Available online at www.sciencedirect.com

ScienceDirect

Procedia Technology 10 (2013) 76 – 84

Procedia Technology

International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013

Cardiac Arrhythmia Classification Using Neural Networks with Selected Features

Malay Mitra[1] and R. K. Samanta[2]*

[1]*Expert System Laboratory, Department of Computer Science & Application,University Of North Bengal, Raja Rammuhunpur, 734013, West Bengal, India.*

[2]*Expert System Laboratory, Department of Computer Science & Application,University Of North Bengal, Raja Rammuhunpur, 734013, West Bengal, India.*

Abstract

This research is to present a new approach for cardiac arrhythmia disease classification. An early and accurate detection of arrhythmia is highly solicited for augmenting survivability. In this connection, intelligent automated decision support systems have been attempted with varying accuracies tested on UCI arrhythmia data base. One of the attempted tools in this context is neural network for classification. For better classification accuracy, various feature selection techniques have been deployed as prerequisite. This work attempts correlation-based feature selection (CFS) with linear forward selection search. For classification, we use incremental back propagation neural network (IBPLN), and Levenberg-Marquardt (LM) classification tested on UCI data base. We compare classification results in terms of classification accuracy, specificity, sensitivity and AUC. The experimental results presented in this paper show that up to 87.71% testing classification accuracy can be obtained using the average of 100 simulations.

Selection and peer-review under responsibility of the University of Kalyani, Department of Computer Science & Engineering

Keywords- Arrhythmia; UCI database; Neural networks; CFS; Incremental back propagation; Levenberg-Marquardt Classification.

* Corresponding author. Tel.: +0-0353-2776344; fax: +0-0353-2776344.
E-mail address:[1]*malay.mitra68@gmail.com,*[2] rksamantark@gmail.com

Selection and peer-review under responsibility of the University of Kalyani, Department of Computer Science & Engineering
doi:10.1016/j.protcy.2013.12.339

1. Introduction

Heart disease is one of the major problems world-wide. Early detection of heart diseases and proper medical treatment can save lives in cases of sudden deaths [1]. One of the important diagnostic tools of heart function is ECG or EKG (Electrocardiogram) which creates a graphic record of the heart's electrical impulses. Fig. 1. shows the ECG patterns of normal heartbeat and abnormal heart beat. An arrhythmia is an abnormal heart rhythm. There are two basic kinds of arrhythmias. Bradycardia is when the heart rate is too low – less than 60 beats per minute [fig. 1]. Tachycardia is when the heart rate is too fast – more than 100 beats per minute [fig. 1]. When arrhythmias are severe, the heart's ability to pump blood may be reduced causing shortness of breath, chest pain, feeling tired, loss of consciousness. If more severe, it can cause heart attack or death.

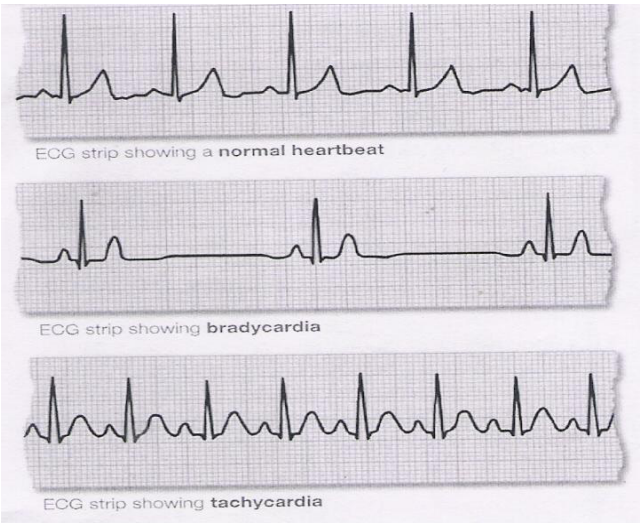

Fig. 1. Human ECG pattern

So, to avoid any bad happenings, we are to detect and classify the cardiac arrhythmia. Approaches have already been developed for classifying cardiac arrhythmias based on ECG signal data but still show poor performance. Various machine learning and data mining methods are being deployed to improve the detection of cardiac arrhythmia. But, however, there are differences between the cardiolog's and the program classification. Taking the cardiolog's as gold standard, we are to minimize the difference by means of machine learning tools [2].

Different approaches have been proposed to develop automated recognition and classification of ECG. Self-Organizing Maps (SOM), Support Vector Machines (SVM), Multilayer Perceptron (MLP), Markov Models, Fuzzy or Neuro-fuzzy Systems and combinations of different approaches have been proposed to improve performance [3-9].

In this study, we attempt two combinations (i) a combination of CFS + IBPLN, (ii) a combination of CFS + LM. We focus on the worst simulation result, the best simulation result as well as an average of 100 simulations. Experiments have been carried out using UCI data set [2]. The aim is to distinguish between the presence and absence of cardiac arrhythmia and to classify - normal or cardiac arrhythmia.

The rest of the paper is organized as follows. Section II discusses the UCI arrhythmia data set. Section III discusses the preliminaries of feature selection nd reduction and CFS as implementation technique. Section IV discusses the preliminaries of artificial neural networks (ANN), incremental back propagation learning networks (

IBPLN), and Levenberg-Marquardt (LM) classification. Section V discusses the applications. Section VI presents the modeling results. Lastly, our conclusions are summarized.

2. UCI ARRHYTHMIA DATA SET

This database was contributed by Dr. H. Altay Guvenir (1998) [2]. The total number of instances is 452. The number of attributes is 279, 206 of which are linear valued and the rest are nominal. For missing values, in our study, we discard these 32 instances with missing values and use 420 instances.

3. FEATURE EXTRACTION AND REDUCTION

Feature extraction and reduction is one of the important steps for classification since even the best classifier may perform poorly if the features are not will chosen [10], [11]. The reduced feature vector includes most of the useful information of the original feature vector. This reduced dimensionality helps reducing the database size as well as speeds up the inference procedure especially for a large data base. There are different algorithms for the purpose. Correlation-based feature subset selection (CFS) [12], Principal Component Analysis (PCA) [10], Association Rules (AR) [13], Rough Set theory [14] are some of the techniques to mention. This work uses CFS..

3.1. Correlation based feature subset selection (CFS)

The central hypothesis of CFS is

"A good feature subset is one that contains features highly correlated with (predictive of) the class, yet uncorrelated with (not predictive of) each other" [12].

A feature evaluation formula, based on ideas from test theory [18], provides an operational definition of the above hypothesis as follows:

$$r_{fc} = \frac{k\overline{r_{fc}}}{\sqrt{\{k + k(k-1)\overline{r_{ff}}\}}} \qquad (1)$$

where r_{fc} is the correlation between the summed features and the class variable, k is the number of features, $\overline{r_{fc}}$ is the average of the correlation between the features and class variable, and $\overline{r_{ff}}$ is the average inter-correlation between features. CFS is an algorithm that couples this evaluation formula with an appropriate correlation measure and a heuristic search strategy.

To accommodate nominal or categorical as well as continuous or ordinal features in Equation 1, continuous features are transformed to categorical features using the supervised discretisation method of Fayyad and Irani [15] as a preprocessing step before applying in classification task. The theory of information gain [16] is applied estimating the degree of associations between nominal features. Moreover, there are 2^n (n is the number of possible initial features initially) possible subsets of reduced features are possible. It would be impractical especially for a large feature set to explore each and every such subset for finding the best subset. Heuristic search strategies, such as best first and hill-climbing [17] are often used to search the feature subset in reasonable amount of time. Moreover, both filter type as well as wrapper types of feature selection methods use correlation-based approach in different applications [18], [19].

4. ARTIFICIAL NEURAL NETWORKS

4.1. Preliminaries

Artificial neural networks (ANN) mimic the workings of the neurons of human brain. The neurons are connected to one another by connection links. Each link has a weight. A simple McCulloch-Pitts model of a neuron [20] is presented in Fig.2. which was presented in the year 1943.

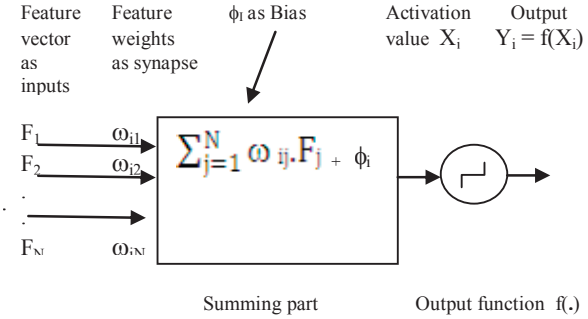

Fig. 2. McCulloch-Pitts model of a neuron.

This model of neuron is the basis of the discipline of artificial neural networks. In the literatures, different forms of ANNs are there for modeling different tasks. Depending upon the function to be performed, different neural network models assume different modes of operation for the network. Basically, they can be either feedforward type or feedback type.

Modelling with ANN involves two important tasks, namely, *design* and *training* the network. The design of a networks involves (1) fixing the number of layers, (2) the number of neurons for each layer, (3) the node function for each neuron, (4) whether feedback or feedforward, and (5) the connectivity pattern between the layers and the neurons. All these adjustments are to be taken care of for improved performance of the system. The training phase or the learning phase involves adjustments of weights as well as threshold values from a set of training examples. The kind of learning law was first proposed by Donald Hebb [21]. Currently, there are hundreds of such leaning algorithms in the literature [22], but the most well-known among them are backpropagation [23], [24], ART [25], and RBF networks [26].

4.2. Incremental Backpropagation Learning Networks

The normal backpropagation network is not an incremental by its nature [27]. The network learns by the backpropagation rule of Rumelhart et al.[28] under the constraint that the change to each weight for each instance is bounded. With this learning rule, it is likely that adjustments of different weights may be truncated at different proportions. As a result, the network weight vector may not move in the steepest descent during error minimization. In IBPLN, this problem is dealt with by introducing a scaling factor *s* which scales down all weight adjustments so that all of them are within bounds. The learning rule is now

$$\Delta W_{ij}(k) = s(k)\, \eta \delta_j(k) O_i(k) \qquad (2)$$

where W_{ij} is the weight from unit i to unit j, η($0< \eta < 1$) is a trial-independent learning rate, δ_j is the error gradient at unit j, O_i is the activation level at unit i, and the parameter k denotes the k-th iteration. In the incremental learning scheme, initial weights prior to learning any new instance represent knowledge accumulated so far. IBPLN introduced *two* structural adaptations; neuron generation and neuron elimination. The IBPLN proceeds as follows [27]:

Given a single misclassified instance:

Begin
    Repeatedly apply the bounded weight adaptation
    learning rule (2) on the instance until stopping
    criteria are met.
If
    the instance can be correctly learned, then restore
    the old weights and apply the bounded weight
    adaptation learning rule once;
Else
    restore the old weights and apply the structural adaptation learning rules.
End.

The stopping criteria are: The instance can be correctly learned or the output error fluctuates in a small range.

4.3. Levenberg-Marquardt(LM) Algorithm

The Levenberg-Marquardt (LM) algorithm is basically an iterative method that locates the minimum of a multivariate function that is expressed as the sum of squares of non-linear real-valued functions [29], [30]. LM can be thought of as a combination of steepest descent and the Gauss-Newton (GN) method. LM algorithm is more robust than GN algorithm which essentially means that it finds a solution even if it starts far off the final minimum. During the iterations, the new configuration of weights in step k+1 is calculated as follows

$$w(k+1) = w(k) - (J^T J + \lambda I)^{-1} J^T \varepsilon(k) \qquad (3)$$

where J – the Jacobian matrix, λ - adjustable parameter, ε - error vector. The parameter λ is modified based on the development of error function E. If the step causes a reduction of E, we accept it. Otherwise, λ is changed; reset the original value and recalculate w(k+1).

5. APPLICATIONS

Basically, this study consists of two stages: The feature extraction and reduction phase by correlation-based feature selection (CFS); and classification phase by incremental back propagation neural networks (IBPLN), and Levenberg-Marquardt (LM) algorithm. The schematic view of our system is shown in Fig. 3.

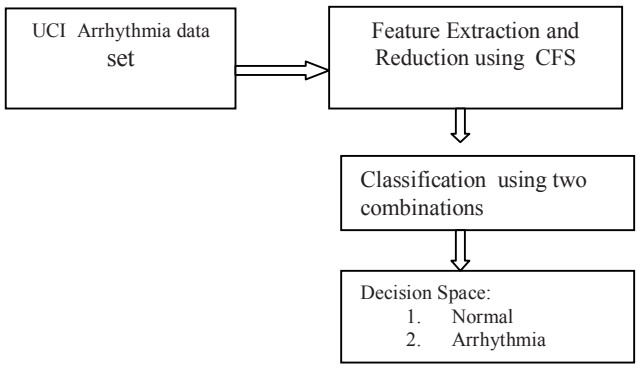

Fig. 3. Block diagram of a system for arrhythmia diagnosis

5.1. Data Preprocessing

Data preprocessing is the primary step for any model development. We deleted columns with all 0's and missing values and we also deleted columns of which most of the elements are 0's. We got 182 columns of which 9 are categorical and 173 are numerical. Next, we deleted 32 rows with missing values and the rest 420 records were considered for analysis. We completely randomize the data sets after missing records deletion. There is no outlier in our data. The data set is partitioned into three: Training set (68%), Validation set (16%), and Test set (16%).

5.2. Feature Selection and Extraction

We apply CFS with linear forward selection search method as attribute evaluators. We got 18 attributes as reduced feature set as follows: Sex, QRS duration, DII 49, DII 76, DII91, DII103, DII112, DI163, DI167, DI169, DII173, DII 199, DII207, DII211, DII261, DII267, DII271, and DII277.

5.3. Network Architecture

In general, balancing the trade-off between accuracy and generalizability is the prime characteristic of selecting a model. The ANN model selection includes choice of network architecture and feature selection. The hold-out data set called the *validation set* would be useful helping all these decisions successful [31]. Validation set is a part of our data used to tune the network topology or network parameters other than weights. In our networks, we use logistic function of the form $F(x) = 1/ (1+e^{-x})$ in the hidden and output nodes. Theoretically, a network with one hidden layer and logistic function as the activation function at the hidden and output nodes is capable of approximating any function arbitrarily closely, provided that the number of hidden nodes are large enough [32]. So, we use one input layer, one hidden layer, and one output layer. The number of neurons in ANN is always a problem; too few hidden nodes as well as too many hidden nodes have certain problems. To overcome the problem, the formula proposed by Goa [33] was used in our study. The said formula is as follows:

$$s = \sqrt{(a_1m^2 + a_2mn + a_3n^2 + a_4m + a_5n + a_6)} + a_7 \qquad (4)$$

where s is the number of neuron, m is the number of input, n is the number of output, $a_1 \sim a_7$ are undefined coefficients. Using least mean square (LMS), Huang et al. [34] derived the following formula:

$$s= \sqrt{(0.43mn + 0.12n^2 + 2.54m + 0.77n + 0.35)} + 0.51 \qquad (5)$$

In the present study, m = 18, n =2; and hence s= 8 after round off. So, in our study, we use eight neurons at the hidden layer for all combinations. So, our network becomes 18-8-1. As a overtraining control measure, we retain the copy of the network with the lowest validation error.

6. MODELING RESULTS

WEKA [35] was used for feature set reduction using CFS. The classification algorithms using two combinations were implemented in Alyuda NeuroIntelligence [36]. All the mentioned tasks were executed on Intel Core Solo T1350 CPU(1.86GHz, 533MHz FSB, 2MB L2 cache) with 512MB DDR2 RAM.

6.1. Performance Evaluation Method

As the performance measure, we compute the classification accuracy, sensitivity, specificity and AUC. The formulations are as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \qquad (6)$$

$$\text{Sensitivity} = \frac{TP}{TP+TN} \text{ x } 100\% \quad (7)$$

$$\text{Specificity} = \frac{TN}{FP+TN} \text{ x } 100\% \quad (8)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives respectively. Moreover, the area under ROC curve (AUC) is an important measure of classification performance that is being used in biomedical research to assess the performance of diagnostic tests [37]. AUC close to one indicates more reliable diagnostic result [37] and it is considered one of the best methods for comparing classifiers in two-class problems.

6.2. Experimental Results

The compiled results from 100 simulations of our studies are shown in table IA and table IB.
The following observations are noted below:

- Out of the four methods, CFS+LM shows best performance in terms of CCR, Specificity, Sensitivity, and AUC.
- All of the methods provide 100% sensitivity as the highest performance using 68% training, 16% validation, and 16% testing data sets.

Table IA: Results from 100 simulations

Methods	Sensitivity			AUC		
	Highest (freq)	Lowest (freq)	Avg	Highest (freq)	Lowest (freq)	Avg
CFS + IBPLN	100 (31)	60(2)	84.7	95(1)	62(1)	79.26
CFS + LM	**100 (33)**	**55.56 (1)**	**86.72**	**96(1)**	**60(1)**	**81.88**

Table IB: Results from 100 simulations

Methods	Test set (CCR%)			Specificity		
	Highest (freq)	Lowest (freq)	Avg	Highest (freq)	Lowest (freq)	Avg
CFS + IBPLN	92.68(5)	73.17 (1)	86.02	96.77(1)	72.97(1)	86.65
CFS + LM	**95.12(3)**	**80.49 (3)**	**87.71**	**100(1)**	**79.17(1)**	**88.38**

7. CONCLUSIONS

This work has explored two combinations of intelligent diagnostic systems for cardiac arrhythmia diagnosis. We argue that the lowest performance should also be a judging parameter for the performance of a system. So, we present here the highest, lowest, and the average behavior of the methods used.

This work provides a better result in terms of average classification accuracy compared to the results of some recent works [38, 39] on the same data set. In a significant number of cases, sensitivity has reached 100%. It is proposed that CFS derived set of reduced features would have been worthwhile when the final decision is made by the doctors.

References

[1]Y. Ozbay and B. Karlik, *A Recognition of ECG Arrhythmia using Artificial Neural Networks,* Proc. 23rd Annual Conference, IEEE/EMBS, Istanbul, Turkey, 2001.

[2]http://archive.ics.uci.edu/ml/datasets/ UCI Machine Learning Repository.

[3]K. Acharya and Bhatt, *Classification of cardiac abnormalities using heart rate signals,* Medical & Biological Engineering & Computing, 42(3), 288-293, 2004.

[4]N. Belgacem et. al., *Supervised classification of ECG using neural networks,* www.univ-tlemcen.dz/manifest/CISTEMA2003/../GBM8.pdf

[5[R. Ceylan, Y. Ozbay and B. Karlik, *A novel approach for classification of ECG arrhythmias: Type-2 fuzzy clustering neural network,* Expert Systems with Applications, 6721-6726, 2009.

[6]M. Engin, *ECG beat classification using neuro-fuzzy network,* Pattern Recognition Letters, 25, 1715-1722, 2004.

[7]Al-Fahoum, Khadra & Al-Nashash, *Detection of life-threatening cardiac arrhythmias using wavelets transformation,* Medical and Biological Engineering Computing, 35(6), 626-632, 1997.

[8]D. Patra, M. K. Das and S. Pradhan, *Integration of FCM, PCA and neural networks for classification of ECG arrhythmias,* IAENG Int. J. of Computer Science, 36(3), 2009.

[9]Y. Sun and K. Chan, *Arrhythmia detection and recognition in ECG signals using nonlinear techniques,* Annals of Biomedical Engineering, 28, 37, 2000.

[10]D. Calisir, and E. Dogantekin, *A new intelligent hepatitis diagnosis system: PCA-LSSVM,* Expert Systems with Applications, vol. 38, pp. 10705-10708, 2011.

[11]E. Avci, *A new optimum feature extraction and classification method for speaker recognition: GWPNN*, Expert Systems with Applications, vol. 32, no. 2, pp. 485-498, 2007.

[12]M. A. Hall, *Correlation-based Feature Subset Selection for Machine Learning,* Hamilton, New Zealand, 1998.

[13]M. Karabatak, and M. C. Ince, *An expert system for detection of breast cancer based on association rules and neural network,* Expert Syst Appl, vol. 36, no. 2, pp. 3465-3469, 2009.

[14]H. L. Chen et al. A *support vector machine classifier with rough set based feature selection for breast cancer diagnosis,* Expert Syst Appl, vol. 38, no. 7, pp. 9014-9022, 2011.

[15]U. M. Fayyad, and K. B. Irani, *Multi-interval Discretisation of Continuous-valued Attributes for Classification learning,* Proc. of the XIIIth Int. Joint Conf. on AI, Morgan Kaufmann, 1993.

[16]J. R. Quinlan, C 4.5: Programs for Machine Learning, Morgan Kaufmann, 1993.

[17]E. Rich, and K. Knight, Artificial Intelligence, McGraw-Hill, 1991.

[18]L. Yu, and H. Liu, *Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution,* Machine Learning, vol. 20, issue 2, pp 856, 2003.

[19]K. Michalak, and H. Kwasnicka, *Correlation-Based Feature Selection Strategy in Classification Problems*, Int. J. Appl. Math. Comput. Sci, vol. 16, no. 4, pp. 503-511, 2006.

[20]W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity,* Bulletin of Mathematical Biophysics, vol. 7, pp. 115-133, 1943.

[21]D. O. Hebb, The Organization of Behavior, a Neuropsychological Theory, New York, John Wiley, 1949.

[22]A. Roy, *Artificial Neural Networks- A Science in Trouble,* SIGKDD Explorations, vol. 1, issue 2, pp. 33-38, 2000.

[23]D. E. Rumelhart, J. L. McClelland (eds.), Parallel Distributed Processing: Explorations in Microstructures of Cognition, vol. 1: Foundations, MIT Press, Cambridge, M.A., pp. 318-362, 1986.

[24]D. E. Rumelhart, The Architecture of Mind: A Connectionist Approach, Chapter 8 in J. Haugeland (ed.), Mind_design II, 1997, MIT Press, pp. 205-232, 1986.

[25]S. Grossberg, *Nonlinear Neural Networks: Principles, Mechanisms, and Architectures,* Neural Networks, vol. 1 , pp. 17 -61, 1988.

[26]J. Moody and C. Darken, *Fast Learning in Networks of Locally-Tuned Processing Units,* Neural Computation, vol. 1, pp. 281-294, 1989.

[27]L. Fu, H. Hsu, and J. C. Principe, *Incremental Backpropagation Learning Networks,* IEEE Trans. on Neural Networks, vol. 7, no.3, pp. 757-761, 1996.

[28]D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning internal representation by error propagation,* in Parallel Distributed Processing: Explorations in the Microstructures of Cognition, MA, MIT Press, vol. 1, 1986.

[29]K. Levenberg, *A method for the solution of certain non-linear problems in least squares,* Quarterly in Applied Mathematics, vol. 2, no. 2, pp. 164-168, 1944.

[30]D. W. Marquardt, *An algorithm for the least-squares estimation of nonlinear parameters,* SIAM Journal of Applied Mathematics, vol. 11. No. 2, pp. 431-441, 1963.

[31]M. S. Hung, M. Shankar, M. Y. Hu, *Estimating Breast Cancer Risks Using Neural Networks,* J. Operational Research Society, Vol. 52, pp. 1-10, 2001.

[32]K. Hornik, M. Stinchcombe, H. White, *Multilayer feedforward networks are universal approximator,* Neural Network, Vol.2, pp. 359-366, 1991.

[33]D. Goa, *On structures of supervised linear basis function feedforward three-layered neural networks*, Chin. J. Comput., vol. 21, no. 1, pp. 80-86, 1998.

[34]M.L. Huang, Y.H. Hung, and W.Y. Chen, *Neural network classifier with entropy based feature selection on breast cancer diagnosis,* J. Med. Syst, 34(5), pp. 865-873, 2010.

[35]E. Hall, G. Frank, B. Holmes, P. Pfahringer, I. Reutemann, H. Witten, *The WEKA Data Mining Software: An Update*, SIGKDD Explorations, vol. 11, no. 1, 2009.

[36]Alyuda NeuroIntelligence 2.2, http://www.alyuda.com.

[37]A. P. Bradley, *The use of the area under the ROC curve in the evaluation of machine learning algorithms,* Pattern Recognition, vol. 30, no. 7, pp. 1145-1159, 1997.

[38]S. M. Jadhav, S.L. Nalbalwar and A. A. Ghatol, *Artificial Neural Network Models based Cardiac Arrhythmia Disease Diagnosis from ECG Signal Data,*Int. J. of Computer Applications, 44(15), 8-13, 2012.
[39]S. M. Jadhav, S.L. Nalbalwar and A. A. Ghatol, *Performance Evaluation of Generalized Feedforward Neural Network based ECG*