IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011
ISSN (Online): 1694-0814
www.IJCSI.org

Classification of Electrocardiogram Signals With Extreme Learning Machine and Relevance Vector Machine

S. Karpagachelvi,

Doctoral Research Scholar, Mother Teresa Women's University,

Kodaikanal, Tamilnadu, India.

Dr. M. Arthanari,

Director, Bharathidasan School of Computer Applications,

Erode- 638 116, Tamilnadu, India.

M.Sivakumar,

Doctoral Research Scholar,Anna University – Coimbatore,
Tamilnadu, India

***Abstract*—**The ECG is one of the most effective diagnostic tools to detect cardiac diseases. It is a method to measure and record different electrical potentials of the heart. The electrical potential generated by electrical activity in cardiac tissue is measured on the surface of the human body. Current flow, in the form of ions, signals contraction of cardiac muscle fibers leading to the heart's pumping action. This ECG can be classified as normal and abnormal signals. In this paper, a thorough experimental study was conducted to show the superiority of the generalization capability of the Relevance Vector Machine (RVM) compared with Extreme Learning Machine (ELM) approach in the automatic classification of ECG beats. The generalization performance of the ELM classifier has not achieved the nearest maximum accuracy of ECG signal classsification. To achieve the maximum accuracy the RVM classifier design by searching for the best value of the parameters that tune its discriminant function, and upstream by looking for the best subset of features that feed the classifier. The experiments were conducted on the ECG data from the Massachusetts Institute of Technology–Beth Israel Hospital (MIT– BIH) arrhythmia database to classify five kinds of abnormal waveforms and normal beats. In particular, the sensitivity of the RVM classifier is tested and that is compared with ELM. Both the approaches are compared by giving raw input data and preprocessed data. The obtained results clearly confirm the superiority of the RVM approach when compared to traditional classifiers.

***Index Terms*—**Electrocardiogram (ECG) signals classification, feature detection, feature reduction, generalization capability, model selection issue, extreme learning machine (ELM), relevance vector machine (RVM).

I. INTRODUCTION

The recognition of the ECG beats is an extremely important task in the coronary intensive unit, where the classification of the ECG beats is essential tool for the diagnosis. ECG is a technique which captures transthoracic interpretation of the electrical activity of the heart over time and externally recorded by skin electrodes. It is a non persistent recording produced by an electrocardiographic device. ECG offers cardiologists with useful information about the rhythm and functioning of the heart. Therefore, its analysis represents an efficient way to detect and treat different kinds of cardiac diseases.A typical structure of the ECG signal is shown in Figure 1. The ECG signal is usually divided into two phases: depolarization and repolarization phases. The depolarization phase corresponds to the P-wave and QRS-wave while repolarization phase corresponds to the T-wave and U-wave. The ECG is measured by placing ten electrodes on preferred spots on the human body surface. For regular ECG recordings, the deviations in electrical potentials in 12 different directions out of the ten electrodes are measured. These 12 different electrical observations of the activity in the heart are normally referred to as leads. Trained physicians are able to recognize certain patterns in a patient's ECG signal and use them as the basis for diagnosis. Researchers have tried as the inception of computers to develop techniques and algorithms for automated processing of ECG signals for various medical applications.

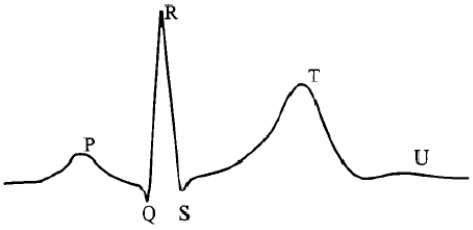

Figure 1: Structure of the ECG signal

Up to now; numerous algorithms have been introduced for the recognition and classification of ECG signal. Some of them use time and some use frequency domain for depiction. Based on that many specific attributes are defined, allowing the recognition between the beats belonging to different pathological classes. The ECG waveforms may be different for the same patient to such extent that they are unlike each other and at the same time alike for different types of beats [1]. Artificial neural network (ANN) and fuzzy-based techniques were also employed to exploit their natural ability in pattern recognition task for successful classification of ECG beats [2].

In this paper, the approach to ECG beat classification presented thorough experimental exploration of the RVM capabilities for ECG classification. Further the performances of the RVM approach in terms of classification accuracy are evaluated: 1) by automatically detecting the best discriminating features from the whole considered feature space and 2) by solving the model selection issue. Unlike traditional feature selection methods, where the user has to specify the number of desired features, the proposed system gives a method for extraction of features called as "feature detection". Feature selection and feature detection have the common characteristic of searching for the best discriminative features. The latter, however, has the advantage of determining their number automatically. In other words, feature detection does not require the desired number of most discriminative features from the user apriori. The detection process is implemented through AR Modeling framework that exploits a criterion intrinsically related to RVM classifier properties. This framework is formulated in such a way that it also solves the model selection issue, i.e., to estimate the best values of the RVM classifier parameters, which are the regularization and kernel parameters.

The rest of the paper is organized as follows. The AR method for ECG feature extraction, the basic mathematical formulation of ELMs for solving binary and multiclass classification problems and the working methodology of RVM is given in Section III. The experimental results obtained on ECG data from the Massachusetts Institute of Technology–Beth Israel Hospital (MIT–BIH) arrhythmia database [9] are reported in Sections IV. Finally, conclusions are drawn in Section V.

II. LITERATURE STUDY

In the literature survey, several methods have been proposed for the automatic classification of ECG signals. Among the most recently published works are those presented as follows

F. de Chazal et.al,[4] investigates the design of an efficient system for recognition of the premature ventricular contraction from the normal beats and other heart diseases. This system comprises three main modules: denoising module, feature extraction module and classifier module. In the denoising module of the system, it proposed the stationary wavelet transform for noise reduction of the electrocardiogram signals. In the feature extraction of the ECG module a proper combination of the morphological-based features and timing interval-based features are proposed. As the classifier, many supervised classifiers are investigated; they are: a number of multi-layer perceptron neural networks with different number of layers and training algorithms, support vector machines with different kernel types, radial basis function and probabilistic neural networks. Also, for comparison the proposed features, the author has considered the wavelet-based features. It has done complete simulations in order to achieve a high efficient system for ECG beat classification from 12 files obtained from the MIT–BIH arrhythmia database. Simulation results show that best results are achieved about 97.14% for classification of ECG beats.

R. V. Andreao *et.al*,[5] proposed a novel embedded mobile ECG reasoning system that integrates ECG signal reasoning and RF identification together to monitor an elderly patient. As a result, the proposed method by andreao has a good accuracy in heart beat recognition, and enables continuous monitoring and identification of the elderly patient when alone. Furthermore, in order to examine and validate our proposed system, the author proposes a managerial research model to test whether it can be implemented in a medical organization. The results prove that the mobility, usability, and performance of author's proposed system have impacts on the user's attitude, and there is a significant positive relation between the user's attitude and the intent to use the proposed system.

L. Khadra *et.al*,[3] proposed a high order spectral analysis technique for quantitative analysis and classification of cardiac arrhythmias. The algorithm is based upon bispectral analysis techniques. The bispectrum is estimated with the use of an autoregressive model, and the frequency support of the bispectrum is extracted as a quantitative measure to classify a trial and ventricular tachyarrhythmias. Results illustrate a significant difference in the parameter values for different arrhythmias. Furthermore, the bicoherency spectrum shows different bicoherency values for normal and tachycardia patients. In particular, the bicoherency points out that phase coupling decreases as arrhythmia kicks in. The ease of the classification parameter and the attained specificity and sensitivity of the classification scheme reveal the importance of higher order spectral analysis in the classification of life threatening arrhythmias.

S. Mitra et.al, [6] puts forth a three stage technique for detection of premature ventricular contraction (PVC) from normal beats and other heart diseases. This method comprises a denoising module, a feature extraction module and a classification module. In the first module the author investigates the application of stationary wavelet transform (SWT) for noise reduction of the electrocardiogram (ECG) signals. The feature extraction module obtains 10 ECG morphological features and one timing interval feature. Then several multilayer perceptron (MLP) neural networks with different number of layers and nine training algorithms are designed. The performances of the networks for the speed of convergence and its accuracy classifications are evaluated for seven files from the MIT–BIH arrhythmia database. Among the various training algorithms, the resilient back-propagation (RP) algorithm illustrated the best convergence rate and the Levenberg–Marquardt (LM) algorithm achieved the best overall detection accuracy.

Chuan-Min Zhai et al, [20] presented a classification approach using ELM. In his paper, a machine learning algorithm referred to as the extreme learning machine (ELM) is used to classify plant species through plant leaf Gabor texture feature. A comparative study on system performance is

carried out among ELM and the main conventional neural network classifier - backpropagation neural networks. Results illustrate that the classification accuracy of ELM is higher than that of BP network. For given network architecture, ELM doesn't have any control parameters (i.e, stopping criteria, learning rate, learning epoches, etc.) to be manually tuned and can be implemented easily.

Sheng-Wu Xiong et.al.,[8] proposed in their paper that fuzzy support vector machines based on fuzzy c-means clustering. They employed the fuzzy c-means clustering technique to each class of the training set. At the time of clustering with a appropriate fuzziness parameter q, the more important samples, such as support vectors, become the cluster centers respectively.

Xin Zhou et al, [16] presented a novel approach on modulation classification using RVM. In his paper, a novel classification method based on relevance vector machine (RVM) is used in the MPSK signals classification. Compared with the SVM, RVM is sparse model in the Bayesian framework, not only the solution is highly sparse, but also it does not need to adjust model parameter and its kernel functions don't need to satisfy Mercer's condition. The fourth order cumulants of received signals are used as the classification vector firstly, and then multi-class classifier of RVM is designed. The authors first introduce the sparse Bayesian classification model, then transform the RVM learning to the maximization of marginal likelihood, and select the fast sequential sparse Bayesian learning algorithm. With the results of the experiment compared with SVM classifier proves the advantage of RVM.

Ke Wang et al, [15] discussed on the image classification technique RVM. In his paper, four building categories for database is prepared. Firstly the author uses the Gabor filter for image processing to extract the image features, and then divide the images to different subregions for histogram-based Gabor features. At last, for image classification, Support Vector Machine (SVM) and Relevance Vector Machine (RVM) are known to outperform classical supervised classification algorithms. SVM has excellent performance to solve binary classification problems. RVM could be more sparsity than SVM. A new method based on relevance vector machine- No-balance Binary Tree Relevance Vector Machine (NBBTRVM) is proposed to define a class in this classification task. NBBTRVM could do a good performance according to our experiment results.

III. METHODOLOGY

3.1. Feature extraction

Automatic ECG beat recognition and classification is performed in the part either by the neural network or by the other recognition systems relying in various features, time domain representation, extracted from the ECG beat [2], or the measure of energy in a band of frequencies in the spectrum (frequency domain representation) [10]. Since these features are very susceptible to variations of ECG morphology and the temporal characteristics of ECG, it is difficult to distinguish one from the other on the basis of the time waveform or frequency representation. In this paper three different classes of feature set are used belonging to the isolated ECG beats including; third-order cumulant, auto-regressive model parameters and the variance of discrete wavelet transform detail coefficients for the different scales (1–6 scales).

3.1.1. Wavelet transformation

Physiological signals used for diagnosis are frequently characterized by a non-stationary time behavior. For such patterns, time and frequency representations are desirable. The frequency characteristics in addition to the temporal behavior can be described with respect to uncertainty principle. The wavelet transform can represent signals in different resolutions by dilating and compressing its basis functions. While the dilated functions adapt to slow wave activity, the compressed functions captures fast activity and sharp spikes. The most favorable choice of types of wavelet functions for pre-processing is problem dependent. In this paper Daubechies wavelet function (db5) which is called compactly supported orthonormal wavelets [11]. By making discretization the scaling factor and position factor the DWT is obtained. For orthonormal wavelet transform, x(n) the discrete signal can be expanded in to the scaling function at j level, as follows:

$$x(n) = D_{j,k}[x(n)] + A_{j,k}[x(n)], n \in Z \qquad (1)$$

where $D_{j,k}$ represents the detailed signal at j level. Note that j controls the dilation or contraction of the scale function $\Phi(t)$ and k denotes the position of the wavelet function $\Psi(t)$, and n represents the sample number of the x(n). Here $n \in Z$ represents the set of integers. The frequency spectrum of the signal is classified into high frequency and low frequency for wavelet decomposition as the band increases (j = 1, . . . , 6). Wavelet transform is a two-dimensional timescale processing method for non-stationary signals with adequate scale values and shifting in time [12].

Multi resolution decomposition can efficiently provide simultaneous characteristics, in term of the representation of the signal at multiple resolutions corresponding to different time scales. Feature vectors are constructed by the normalized variances of detail coefficients of the DWT which belongs to the related scales.

3.1.2. Higher-order statistics and AR modeling

The main problem in automatic ECG beat recognition and classification is that related features are very susceptible to variations of ECG morphology and temporal characteristics of ECG. In the study [1] the set of original QRS complexes typical for six types of arrhytmia taken from the MIT/BIH arrhytmia database, there is a great variations of signal among the same type of beats belonging to the same type of arrhytmia. Therefore, in order to solve such problem, the author will rely on the statistical features of the ECG beats. In this paper for this aim, third-order cumulant has been taken

into account, which can be determined (for zero mean signals) as follows

$$C_{2x}(k) = E\{x(n)x(n+k)\} \quad (2)$$

$$C_{3x}(k,l) = E\{x(n)x(n+k)x(n+l)\} \quad (3)$$

$$C_{4x}(k,l,m) = E\{x(n)x(n+k)x(n+l)x(n+m)\} - C_{2x}(k)C_{2x}(m-l) - C_{2x}(l)C_{2x}(m-k) - C_{2x}(m)C_{2x}(l-k) \quad (4)$$

Where E represents the expectation operator, and k, l, and m are the time lags. In this paper, third-order cumulant of selected ECG beats is used. Normalized ten points represents the cumulant evenly distributed with in the range of 25 lags. Each succeeding samples of a signal as a linear combination of previous samples, that is, as the output of an all-pole IIR filter is modeled by linear prediction. This process locates the coefficients of an n^{th} order auto-regressive linear process that models the time series x as

$$x(k) = -a(2)x(k-1) - a(3)x(k-2) - \cdots - a(n+1)x(k-n-1) \quad (5)$$

where x represents the real input time series (a vector), and n is the order of the denominator polynomial a(z). In the block processing, autocorrelation method is one of the modeling methods of all-pole modeling to find the linear prediction coefficients. This method is as well called as the maximum entropy method (MEM) of spectral analysis.

3.2. Extreme Learning Machine

A new learning algorithm called the Extreme Learning Machine for Single-hidden Layer Feed forward neural Networks (SLFNs) supervised batch learning. The output of an SLFN with ~N hidden nodes (additive or RBF nodes) can be represented by

$$f_{\tilde{N}}(X) = \sum_{i=1}^{\tilde{N}} \beta_i G(a_i, b_i, X), \quad X \in Rn, \; ai \in Rn, \quad (6)$$

where a_i and b_i are the learning parameters of hidden nodes and β_i is the weight connecting the ith hidden node to the output node. $G(a_i,b_i,X)$ is the output of the ith hidden node with respect to the input x. For the additive hidden node with the activation function $g(x):R \rightarrow R$ (e.g., sigmoid or threshold), $G(a_i,b_i,X)$ is given by

$$G(ai, bi, X) = g(a_i . X + b_i), b_i \in R \quad (7)$$

Where ai represents the weight vector connecting the input layer to the i^{th} hidden node and b_i is the bias of the i^{th} hidden node. $a_i.x$ denotes the inner product of vectors a_i and x in R^n. For an RBF hidden node with an activation function $g(x):R \rightarrow R$(e.g., Gaussian), $G(a_i,b_i,X)$ is given by

$$G(ai, bi, X) = g\left(b_i \left| |x - a_i| \right|\right), b_i \in R^+ \quad (8)$$

Where a_i and b_i are the i^{th} RBF node's center and impact factor. R^+ indicates the set of all positive real values. The RBF network is a special case of the SLFN with RBF nodes in its hidden layer. Each RBF node has its own centroid and impact factor and output of it is given by a radially symmetric function of the distance between the input and the center.

In the learning algorithms it uses a finite number of input-output samples for training. Here, N arbitrary distinct samples are considered $(x_i,t_i) \in R^n \times R^m$, where x_i is an n x 1 input vector and t_i is an m x 1 target vector. If an SLFN with $\tilde{N}$ hidden nodes can approximate N samples with zero error, it then implies that there exist β_i, a_i, and b_i such that

$$f_{\tilde{N}}(X_j) = \sum_{i=1}^{\tilde{N}} \beta_i G(a_i, b_j, X_j) = t_j, \; j = 1, \ldots, N \quad (9)$$

Equation (9) can be written compactly as

$$H\beta = T \quad (10)$$

Where

$$H(a_1, \ldots, a_{\tilde{N}}, b_1, \ldots, b_{\tilde{N}}, X_1, \ldots, X_{\tilde{N}}) = \quad (11)$$

$$\begin{bmatrix} G(a_1,b_1,X_1) & \cdots & G(a_{\tilde{N}},b_{\tilde{N}},X_1) \\ \vdots & \ddots & \vdots \\ G(a_1,b_1,X_N) & \cdots & G(a_{\tilde{N}},b_{\tilde{N}},X_N) \end{bmatrix}_{N \times \tilde{N}}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \quad \text{and} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \quad (12)$$

H is called the hidden layer output matrix of the network [15]; the i^{th} column of H is the i^{th} hidden node's output vector with respect to inputs x_1, x_2,…, x_N and the j^{th} row of H is the output vector of the hidden layer with respect to input x_j.

In real applications, the number of hidden nodes, $\tilde{N}$, will always be less than the number of training samples, N, and, hence, the training error cannot be made exactly zero but can approach a nonzero training error. The hidden node parameters a_i and b_i (input weights and biases or centers and impact factors) of SLFNs need not be tuned during training and may simply be assigned with random values according to any continuous sampling distribution. Equation (12) then becomes a linear system and the output weights are estimated as

$$\tilde{\beta} = H \dagger T \quad (13)$$

Where $H \dagger$ the Moore-Penrose is generalized inverse [15] of the hidden layer output matrix H. The ELM algorithm which consists of only three steps, can then be summarized as

ELM Algorithm: Given a training set $\aleph = \{(X_i, t_i) | X_i \in R^n, t_i \in R^m, i = 1, \ldots, N\}$ activation function $g(x)$, and hidden node number $\tilde{N}$,

1) Assign random hidden nodes by randomly generating parameters (a_i,b_i) according to any continuous sampling distribution, i=1,…., $\widetilde{N}$

2) Calculate the hidden layer output matrix H.

3) Calculate the output weightβ: $\tilde{\beta} = H \dagger T$

The universal approximation capability of ELM has been analyzed by Huang et al. [7] using an incremental method and it shows that single SLFNs with randomly generated additive or RBF nodes with a wide range of activation functions can universally approximate any continuous target functions in any compact subset of the Euclidean space R^n. $g(x) = \frac{1}{1+e^{-\lambda x}}$ is the sigmoidal function used as activation function in ELM.

3.3. *Relevance Vector Machine*

The relevance vector machine (RVM) classifier [14], is a probabilistic extension of the linear regression model, which provides sparse solutions. It is analogous to the SVM, since it computes the decision function using only few of the training examples, which are now called relevance vectors. However training is based on different objectives.

The RVM model y(x ; w) is output of a linear model with parameters w = (w1, . . . , wN)^T , with application of a sigmoid function for the case of classification:

$$y_{RVM}(x) = \sigma(\sum_{n=1}^{N} \omega_n K(x, x_n)) \quad (14)$$

where σ(x) = 1/(1+exp(x)). In the RVM, sparseness is achieved by assuming a suitable prior distribution on the weights, specifically a zero-mean, Gaussian distribution with distinct inverse variance α_n for each weight ω_n:

$$p(\omega|\alpha) = \prod_{n=1}^{N} N(\omega_n|0, \alpha_n^{-1}) \quad (15)$$

The variance hyperparameters α = (α_1,..., α_N) are assumed to be Gamma distributed random variables:

$$p(\alpha) = \prod_{n=1}^{N} Gamma(\alpha_n|a, b) \quad (16)$$

The parameters a and b are implicitly fixed and usually they are set to zero (a = b = 0), which provides sparse solutions.

Given a training set $\{x_n, t_n\}_{n=1}^{N}$ with $t_n \in \{0,1\}$ training in RVM is equivalent to compute the posterior distribution $p(\omega, \alpha|t)$. However, since this computation is intractable, a quadratic approximation $\log p(\omega, \alpha|t) \approx (\omega - \mu)^T \sum^{-1} (\omega - \mu)$ is assumed and computed matrix Σ and vector μ as:

$$\Sigma = (\Phi^T B \Phi + A)^{-1} \quad (17)$$

$$\mu = \Sigma \Phi^T B \hat{t} \quad (18)$$

with the N × N matrix Φ described as [Φ]ij = K(xi, xj), A = diag(α1, . . . , αN), B = diag(β1, . . . , βN), βn = yRVM(xn)[1−yRVM(xn)] and ˆt = Φμ+B¡1(t−y). The parameters α are set to the values αMP that maximize the logarithm of the following marginal likelihood

$$L(\alpha) = logp(\alpha|t) = -\frac{1}{2}[Nlog2\pi + \log|C| + t^T C^{-1} t], \quad (19)$$

with C = B^{-1} + $\Phi A^{-1} \Phi^T$. This, gives the following update formula:

$$\alpha_n = \frac{1 - \alpha_n \Sigma_{nn}}{\mu_n^2} \quad (20)$$

The RVM learning algorithm iteratively evaluates formulas (15),(16) and (18). After training, the value of yRVM(x) = y(x; μ) can be used to estimate the reliability of the classification decision for input x. Values close to 0.5 are near the decision boundary and consequently are unreliable classifications, while values near 0 and near 1 should correspond to reliable classifications. In this experiment, the reliability measure is used

$$RE_{RVM} = |2_{yRVM}(x) - 1| \quad (21)$$

which uses values near 0 for unreliable classifications and near 1 for reliable classifications.

IV. EXPERIMENTAL RESULTS

4.1. Dataset Description

The experiment conducted on the basis of ECG data from the MIT–BIH arrhythmia database [9]. In particular, the considered beats refer to the following classes: normal sinus rhythm (N), atrial premature beat (A), ventricular premature beat (V), right bundle branch block (RB), left bundle branch block (LB), and paced beat (/). The beats were selected from the recordings of 20 patients, which correspond to the following files: 100, 102, 104, 105, 106, 107, 118, 119, 200, 201, 202, 203, 205, 208, 209, 212, 213, 214, 215, and 217. In order to feed the classification process, in this paper, the two following kinds of features are adopted: 1) ECG morphology features and 2) three ECG temporal features, i.e., the QRS complex duration, the RR interval (the time span between two consecutive R points representing the distance between the QRS peaks of the present and previous beats), and the RR interval averaged over the ten last beats [4]. In order to extract these features, first the QRS detection is performed and ECG wave boundary recognition tasks by means of the well-known ecgpuwave software available on [17]. Then, after extracting the three temporal features of interest, normalized to the same periodic length the duration of the segmented ECG cycles according to the procedure reported. To this purpose, the mean

beat period was chosen as the normalized periodic length, which was represented by 300 uniformly distributed samples. Consequently, the total number of morphology and temporal features equals 303 for each beat.

TABLE1
NUMBERS OF TRAINING AND TEST BEATS USED IN THE EXPERIMENTS

Class	N	A	V	RB	/	LB	Total
Training beats	150	100	100	50	50	50	500
Test beats	24000	245	3789	3893	6689	1800	40416

In order to obtain reliable assessments of the classification accuracy of the investigated classifiers, in all the following experiments, three different trials are performed, each with a new set of randomly selected training beats, while the test set was kept unchanged. The results of these three trials obtained on the test set were thus averaged. The detailed numbers of training and test beats are reported for each class in Table 1. Classification performance was evaluated in terms of four measures, which are: 1) the overall accuracy (OA), which is the percentage of correctly classified beats among all the beats considered (independently of the classes they belong to); 2) the accuracy of each class that is the percentage of correctly classified beats among the beats of the considered class; 3) the average accuracy (AA), which is the average over the classification accuracies obtained for the different classes; 4) theMcNemar's test that gives the statistical significance of differences between the accuracies achieved by the different classification approaches. This test is based on the standardized normal test statistic [16]

$$Z_{ij} = \frac{f_{ij} - f_{ji}}{\sqrt{f_{ij} - f_{ji}}} \qquad (22)$$

where Z_{ij} measures the pair wise statistical significance of the difference between the accuracies of the i^{th} and j^{th} classifiers. f_{ij} stands for the number of beats classified correctly and wrongly by the i_{th} and j_{th} classifiers, respectively. Accordingly, f_{ij} and f_{ji} are the counts of classified beats on which the considered i_{th} and j_{th} classifiers disagree. At the commonly used 5% level of significance, the difference of accuracies between the i_{th} and j_{th} classifiers is said statistically significant if $|Z_{ij}| > 1.96$.

4.2. Experimental Scheme

The proposed experimental framework was performed around the following four main experiments. The first experiment aimed at assessing the effectiveness of the RVM approach in classifying ECG signals directly in the whole original hyper dimensional feature space (i.e., by means of all the 303 available features). The total number of training beats was fixed to 500, as reported in Table 1. In the second experiment, it was desired to explore the behavior of the ELM classifier (compared to the two reference classifiers) when integrated within a standard classification scheme based on an AR feature reduction. In particular, the number of features was changed from 10 to 50 with a step of 10 so as to test this classifier in small as well as high-dimensional feature subspaces. The third experimental part had for objective to assess the capability of the proposed RVM classification system to boost further the accuracy of the ELM classifier. The fourth experiment was devoted to analyze the generalization capability of the ELM, with and without feature reduction, and of the RVM classification system by decreasing/increasing the number of available training beats. This analysis was done through two experimental scenarios, which consisted in passing from 500 to 250 and 750 training beats, respectively. Finally, the sensitivity of the RVM classification system is analyzed.

4.3. Experimental settings

In this experiment, the ELM classifier is first used to classify the signals. The desired number of features varied from 10 to 50 with a step of 10, namely, from small to high-dimensional feature subspaces. Feature reduction was achieved by the traditional AR modeling, commonly used in ECG signal classification. In particular, it can be seen that for all feature subspace dimensionalities except the lowest (i.e., 10 feature), the RVM classifier maintains a clear superiority over the ELM. Its best accuracy was established using a feature subspace made up of the first 30 components. The corresponding OA and AA accuracies were 95.67% and 95.33%, respectively. From this experiment, three observations can be made: 1) the ELM classifier shows a relatively low sensitivity to the curse of dimensionality as compared to RVM 2) the ELM classifier still preserve its superiority when integrated in a feature reduction-based classification scheme; and 3) RVM performs better results than ELM and provides very high result than ELM when the data are preprocessed with AR modeling technique.

TABLE2
OVERALL (OA), AVERAGE (AA), AND CLASS PERCENTAGE ACCURACIES ACHIEVED ON THE TEST BEATS WITH THE DIFFERENT INVESTIGATED CLASSIFIERS WITH A TOTAL NUMBER OF 500 TRAINING BEATS

Method	OA	AA	N	A	V	RB	/	LB
ELM	88.76	88.48	88.44	87.39	83.48	95.98	84.47	88.76
ELM classification after Preprocessing	89.74	89.78	89.69	88.96	85.18	97.69	86.58	89.74
RVM	92.67	93.33	90.52	93.65	90.97	96.58	92.99	92.67
RVM classification after Preprocessing	95.67	95.33	92.52	94.65	93.97	98.28	94.19	97.16

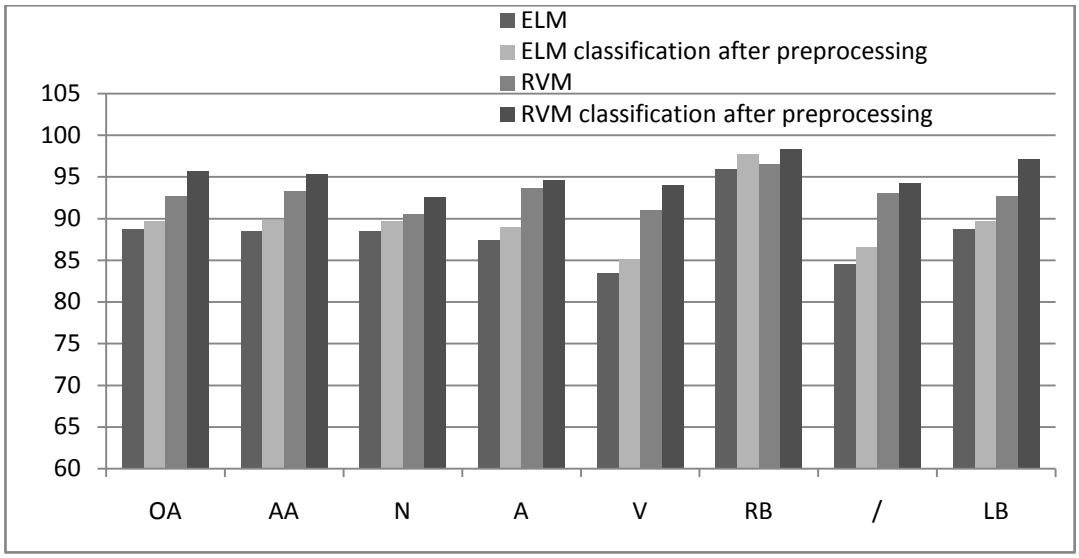

Figure2: Comparison of ELM and RVM accuracy for different datasets

The Figure 2 gives the comparison of the accuracy of classifying the ECG signals by using ELM and RVM. This shows that RVM gives much better accuracy for all datasets given as input. In which RB dataset achieves the maximum accuracy of 98.28%.

TABLE3
NUMBER OF FEATURES DETECTED FOR EACH CLASS WITH THE ELM/RVM CLASSIFICATION SYSTEM TRAINED ON 500 BEATS

Class	N	A	V	RB	/	LB	AVERAGE
#Detected Features	68	49	32	50	47	41	47

Table 3 shows the number of features detected automatically to discriminate each class from the others. The average number of features required by the RVM classifier is 47, while the minimum and maximum numbers of features were obtained for the ventricular premature (V) and normal (N) classes with 32 and 68 features, respectively.

V. CONCLUSION

In this paper, a novel ECG beat classification system using RVM is proposed and applied to MIT/BIH data base. The wavelet transforms variance and AR model parameters have been used for the features selection. From the obtained experimental results, it can be strongly recommended that the

use of the RVM approach for classifying ECG signals on account of their superior generalization capability as compared to traditional classification techniques. This capability generally provides them with higher classification accuracies and a lower sensitivity to the curse of dimensionality. The results confirm that the RVM classification system substantially boosts the generalization capability achievable with the ELM classifier, and its robustness against the problem of limited training beat availability, which may characterize pathologies of rare occurrence. Another advantage of the RVM approach can be found in its high sparseness, which is explained by the fact that the adopted optimization criterion is based on minimizing the number of SVs. It can also be seen that RVM accomplishes better and more balanced classification for individual categories as well in very less training time comparative to ELM. In future some advanced neural network techniques can be used to train the RVM classifier and it may enhance the classification accuracy of the ECG and reduce the training time.

REFERENCES

[1] Osowski, S., Linh, T.H., 2001. ECG beat recognition using fuzzy hybrid neural network. IEEE Trans. Biomed. Eng. 48 (11), 1265–1271.
[2] Hu, Y.H., Palreddy, S., Tompkins, W., 1997. A patient adaptable ECG beat classifier using a mixture of experts approach. IEEE Trans. Biomed. Eng. 44 (9), 891–900.
[3] L. Khadra, A. S. Al-Fahoum, and S. Binajjaj, “A quantitative analysis approach for cardiac arrhythmia classification using higher order spectral techniques,” IEEE Trans. Biomed. Eng., vol. 52, no. 11, pp. 1840–1845, Nov. 2005.
[4] F. de Chazal and R. B. Reilly, “A patient adapting heart beat classifier using ECG morphology and heartbeat interval features,” IEEE Trans. Biomed. Eng., vol. 53, no. 12, pp. 2535–2543, Dec. 2006.
[5] R. V. Andreao, B. Dorizzi, and J. Boudy, “ECG signal analysis through hidden Markov models,” IEEE Trans. Biomed. Eng., vol. 53, no. 8, pp. 1541–1549, Aug. 2006.
[6] S. Mitra, M. Mitra, and B. B. Chaudhuri, “A rough set-based inference engine for ECG classification,” IEEE Trans. Instrum. Meas., vol. 55, no. 6, pp. 2198–2206, Dec. 2006.
[7] C.-K. Siew and G.-B. Huang, “Extreme Learning Machine with Randomly Assigned RBF Kernels,” Int’l J. Information Technology, vol. 11, no. 1, 2005.
[8] Sheng-Wu Xiong, Hong-Bing Liu and Xiao-Xiao Niu, Fuzzy support vector machines based on FCM clustering. Proceedings of 2005 International Conference on Machine Learning and Cybernetics, Aug. 2005, vol. 5, pp. 2608- 2613.
[9] R. Mark and G. Moody MIT-BIH Arrhythmia Database 1997 [Online]. Available http://ecg. mit.edu/dbinfo.html.
[10] Minami, K., Nakajima, H., Toyoshima, T., 1999. Real-time discrimination of ventricular tachyarrhytmia with Fouriertransform neural network. IEEE Trans. Biomed. Eng. 46, 179–185.
[11] Daubechies, I., 1998. Orthonormal bases of compactly wavelets. Commun. Pure Appl. Math. (41), 909–996.
[12] Thakor, N.V., 1993. Multiresolution wavelet analysis of evoked potentials. IEEE Trans. Biomedical. Eng. 40 (11), 1085– 1093.
[13] C.-W.Hsu and C.-J. Lin, “A comparison of methods formulticlass support vector machines,” IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 415–425, Mar. 2002.
[14] Tipping, M.E.: Sparse Bayesian learning and the Relevance Vector Machine. Journal of Machine Learning Research 1 (2001) 211–244
[15] Ke Wang and Haitao Jia, "Image Classification Using No-balance Binary Tree Relevance Vector Machine", Intelligent Interaction and Affective Computing, 2009. ASIA '09, Page(s): 79 – 82
[16] Xin Zhou, Ying Wu and Guopeng Yang, "Modulation Classification of MPSK Signals Based on Relevance Vector Machines", Information Engineering and Computer Science, 2009, Page(s): 1 – 5
[17] http://www.physionet.org/physiotools/ecgpuwave/src/.
[18] J. J. Wei, C. J. Chang, N. K. Shou, and G. J. Jan, “ECG data compression using truncated singular value decomposition,” IEEE Trans. Biomed. Eng., vol. 5, no. 4, pp. 290–299, Dec. 2001.
[19] A. Agresti, Categorical Data Analysis, 2nd ed. New York:Wiley, 2002.
[20] Chuan-Min Zhai and Ji-Xiang Du, "Applying extreme learning machine to plant species identification", Information and Automation, 2008. ICIA 2008, Page(s): 879 – 884

AUTHORS PROFILE

Karpagachelvi.S: She received the BSc degree in physics from Bharathiar University in 1993 and Masters in Computer Applications from Madras University in 1996. She has 13 years of teaching experience. She is currently a PhD student with the Department of Computer Science at Mother Teresa University. She has published four papers in international journals.

Dr.M.Arthanari: He has obtained Doctorate in Mathematics in Madras University in the year 1981. He has 35 years of teaching experience and 25 years of research experience. He has a Patent in Computer Science approved by Govt. of India.

Sivakumar M : He has 10+ years of experience in the software industry including Oracle Corporation. He received his Bachelor degree in Physics and Masters in Computer Applications from the Bharathiar University, India. He holds patent for the invention in embedded technology. He is technically certified by various professional bodies like ITIL, IBM Rational Clearcase Administrator, OCP – Oracle Certified Professional 10g and ISTQB.