Heart Disease Prediction Using Integer-Coded Genetic Algorithm (ICGA) Based Particle Clonal Neural Network (ICGA-PCNN)

S. Silvia Priscila and Dr.M. Hemalatha

Abstract--- *Nowadays, cardiovascular diseases are known as the one of the most dangerous and common problems in modern society. Analyzing and classifying the ECG signal will yield an accurate detection of different arrhythmias. Hence this paper proposes an Integer-Coded Genetic Algorithm (ICGA) based Particle Clonal Neural Network (ICGA-PCNN) for classifying the different ECG arrhythmias. Initially the histogram features and morphological features are extracted from the Pan-Tompkins based QRS complex. After that the optimal set of features has been selected using the ICGA for the extracted features. Then Multilayer feed forward neural network (MFNN) is used as a classifier to classify the ECG signal, where the weight and the biased are trained using the Particle based clonal selection. The MIT-BIH arrhythmias ECG Database has been presented as a database to train and test the proposed ICGA-PCNN classifier. The experimental results show that the proposed classifier approach performs better than the existing classification approaches in terms of classification accuracy, sensitivity and specificity.*

Keywords--- *ECG, Cardiovascular Diseases, Pan-Tompkins, ICGA-PCNN, Integer-Coded Genetic Algorithm, MFNN, Particle Swarm Optimization, Clonal Selection.*

I. INTRODUCTION

DATA MINING is the process of extracting meaningful information from various resources which is used to analyze the decision about particular data. The mining process used in several applications such as prediction, emotion classification, hand gesture recognition, heart disease classification, iris recognition and cancer identification. From the above applications heart diseases are the most important, because heart diseases foremost cause of death. Heart disease factors can be split in two ways such as modifiable and non modifiable. Modifiable risk factors include obesity, smoking, lack of physical activity and so on. The non modifiable risk factors for heart disease are like age, gender, and family history [1].

Some kind of heart diseases are irregular heartbeat (arrhythmias), congenital heart defects, weak heart, muscles (cardiomyopathy), heart valve problems, heart infections and cardiovascular disease. Cardiovascular disease is the important cause of death. Cardiovascular disease raised blood pressure, peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure problems [2]. In the cardiovascular diseases may display various syndromes. By reason of this complexity, there is a need to detect diseases using diagnostic process. In the cardiovascular diseases may display various syndromes [3]. Hence, early detection is considerable to improve the diagnosis of cardiovascular diseases. As the symptoms are unsteady it is very challenging to the physicians or radiologists to identify abnormalities if they diagnose only by their experience. In EU country due to cardiovascular disease dearth percent in roughly is 45%, and pay hundred billion Euros for cardiovascular diseases [4]. Several cardiovascular disease cannot be find accurately by chest X-rays. The Cardiac abnormalities can be identified fast and efficiently using ECG signals.

Thus, it is important to develop a fully automatic technique with high sensitivity to assist the early detection of cardiovascular disease with ECG. In our proposed based Particle Clonal Neural Network (ICGA-PCNN) for classifying the different ECG arrhythmias such as Normal, LBBB, APB, VE and VF to aid in diagnosis. The rest of the paper is organized as follows: Section 2 describes the recent related work of automated diagnosis of ECG signals. Section 3 describes the methods and materials. The experimental results are discussed in the section 4. Finally, section 5 renders the conclusion.

II. RELATED WORK

Lovepreet Kaur [5] proposed prototype Intelligent Heart Disease Prediction System with Fuzzy C Means Clustering algorithm. An intelligent Heart Disease Prediction System built with the help of miming technique like decision trees, naive bayes and neural network. The proposed system using data acquisition, pre-processing, feature extraction, and classification methods which is used to analysis the heart diseases in terms of accuracy, time, specificity and sensitivity, and to make intelligent medical decisions from the traditional decision support system. This paper have58 records and 14 attributes are used to predict the heart diseases. The prediction system displays 86.6% accuracy, 32 milliseconds time, 0.44 specificity and 0.45 sensitivity.

Chaitrali S. Dangare et al [6]., proposed mining techniques such as decision trees, naive bayes and neural networks, in which to classify the heart diseases from the historical heart database based on Quality of Service (QoS). Based on the Quality of Service provides correct and effective treatment. In this paper takes two input attributes from 13 input attributes

S. Silvia Priscila, Research Scholar, Bharathiar University, Coimbatore, India. E-mail:sylviaprem2010@gmail.com

Dr.M. Hemalatha, Research Supervisor, Bharathiar University, Coimbatore, India. E-mail:hema.bioinf@gmail.com

DOI:10.9756/BIJIEMS.8394

such as obesity and smoking, which provides the more accuracy result as compared to other attributes.

Manjusha B et al [7]., proposed classification of diseases in multilayer perception with the help of artificial neural network (ANN) with back propagation algorithm. Multilayer perception contains 15 attributes and one output. Classification of 15 attribute is not possible so in this paper take two attribute like smoke and family history. These smoke and family history are classified via Framework for the classification of heart disease dataset and GUI for the classification of heart disease dataset technique. Using these techniques it provides advantages such as it automatically allow arbitrary nonlinear relation between both the independent and dependent variable and increasing the classification process efficiency.

Salha M et al[8]., proposed system to analysis the coronary arteries diseases (CAD), cardiovascular disease (CVD) and coronary heart diseases (CHD)by using some data mining techniques such as decision trees, C4.5 classification algorithm, RIPPER classification algorithm, Naive Bayes classifiers, K-nearest neighbor classification (KNN), support vector machine (SVM), and artificial neural networks techniques. The cardiovascular disease (CVD) provides more than 95% accuracy with the help of decision trees and Naive Bayes classifiers. Further C5, support vector machine (SVM) and neural network used for coronary heart diseases (CHD) prediction. Thus, these methods are used to predict the diseases in early stages with more accuracy.

Srinivas, K. et al [9] proposed to predict the heart diseases in terms of cost, because several analysis systems are used to identify the cardiovascular disease (CVD). Sometimes prediction of cardiovascular disease (CVD) does not provide accurate decision. In that scenario, the data mining techniques such as Decision Trees, Naive Bayes and Neural Network are used to improve the medical care and reduce the cost.

Yang Wu, et, al [10]., proposed statistical approach, which improves the ECG classification accuracy, and also segment the heart beats from the ECG signal by temporal and independent component analysis features. A minimal-redundancy-maximal-relevance based feature selection has been introduced to obtain the most significant features for diagnosis. At last a voting strategy has been designed for decision taking from various classifiers.

III. Methodology and Algorithm

Proposed Integer-Coded Genetic Algorithm based Particle Clonal Neural Network (ICGA-PCNN) analysis and classifies the different types of arrhythmias via ECG and the block diagram for the proposed approach has been shown in figure 1. The input ECG signals are preprocessed by Infinity Impulse Response (IIR) filter which is used to remove the noise like a base line wandering. The histogram and morphological features are extracted from the Pan-Tompkins based QRS complex.

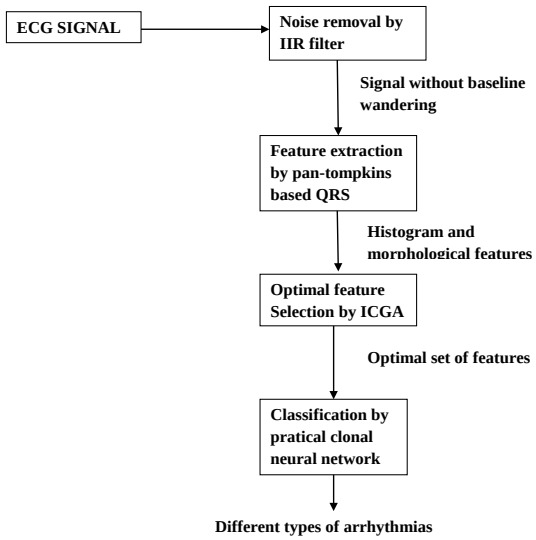

Figure 1: Proposed ICGA-PCNN Approach for Discovering Heart Disease from ECG

After that, using the Integer coded genetic algorithm the optimal features set will be extracted from the complex QRS wave in the de-noised signal. Finally, the selected features will be fed as an input to the Multilayer feed forward neural network classifier to effectively classify the different types of arrhythmias Normal (N), Left Bundle Branch Block (LBBB), Arterial Premature beat (APB), Ventricular Escape beat (VE) and Ventricular Flutter (VF). The clonal selection algorithm is used to train the classifier in order to improve the classification accuracy.

A. *Preprocessing*

IIR filter is the most effective approach, in which to remove the baseline wandering from the ECG signal in order to reduce the misreport data. The IIR filter is used to remove the noise in ECG using some windows such as Butterworth filter, Chebyshev filter, and Inverse Chebyshev filter. The Butterworth filter produces better performance between attenuation and phase response as compared to other IIR filter [11]. The Low pass butter filter is defined as mathematically is given by,

$$H(j\omega) = \frac{1}{1+(\frac{\omega}{\omega_c})^{2N}} \quad (1)$$

Filter selectivity,

$$F_s = N/2\sqrt{2}\omega_c \quad (2)$$

Attenuation,

$$A = 10\ log\ (1 + (\ 1 + (\frac{\omega}{\omega_c})^{2N}) \quad (3)$$

Where ω_c is the cutoff frequency, N is the filter order

B. *Feature Extraction*

The Feature extraction is the significant task in ECG classification to describe the features of signal. This paper uses Pan-Tompkins based approach to detect the R peak signal in ECG and also use step detection accelerometer signal. This signal contains low-pass/high-pass filter, derivative, squaring, and integration for pre-process and adaptive threshold for peak detection.

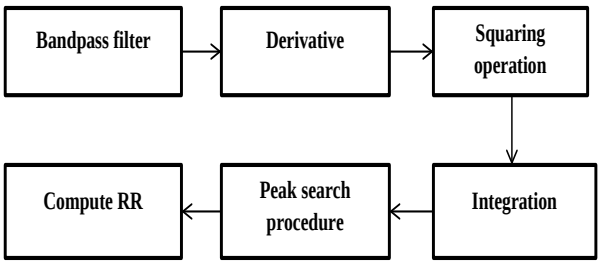

Figure 2: Block Diagram of Pan Tompkins

Bandpass Filter

Band pass filter is a combination of low and high pass filter, which is used to reduce the effect artifacts in the ECG signal. The high pass filter frequency is calculated by,

$$f_c = \frac{1}{2\pi r} = \frac{1}{2\pi RC"} \quad (4)$$

And the low pass filter frequency is calculated by

$$f_c = \frac{1}{2\pi r} = \frac{1}{2\pi RC} \quad (5)$$

Derivative

The derivative operator is defined by,

$$y(n) = 1/8[2x(n) + x(n-1) - x(n-3) - 2x(n \quad (-4)] \quad (6)$$

The derivation operator is formed by low-frequency and high-frequency components

Squaring

The squaring operation conducts the positive result and improves the high frequency components

Integration

The preceding output produces multiple peaks and hence needs to be smoothed by using the equation below

$$y(n) = 1/N[x(n-(N-1)) + \cdots x(n(N-2)) + \cdots + x(n)] \quad (7)$$

Peak search procedure

Peak search procedure is used to detect the R peak. To find the peak wave, the following equation has been used

$$Slope\ (n) = -2x(n-2) - x(n-1) + x(n+1) + 2x(n+2) \quad (8)$$

The scope-threshold value is calculated by

$$Slope\_threshold = \frac{threshhold\_param}{16} * \max i \quad (9)$$

To detect the one set QRS complex by the condition,

$$slope\ (n) > slope\_threshold \quad (10)$$

After finding the onset of the QRS complex the next is to find the RR interval by using the formula

$$RR\ (i) = R_x\ (i+1) - R_x\ (i) \quad (11)$$

Rx (i) denotes the index of i^{th} peak and Rx (i + 1) denotes the index of $(i + 1)^{th}$ R peak. Minimum and maximum value of R-R interval is used to decide the Q and S points. Histogram feature parameters are Average RR signal Morphology (ARSM) and RR interval frequency histogram (RIFH).

Average RR Signal Morphology (ARSM)

The each RR is divided into 300 equal parts and each part is signaled and the amplitude is calculated by the spline interpolation method. After that, the average of RSM is used to extract from the ECG signals and produce the ARSM.

RR Interval Frequency Histogram (RIFH)

The RR interval frequency histogram has to be extracted by following procedure. Initially total range of frequency histogram has selected the time from zero to three seconds. At the time the whole range is divided into 30 equal ranges, which is used to calculate the RR interval. These RR different time interval frequencies produce the RIFH parameter.

C. Feature Selection Using the Integer Coded Genetic Algorithm

Integer Coded Genetic Algorithm is used to select the best feature for classifying different types of arrhythmias. Selecting the optimal feature using the Integer Coded Genetic Algorithm processes like selection, mutation and crossover function. The human body consists of chromosome combination of NG genes corresponding to NG units. It is calculated by using the formula,

$$\sum_{c=1}^{3} |T_i^c| = 24 \quad i \in \{Gen.units\} \quad (12)$$

Here 24 is the hour of a day and 3 is the number of NG genes. After finding the NG unit to calculate the fitness by

$$SU_{it} = \sum_{i=1}^{N_G} \sum_{c=2}^{3} V(T_i^c).SUi(-T_i^{c-1}) \quad (13)$$

$$SD_{it} = [1 - V(T_i^c)].SD_i \quad (14)$$

The selection process is the first process of optimization technique. In the selection method is used to select the N features from the M features by

$$\sum_{i=1}^{N_G} p_{i,max} * I_{it} \geq P_{D,t} + P_{L,t} \text{ for maximum} \quad (15)$$

$$\sum_{i=1}^{N_G} p_{i,min} * I_{it} \leq P_{D,t} + P_{L,t} \text{ for minimum} \quad (16)$$

After selection process the result is applied to cross over technique. The crossover is the second step of the optimal selection method. The general crossover is to select one or more crossing point in the chromosomes. The crossover technique has various types such as, Single point crossover, Two point crossover, Uniform crossover, Arithmetic crossover. Mostly single point cross over are using in feature selection, because single point cross over is avoiding the replication during the optimal selection process. Third process is mutation, which is used to change the selected features based on the crossover function.

IV. CLASSIFICATION

The extracted features are fed into the multi-layer feed forward neural network (MFNN). MFNN is a type of feed-forward neural network. The network consists of one input layer, one or more than one hidden layers and one output layer. Then Multilayer feed forward neural network (MFNN) is used to classify the ECG signal, where the weight and the bias are trained using the Particle clonal neural network selection. Figure 3 shows the Representation of Antibody in MFNN.

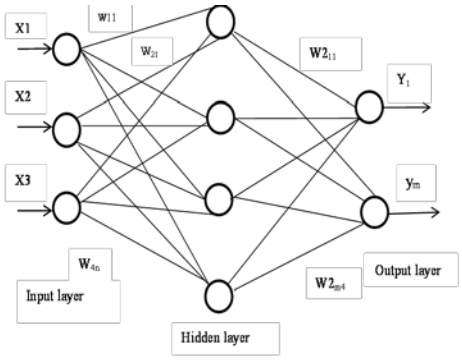

Figure 3: Representation of Antibody in MFNN

An Antibody (Ab) is a vector represents the neural network. Figure 2 shows the antibody representation in MLP. The length of the vector can be given as follows,

$$l = (n+2)h + 1 \quad (17)$$

where n is the number of nodes in input layer, h is the number of nodes in hidden layer. Let i^{th} node input value be x_j, j^{th} hidden node value be y_j , o be the output, w_{ij} is the weight of the linked the I and j, v_{ij} is the weight linked the j and o, j^{th} hidden node bias be b_j and b_o be the output node bias. The biases and the weights are coded into antibodies

$$w_{ij} = Ab_{(i-1)h+j} \quad (18)$$
$$b_j = Ab_{nh+j} \quad (19)$$
$$v_j = Ab_{(n+1)p+j} \quad (20)$$
$$b_o = Ab_i \quad (21)$$

Training Using Particle Clonal Neural Network Selection

Each antibody represents a class, which has ANN encoded data and for these antibody an affinity value is calculated with respect to current antigen. The extracted generic feature vectors is given as input to the MLP, which is designed using the antibody information to obtain the output as well as error value. Calculating the average error and the false positive rate of the Ab with respect to antigen, which is the cluster of similar characteristics. The affinity value is based on the error vale ev and the false positive rate fpr.

$$av = \frac{2}{(\alpha.ev + \beta.fpr)} \quad (22)$$

Where α and β are significant value and its value be 1.

In total set, select the x highest affinity Ab and clone them independently. In clonal set, all antibodies will be performs affinity maturation.

Particle clonal neural network selection is to avoid the local optimal solution and to ensure the global optimal solution; mutation has the major role in immune-based particle swarm optimization, which is used to improve the mutation mechanism. The mutation provides best solution from the N number solution. PSO is a high performance optimizer that is easy to use in clonal neural network selection. The velocity of each particle is updated by,

$$\vec{v_i}(k+1) = \omega * \vec{v_i}(k) + \phi_1 * rand() * \left(Lbest - \vec{x_i}(k)\right) + \phi_2 * rand() * \left(Gbest - \vec{x_i}(k)\right) \quad (23)$$

where ω is the coefficient of inertia, φ1 is the cognitive study, and φ2 is the group study. The rand () is uniformly distributed random numbers like [0, 1]. The term $\vec{v_i}$ is limited to the range max ± $\vec{v_i}$ max. Changing the velocity provides the best position in global best position the each particle position change is calculated by,

$$\vec{x_i}(k+1) = \vec{x_i}(k) + \vec{v_i}(k+1) \quad (24)$$

Updating of each particle value is necessary, which is used to calculate further particle value, because the new particle is higher than the local particle value so the local best particle is replaced by new particle. Moreover, in the mutation step is important because each antibody population is mutated only one time by PSO in each generation.

V. Experimental Results

MIT-BIH database was utilized as information of the ECG signal for examining the performance of the proposed **ICGA-PCNN** for arrhythmia classification. In the examination, the training data set and the testing data set include 20 records. The performance of the approach is measured in terms of accuracy, sensitivity and specificity.

A. Performance Metrics

Sensitivity (true positive rate) is proportion of actual positives that are discovered correctly

$$\text{Sensitivity} = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (25)$$

where True positive is truly identified and False Negative is rejected falsely. Specificity (true negative rate) measures the proportion of actual negatives that are discovered correctly

$$Specificity = \frac{True\ Positive}{False\ positive + True\ Negative} \quad (26)$$

where True negative is truly rejected and False positive is identified wrongly.The classification accuracy of the proposed approach depends on the number of beats correctly classified (true positives + true negatives) and is determined as follows,

$$Classification\ Accuracy = \frac{Number\ of\ instances\ classified\ correctly}{Total\ number\ of\ instances} \quad (27)$$

B. Discussion

Figure 4 shows the accuracy of different classification techniques with respect to different types of heart beats. The proposed approach CNNDT attains 95.32 % sensitivity overall, while the SVM, BBNN attains 93.64 %, 89.14%.

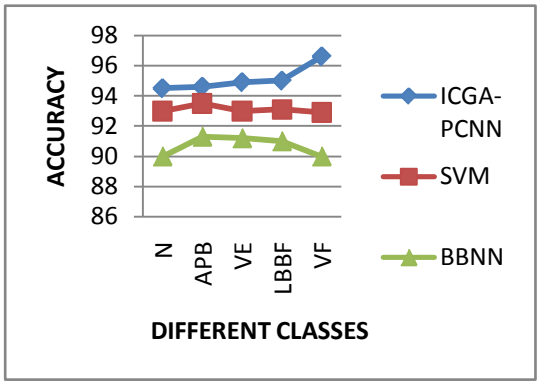

Figure 4: Accuracy

Figure 5 shows the sensitivity of different classification techniques with respect to different types of heart beats. The proposed approach CNNDT attains 95.32 % sensitivity overall, while the SVM, BBNN attains 93.64 %, 89.14%.

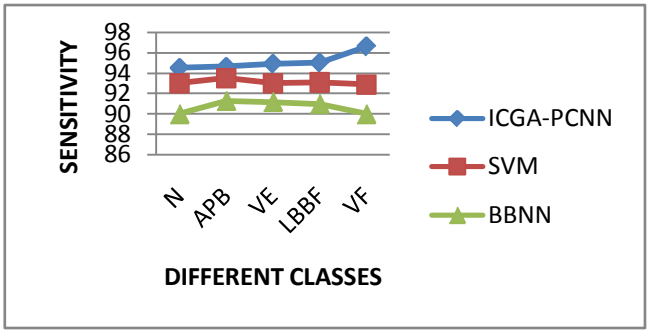

Figure 5: Sensitivity

Figure 6 shows the specificity of different classification techniques with respect to different types of heart beats. The proposed approach CNNDT attains 95.32 % sensitivity overall, while the SVM, BBNN attains 93.64 %, 89.14%.

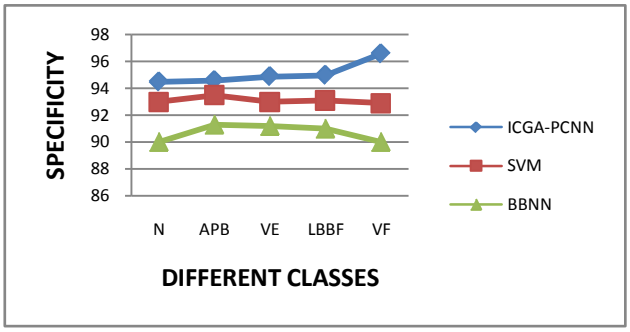

Figure 6: Specificity

VI. Conclusion

This paper proposed an effective ICGA-PCNN approach to classify the different types of arrhythmia based on histogram and morphological features extracted from detected QRS complex. The ECG signal is de-noised with the well-known approach, namely "infinite impulse response", which reduces the distortion of the segment. The histogram and morphological features of each QRS complex from the de-noised signal are extracted using the Pan-Tompkins based QRS detection. The MFNN classifier is used to classify the different types of arrhythmias Normal N, LBBB, APB, VE and VF. The MFNN structure and the error value are tuned by using particle based clonal selection algorithm in training stage in order to increase the robustness of the MFNN classifier. The experimental results show that proposed classification approach performs better than the existing approaches such as BBNN and SVM in terms of accuracy, sensitivity and specificity.

References

[1] R. Chitra and V. Seenivasagam, "Review Of Heart Disease Prediction System Using Data Mining And Hybrid Intelligent Techniques", Intact Journal On Soft Computing, Vol.3, No.04, 2013.

[2] N.G. Bhuvaneswari Amma "Cardiovascular Disease Prediction System using Genetic Algorithm and Neural Network", International Conference on Computing, Communication and Applications, 2012.

[3] C. Cortes and V. Vapnik, "Support-Vector network", Machine Learning, Vol.20, 1995.

[4] J. Habetha, "The MyHeart project-fighting cardiovascular diseases by prevention and early diagnosis", 28th Annual International Conference of the Engineering in Medicine and Biology Society, Pp. 6746-6749, 2006.

[5] L. Kaur, "Predicting Heart Disease Symptoms using Fuzzy C-Means Clustering", International Journal of Advanced Research in Computer Engineering & Technology, Vol.3, No.12, 2014.

[6] C.S. Dangare and S.S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques" International Journal of Computer Applications, Vol.47, No.10, 2012.

[7] M.B. Wadhonkar, P.A. Tijare and S.N. Sawalkar, "Artificial Neural Network Approach for Classification of Heart Disease Dataset", International Journal of Application or Innovation in Engineering & Management, Vol.3, No.4, 2014.

[8] S.M. Alzahani, A. Althopity, A. Alghamdi, B. Alshehri and S. Aljuaid, "An overview of data mining techniques applied for heart disease diagnosis and prediction", Lecture Notes on Information Theory Vol.2, No.4, 2014.

[9] K. Srinivas, G.R. Rao and A. Govardhan, "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", 5th International Conference on Computer Science and Education (ICCSE), Pp. 1344-1349, 2010.

[10] Y. Wu and L. Zhang, "ECG classification using ICA features and support vector machines", International Conference on Neural Information Processing, Pp. 146-154, 2011.

[11] P. Podder, M.M. Hasan, M.R. Islam and M. Sayeed, "Design and implementation of Butterworth, Chebyshev-I and elliptic filter for speech signal analysis", International Journal of Computer Applications, Vol.98, No.7, 2014.

[12] H. Ying, C. Silex, A. Schnitzer, S. Leohardt and M. Schiek, "Automatic step detection in the accelerometer signal", 4th International Workshop on Wearable and Implantable Body Sensor Networks, Vol.13, Pp.80-85, 2007.

[13] J. Pan and W. Tompkins, "A real-time QRS detection algorithm", IEEE Trans. Biomed. Eng., BME, Vol.32, No.3, Pp. 230–236, 1985.

[14] K.D. Chinchkhede, G.S. Yadav, S.R. Hirekhan and D.R. Solanke, "On the Implementation of FIR Filter with Various Windows for Enhancement of ECG signal", International Journal of Engineering Science and Technology (IJEST), Vol.3 No.3, Pp.2031-2040, 2011.