustering neural networks (FCNN), Proceedings of the International XII, TAINN'03, vol.1, no. 1, Canakkale, Turkey, pp. 105–108, 2003.

[36] H. Nikoo, M. Azarpeikan, M. R. Yousefi, R. Ebrahimpour and A. Shahrabadi, Using a trainable neural network ensemble for trend prediction of Tehran stock exchange, International Journal of Computer Science and Network Security, vol. 12, pp. 287–293, 2007.

[37] R.G. Mark and G.B. Moody, MIT–BIH Arrhythmia Database 1997. Available from: <http://www.ecg.mit.edu/dbinfo.html>.

[38] G.B. Moody and R.G. Mark, The impact of the MIT/BIH arrhythmia database, IEEE Eng. Med. Biol. Mag., vol. 20, no. 3, pp. 45–50, 2001.

Atena Sajedin received the B.Sc. degree in Electrical Engineering from, Islamic Azad University Centeral Tehran Branch, Tehran, Iran, in 2005 and she is the student of M.Sc. in Mechatronic Engineering from Islamic Azad University South Tehran Branch, Tehran, Iran. Her research interests include human and machine vision, neural networks and pattern recognition.
e-mail: atena.sajedin@gmail.com

Shokoufeh Zakernejad received the B.Sc. degree in Electronics Engineering from Shahid Rajaee University in 2007 and she is the student of M.Sc. in Electrical Engineering from Shahid Rajaee University, Tehran, Iran. Her research interests include human and machine vision, neural networks and pattern recognition.

Soheil Faridi received the B.Sc. degree in Electrical Engineering from, Islamic Azad University South Tehran Branch, Tehran, Iran, in 2008 and he is the student of M.Sc. in Mechatronic Engineering from Islamic Azad University South Tehran Branch, Tehran, Iran. His research interests include Computational Neuroscience, Biologically Motivated Object Recognition, Neural Networks and Combining Classifiers.

Mehrdad Javadi received the B.Sc. degree in Mechanical Engineering from Sharif University, Tehran, Iran and the MS degree in Mechanical Engineering from Tehran University, Tehran, Iran.
He received his PhD degree from Islamic Azad University, Science and Research Branch, Tehran, Iran.He is head of Mechatronics dept. and his research interest is mechanical –electrical interfaces.

Reza Ebrahimpour received the B.Sc. degree in Electronics Engineering from Mazandaran University, Mazandaran, Iran and the MS degree in Biomedical Engineering from Tarbiat Modarres University, Tehran, Iran, in 1999 and 2001, respectively. He received his PhD degree in July 2007 from the School of Cognitive Science, Institute for Research in Fundamental Sciences (IPM), where he worked on view-independent face recognition with Mixture of Experts. His research interests include human and machine vision, neural networks and pattern recognition. e-mail: ebrahimpour@ipm.ir