ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, April 2013

Patient Adaptable ECG Beat Classifier Using LDC and EMC Algorithms

R.Shantha Selva Kumari[1], J.Ganga Devi[2]

Professor & Head Department of ECE, Mepco Schlenk Engineering College, Sivakasi, India[1]

PG Student, M.E. Communication Systems, Mepco Schlenk Engineering College, Sivakasi, India[2]

Abstract— Electrocardiogram (ECG) waveforms are used to find both regular and irregular patterns in cardiac cycles of patients. In this work, the beats like normal, supraventricular and ventricular beats are identified and classified to the corresponding classes based on combining the decisions of both linear discriminant classifier (LDC) algorithm and Clustering algorithm. Classifications of beats are done by considering both R-R interval features and morphological descriptor features. First the Linear Discriminant Classifier is trained for the three different class and the training beats are taken from three different databases namely, MIT-BIH Arrhythmia, MIT-BIH Supraventricular and MIT-BIH ST Change. Expectation Maximization clustering algorithm (EM) forms patient specific clusters which is based on MOG (mixture of expert) model. Finally the beats in the cluster output is labeled & verified to corresponding classes with the aid of linear discriminant analysis function.

Keywords:*LDC,EM, Feature selection,Wavelet transform, patient adaptable.*

I. INTRODUCTION

In this modern world heart attack death rate increases gradually. With the advanced technology, the heart related problems are detected and cured at earlier stage. To handle thousands of patient's record is a challenging task for any physician. To assist them, a numerous automatic detection algorithms are still developing in the view of patient adaptable. Among the various conditions which cause CVD's, arrhythmia (erratic heartbeat) finds ease way to detect abnormality in the heart function. ECG is the one of the best tool to measure the abnormal beats.

ECG is used to capture transthoracic interpretation of the electrical activity of the heart over time. 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. It is a non persistent recording produced by an electrocardiographic device. The recognition and classification of the ECG beats is a very important task in the coronary intensive unit, where the classification of the ECG beats is essential tool for the diagnosis. 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. Up to now, many algorithms have been developed for the recognition and classification of ECG signal. Some of them use time and some others 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. 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 (Hu, Y.H., Palreddy, S., Tompkins, W., 1997).

Many algorithms for ECG heartbeats classification were developed in the last decades (see references in [1], [2]), but due to the lack of standardization in the development and evaluation criteria, comparison of results across most of these works could not be performed fairly. In order to overcome this problem, some methodological aspects in the development and evaluation of heartbeat classifiers are followed in recent works. It is found that the automatic classification approach has approximated to a performance upper bound. For that the patient adaptation technique was found to be useful. The Association for the Advancement of

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, April 2013

Medical Instrumentation (AAMI) recommends [4] for class labeling and results presentation has eased this problem, and it is broadly accepted. Regarding to the classes of interest, the AAMI recommendation suggests five classes: supraventricular (S), ventricular (V), fusion (F), beats that cannot be classified (Q), and normal (N). In some papers the different classification approaches classify beats without any local expert (LE) assistance [1, [2], and [3], but others take advantage from a LE to improve the classification performance.

II. AIM

The main aim of this work is to correctly detect the R-peaks and QRS points in the ECG waveforms since those are going to help the classification of beats and successively form the clusters with respect to patient specific. The organization of this paper is as follows, Section III explains the materials (i.e.) the description of the dataset used for training and testing. Section IV gives the detailed description about the methodology. Section V provides the results of LDC and EMC algorithms. The overall conclusion of this study is summarized in Section VI.

III. MATERIALS

All the experiments in this work are carried out by considering three different public databases namely MIT-BIH Arrhythmia, MIT-BIH Supraventricular and MIT-BIH ST change which is available freely on Physionet [5]. For all databases the AAMI recommendations for class-labeling are adopted. The database includes different types of ECG recordings. In this work, data with a broad range of normal and pathological ECG recordings are considered to evaluate the algorithm performance.

A. MIT-BIH Arrhythmia Database (MITBIH-AR)

The MIT-BIH Arrhythmia Database contains 48 records which are recorded for 30 minutes studied by the BIH Arrhythmia Laboratory between 1975 and 1979 [6]. Approximately 60% of these recordings are obtained from inpatients. In most records, one signal is a modified limb lead II (MLII), obtained by placing the electrodes on the chest. While the other is usually a modified lead V1. Normal QRS complexes are usually prominent in the first signal. The lead axis for the second signal may be nearly orthogonal to the mean cardiac electrical axis. The recordings are digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. Some authors use annotations provided with the database for training and testing purposes, following the recommendations and class-labeling of AAMI.

B. MIT-BIH Supraventricular Arrhythmia Database

The database consists of 78 two-lead records of approximately 30 minutes and sampled at 128 Hz. The records are chosen to supplement the examples of supraventricular arrhythmias in the MIT-BIH Arrhythmia Database [7]. All records in these databases are first resampled to 360Hz, which is the sampling frequency of the MIT-BIH-Arrhythmia database. The original labeling is adapted to the AAMI recommendations and to the AAMI2 modification. This database will be considered for validation purposes.

C. The MIT-BIH ST Change Database

This database includes 28 ECG recordings of varying lengths, most of which are recorded during exercise stress tests and which exhibit transient ST depression. From this 18 useful recordings are used for both training and evaluation purposes. The recordings are sampled at 360 Hz and 12-bit resolution.

IV. METHODOLOGY

The Proposed method aims to classify the beats into normal, supraventricular and ventricular classes for a patient record [8], [9] and [10]. In this work, annotations provided in the database are not used. Instead Pan-Tompkins's algorithm is used to find the peaks of every beats. The next step is to train the LDC system by deriving ECG feature vectors from the signal sample beats from three different databases. Final step is to form the clusters of different classes for any patient ECG signal using Expectation Maximization Clustering algorithm.

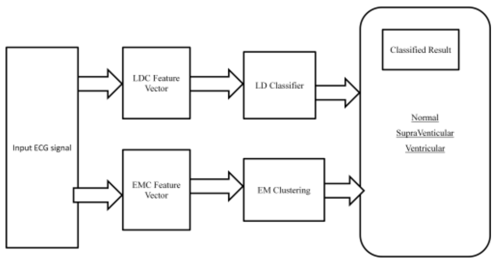

Fig 1. The Proposed Block Diagram

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, April 2013

LDC and EMC work independently and each performs a preliminary classification/clustering task in their feature spaces. Both LDC and EMC's are discussed in the forthcoming sections individually. Finally, the heartbeat labels provided in the clusters are verified by the LDC.As given in figure 1, from the input ECG signal is observed as beat by beat and features like RR interval and QRS positions are extracted. These feature vectors are given as input to both classifiers. The proposed system consists of three stages namely, a preprocessing stage, feature extraction stage and a clustering/classification stage. Let us see the explanations of individual stage as follows.

A. Preprocessing Stage

In this work, only modified lead II signal of MIT-BIH Arrthymia Database is used [11], [12]. The digitized ECG signal is applied as input to the preprocessing stage. These signals consist of baseline wander, power line interference, and high-frequency noises. The preprocessing stage utilizes a filtering unit to remove artifact signals from the ECG signals. The following steps explain the function of preprocessing stage

Step 1: First read the annotation file, header file and data file for a record for a duration of 30min and 5 .556 Seconds.

Step 2: Apply smoothening function to that signal to remove the noise interference.

Step 3: To remove the base line interference, the signal obtained from the step2 is subtracted from original signal.

Step 4: Determine Wavelet coefficients at scale 5 and using that denoise the output of step 3 signals.

Step 5: Finally a Clean ECG signal is obtained and it is considered as input to the next stage.

B. Feature Extraction Stage

The performance of ECG pattern classification strongly depends on the characterization power of the features extracted from the ECG data. In this work, features related to fiducial point intervals and ECG morphology are calculated separately [12], [13].

In every cardiac cycle of heart, R peak occurred due to the contraction of the ventricles. The main reason for considering the R peak for my work is that it has higher amplitude value when compared to other points in the ECG signal. Therefore it can be easily identifiable. The normal duration of between two R peaks (i.e) RR interval varies from 0.6 to 1.2 seconds. Abnormal beats shows variation in their range. That abnormality is due to delay in the contraction of ventricles which leads to a possibility of getting more than one R peak within one cycle.

By considering R-R interval feature alone is not enough to classify the classes of heart beats.QRS morphology descriptor features are also needed to specify the specific abnormality. The normal duration QRS duration ranges from 80 to 120ms as referred in the ECG basics. If that width is less means (i.e.) narrow it comes under supraventricular beats. Else wider means it comes under ventricular beats. Since those abnormalities occur due to different conditions in the heart. They also need different kind of treatment for recovery. Due to its time–frequency localization properties, the wavelet transform is an efficient tool for analyzing non stationary ECG signals. The wavelet transform is used to extract morphological information from the ECG data. Wavelet transformation maps the ECG signal into a time-scale plane. The fourth scale has the good projection between the frequency ranges 12.25 to 22.5Hz. It is worth to use the DWT which can be efficiently implemented as a filter bank. The DWT function of the signal is given as

$$f(t) = \sum_{a} \sum_{b} C_{ab} \, \Psi_{ab}(t) \qquad \text{------------ (1)}$$

Here Ψ is the mother wavelet, *a* is the dilation parameter and *b* is the offset parameter. To find QRS points Pan-Tompkin's algorithm is used. The following steps explain how to obtain the QRS points in beat:

Step 1: Apply DWT at scale 4 to clean ECG signal which is the output of preprocessing stage.

Step2: Reconstruct the signal only using approximate coefficients.

Step 3: As per Pan-Tompkin's algorithm apply band pass filter and moving window integrator. The threshold value is obtained by subtracting the mean and maximum value of the previous step output signal.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, April 2013

Step 4: Compare the peaks of the beats which are greater than the threshold are considered as R peaks.

Step 5: For some records there is a need to change the threshold for finding the missing beats.

Step 6: After finding R peak , set the window 50ms before R peak and 150ms after R peak.

Step 7: To the left of R peak ,within that window find the minimum point that is considered as Q point.

Step 8: To the Right of R peak,within that window find the mininum point that is considered as S point.

Step 9: Using that points find the QRS duration.

C. DISCRIMINANT ANALYSIS

Discriminant analysis is a statistical technique. LDA seeks to reduce dimensionality while preserving as much of the class discriminatory information as possible. Under the assumption of normally distributed data, the MAP (maximum a posteriori) criterion leads to quadratic discriminant functions, for classification purposes .The values for the prior probabilities $P(\omega_i)$ are considered the same for all classes [14]. In the case that the covariance matrix **Σ** is considered to be the same for all classes the Quadratic Discriminant Classifier (QDC) becomes linear in **x** leading to the Linear Discriminant Classifier (LDC).The Linear classifier defined by the discriminant functions is given by

$$g_i(x) = \mu_i^T \Sigma^{-1} x - 0.5\mu_i^T \Sigma^{-1} \mu_i + \log(P(\omega_i)) \quad \text{----- (2)}$$

Where,
i represents the specific class,
x represents feature vector describing each heart beat
$P(\omega_i)$ represents prior probability
Mean vector μ_i is given by

$$\mu_i = 1/M_i \sum_{m=1}^{M_i} x_m \quad \text{---------------- (3)}$$

Weighted covariance matrix expression is given as

$$\Sigma = \frac{1/(\sum_{i=1}^{C} w_i)\sum_{i=1}^{C} w_i \sum_{m=1}^{M_i}(x_m - \mu_i)(x_m - \mu_i)^T}{M_i} \quad \text{----- (4)}$$

The following table illustrates the features considered for discriminant analysis,

TABLE 1.FEATURES FOR LDA

Sl.no	Features	Description
1.	RR(i)	Current RR interval
2.	RR(i+1)	Next RR interval
3.	RR_{20}	Average RR interval for last 20 minutes
4.	Mean QRS	Average of QRS width for 10 minutes

Here in this work three classes are considered for classification of ECG signal beats, they are i) normal, ii) Supraventricular and iii) ventricular. Train these classes with beats from different records as mentioned in section III. Using equation (3) and (4) find the mean and covariance matrix. The value for the prior probabilities $P(\omega i)$ are considered the same for all classes. Using equation (2) find the class of unknown beats, M_i is the number of examples $\mathbf{x}_m$ of the i^{th} class, the rule assigns an unlabeled observation **x** to the class *i* that results in the maximum posterior probability $g_i(\mathbf{x})$.

D. EMC algorithm

As mentioned earlier to handle huge inter patient variability, automatic classification method fails. So patient specific algorithms like expectation maximization clustering algorithm is needed. This algorithm assumes a priori that there are *'K'* Gaussian and then the algorithm tries to fit the data into the *'K'* Gaussian by expecting the classes of all data point and then maximizing the maximum likelihood of Gaussian centers. The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the *E* step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step. It gives extremely useful result for the real world data set. But it is highly complex in nature. It consists of estimating the parameters of a density function and is given as

$$p(x|\psi) = \sum_{k=1}^{K} \pi_k \; \mathcal{N}(x|\mu_k, \Sigma_k) \quad \text{----------- (5)}$$

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, April 2013

Where, K represents number of clusters, the *m*-dimensional vector **x** is modeled by *K* Gaussians with mixing coefficients π_k , in order to retain a more realistic structure of the data. The parameter set $\Psi = \{\pi_k , \boldsymbol{\mu}_k , \boldsymbol{\Sigma}_k \mid k = 1, \ldots , K\}$ is estimated by maximum likelihood criterion. We maximize the log likelihood

$$L(X|\psi) = \ln \prod_{n=1}^{N} p(x_n|\psi) \qquad \text{------------ (6)}$$

for the *N* heartbeats in each recording named $X = \{\mathbf{x}1, \ldots, \mathbf{x}_N\}$.Since there is not a closed form solution for Ψ by maximizing $L(X|\Psi)$, the well-known expectation maximization (EM) algorithm is used to obtain the estimation equations of the parameters Ψ at each iteration *j*, which are the mixing coefficient for each cluster and is given by

$$\pi_k^j = 1/\mathrm{N} \sum_{m=1}^{N} \beta_{m,k}^{j-1} \qquad \text{--------------- (7)}$$

The cluster mean is given a

$$\mu_k^j = 1/(\mathrm{N}\pi_k^j) \sum_{m=1}^{N} \beta_k^{j-1} x_m \qquad \text{-------------- (8)}$$

The cluster covariance matrix is given as

$$\Sigma_k^j = /(\mathrm{N}\pi_k^j)\sum_{m=1}^{N} \beta_{m,k}^{j-1} (x_m - \mu_k^j)(x_m - \mu_k^j)^{\mathrm{T}} \quad \text{-- (9)}$$

Where $\beta^j_{m,k}$ is known as the ownership variable, which indicates the probability of sample $\mathbf{x}_m$ to have been generated by the k^{th} component at iteration j and it is given as $\pi_k^{j-1} \mathcal{N}(x_m | \mu_k^{j-i}, \Sigma_k^{j-1})$

$$\beta_{m,k}^j = \frac{\qquad}{\sum_{n=1}^{K} \pi_n^{j-1} \mathcal{N}(x_m|\mu_n^{j-1}, \Sigma_n^{j-1})} \qquad \text{------ (10)}$$

The EM algorithm iteratively computes the weight, location, and dispersion for each of the *K* clusters, until $\beta^j_{m,k}$ does not change significantly, which is equivalent to obtain stable clusters. The prematurity of heart beat measures how anticipated is a heartbeat respect to the previous and next RR interval and is given by

$$P_{RR}[\mathrm{i}] = \frac{\mathrm{RR}[\mathrm{i}]}{\sum_{k=i-1}^{i+1} RR[k]} \qquad \text{--- -- (11)}$$

The local RR interval variation is defined as

$$\mathrm{dRR_L[i]} = \sum_{k=i-1}^{i+1} |dRR[k]| \qquad \text{----------- (12)}$$

Where dRR[i] = RR[i]-RR [i-1].
The following table gives the features for EMC

TABLE 2: FEATURES FOR EMC

Sl.no	Feature	Description
1.	RR(i)	Current RR interval
2.	RR(i+1)	Next RR interval
3.	RR(i-1)	Previous RR interval
4.	RR_{20}	Average RR interval for last 20 minutes
5.	Mean QRS	Average of QRS width for 10 minutes

V. RESULTS AND DISCUSSION

As stated earlier in the metholody section, the input signal is preprocessed since it is distorted by external noise signals and also by the measuring equipments. First remove the baseline drift by smoothening the input signal and that output is subtracted from the original signal. This base line eliminated signal is denoised to obtain the clean ECG signal. For denoising determine wavelet coefficients at scale 5.

Input signal Baseline Wander Removed

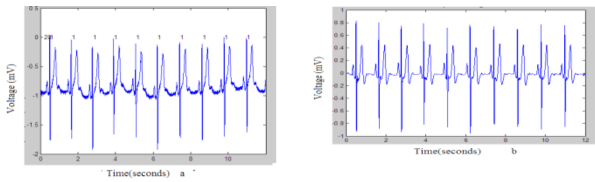

Fig 2: Preprocessing stage output signal

For any algorithm the feature extraction is the predominant one. The best and correct feature leads to good classification performance. According to that, RR interval and QRS morphology features are considered for this work. The following figure shows the R position and QRS points of record 117.

R Peak Detected QRS point Detected

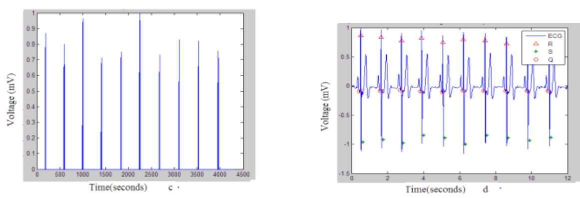

Fig 3: Feature Extraction stage output

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, April 2013

In case of LDA, First the system is trained with the corresponding classes feature vectors. In this work, the ultimate aim is to classify normal, supraventricular and ventricular classes. So that the records are selected from three different databases namely MIT-BIH Arrhythmia, MIT-BIH Supraventricular and MIT-BIH ST change databases with respect to its classes.. For training the normal class, the beats from the following records 100, 209, 121, 117, 103, 113, 118, 302, and 310 are used. For training supraventricular class, the beats from the following records 802, 803, 804, 824, 844, 326, 841, 806, 826, 809, and 890 are used. For training ventricular class, the beats of following records 102, 207, 202, 123, 233, 208, 325, 109, 203, and 213 are used. As mentioned in table 1 those features are considered for discriminant analysis algorithm. For above mentioned beats of the records determine the RR interval and QRS morphology features and train the discriminate function to their specific class using the equation mentioned in the section C. After training the system, the remaining beats are taken for testing. Consider the following table which shows the output classification of particular beats of different records. For a feature vector of a particular beat, it is classified into corresponding class whose discriminant function yields maximum value than other class discriminant functions.

The specification of various discriminant functions as follows,

G1(x) corresponds to Normal class discriminant function.

G2(x) corresponds to Supraventricular class discriminant function.

G3(x) corresponds to Ventricular class discriminant function.

The main objective of the EMC algorithm to retain patient specific information. As Specified in table 2 features are considered for EMC algorithm. Based on the feature vector the beats are clustered. Here in this work, three clusters are formed which contains the beats of similar category. The figure depicts the three different output clusters belongs to Normal, Supraventricular and Ventricular beats of record 203 based on their RR interval and QRS width duration.

Table 3: Classification output for LDC

Particular Beat from Specific Records	G1(x)	G2(x)	G3(x)	Classification output
117	20.3329	20.0409	9.3380	Normal
872	17.5811	17.8568	5.9016	Supraventricular
812	41.278	36.0485	42.4363	Ventricular

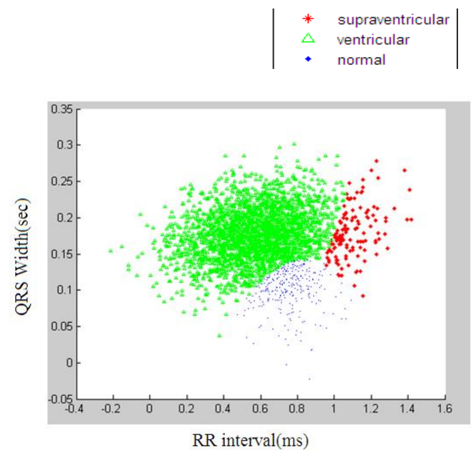

Fig 4. EM Cluster Output-Record 203

Moreover 90% of beats are correctly grouped in the same cluster. Likewise, any patient’s signal beats are given to this clustering algorithm and hence they are classified in their corresponding clusters.

VI. CONCULSION

As far as concerned, in these work two different algorithms namely LDA and EMC are used to identify and classify the different class of beats by considering the RR interval and morphological feature vectors. Linear discriminate analysis classifier is trained individually for three different classes namely Normal, Supraventricular and Ventricular classes. Regarding Expectation Maximization Clustering algorithm, the patient specific information is retrieved (i.e.), for individual patient’s beats are classified according to the classes by forming the clusters. Finally LDC verifies the beat labels present in each clusters. Hence the beats are confirmed with respect to their classes. In future work the beats of the corresponding classes are evaluated based on the ground truth databases.

REFERENCES

[1]. Y. H. Hu, S. Palreddy, and W. Tompkins, " A patient-adaptable ECG beat classifier using mixture of experts approach," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 891–899, Sep. 1997.

[2]. P. de Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

[3]. P. de Chazal and R. B. Reilly, "A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features," *IEEE Trans.Biomed. Eng.*, vol. 53, no. 12, pp. 2535–2543, Dec. 2006.

[4]. Testing and Reporting Performance Results of Cardiac Rhythm and ST-Segment Measurement Algorithms, American National Standard,ANSI/AAMI/ISO EC57, 1998–(R)2008.

[5]. A. L. Goldberger, L. A. N. Amaral, L.Glass, J.M.Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley,"PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[6]. R. Mark and G. Moody. (1997)."Mit-bih arrhythmia database," [Online].Available: http://ecg.mit.edu/dbinfo.html.

[7]. R. Mark, G. Moody, and S. Greenwald. (1990). "Mit-bih Supraventricular arrhythmia database," [Online]. Available: http://www.physionet.org/physiobank/database/svdb.

[8]. M. Llamedo and J. P. Martınez, "An Automatic Patient –Adaptable ECG Heartbeat Classifier Allowing Expert Assistance," *IEEE Transactions on. Biomedical Engg.* vol. 59, No.8 August 2012.

[9]. M. Llamedo and J. Mart´ınez, "Heartbeat classification using feature selection driven by database generalization criteria," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 616–625, Mar. 2011.

[10]. M. Llamedo and J. P. Mart´ınez, "Cross-database evaluation of a multilead heartbeat classifier," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, in press, 2012.

[11]. R. Shantha Selva Kumari and V.Sadasivam "De-Noising and Baseline Wandering Removal of Electrocardiogram using Double Density Discrete Wavelet", International Journal of Wavelets, Multiresolution & Information Processing,Vol. 5 no.1, Jan 2007.

[12]. R. Shantha Selva Kumari and V.Sadasivam,"Wavelet-based base line wandering removal & R peak & QRS complex detection", International Journal of Wavelets, Multiresolution & Information Processing,Vol. 5 no.6, November 2007.

[13]. R. Shantha Selva Kumari , S.Bharathi and V.Sadasivam, "QRS Complex Detection Using Optimal Discrete Wavelet ", International Journal of Computational Intelligence & applications ,vol. 8, issue 2, June 2009.

[14]. V. Chud´acek, G. Georgoulas, L. Lhotsk´a, C. Stylios, M. Petr´ık, and M. Cepek, "Examining cross-database global training to evaluate five different methods for ventricular beat classification," *Physiol. Meas.*,vol. 30, no. 7, p61, 2009. [Online]. Available:http://stacks.iop.org/0967-334/30/i=7/a=010