[Rani, 5(7): July 2018] ISSN 2348 – 8034
DOI- 10.5281/zenodo.1305020 Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

DETECTION AND CLASSIFICATION OF CARDIAC ARRHYTHMIAS USING ECG BIG DATA ANALYSIS

Vanita Rani*1 & Er. Gurjit Singh Bhathal[2]
*1Research Scholar: Computer Engineering, Punjabi University, Patiala
[2]Assistant Professor: Computer Engineering, Punjabi University, Patiala

ABSTRACT

In this work, we have presented a feature extraction method in order to differentiate normal ECG signal from arrhythmias LBBB and RBBB ECG pulses. The method first detect the high amplitude R peaks from the signal, Then interval between two R signals is used to locate other peaks named as P, Q, S and T. After that five different types of time domain features has been extracted which are named as mean NN, SDNN, SDSD, RMSSD and pNN50 from samples of 22 patients. For classification, training and testing of features has been carried out using K-nearest neighbor and decision tree classifiers. Experimental results shows that proposed system effectively classifies the normal, LBBB and RBBB ECG pulses into corresponding classes out of which decision tree gives 95% accuracy in classification whereas k-NN classifier gives 91% accuracy in classification.

Keywords: *Electrocardiogram, Cardiac, Arrhythmias, Diagnostic, QRS complex detection, decision tree, k-nearest neighbor.*

I. INTRODUCTION

Heart rate variability (HRV) analysis attempts to assess cardiac autonomic regulation through quantification of sinus rhythm variability. Sinus rhythm time series is derived from the RR interval (interval between consecutive heartbeats) sequence by extracting only normal sinus to normal sinus (NN) inter-beat intervals.

The traditional analysis of heart rate variability (HRV) in the time and frequency domains seems to be an independent predictive marker for cardiovascular mortality, including sudden cardiac death. Besides linear methods there are a large number of non-linear approaches in HRV analysis where the extracted parameters quantify complicated processes and their complex relationships. The use of nonlinear methods in combination with parameters of the time and frequency domains in HRV offers possibilities for improved classification of HRV behavior. It is suggested that this could lead to a better risk stratification [1].

II. RELATED WORK

Shyjila P.A. et al. [2] proposed the framework for arrangement of arrhythmia is found to deliver preferable outcomes over the current frameworks for grouping. The framework utilizes a self-loader technique for AEA location which is observed to be successful and less tedious than doing it physically totally.

S. Udhaya Kumar et al. [3] proposed a novel bijective soft set based classification method for ECG signal classification from MIT - BIH data base. In the feature extraction module, they have extracted morphological features as the effective features for differentiating various types of ECG beats. Then, for the classification stage improved bijective soft set is applied and evaluated for ECG beats recognition of five different classes of ECG signals.

S. Senthil Kumar et al. [4] proposed MSR based classification of ECG signals for classification of ECG signals. The important morphological features were extracted from ECG signals. The method proposed for feature extraction was an effective method in classifying five different cardiac conditions.

U. Snekhalatha et al. [5] presented ECG signal analysis is done on the data collected from MITBIH database through a number of processing stages like preprocessing and feature extraction. The classification stage is in progress which will be implemented by NN-based MLP technique that is supposed to yield good accuracy and sensitivity measures.

R. Pereira et al. [6] presented a detailed report about the execution and computational time of managed grouping algorithms with respect to the errand of arrhythmia identification in ECG signals. The primary commitments of their work are: to assess the OPF classifier in the assignment of arrhythmia location, to assess six separations with OPF, among which the best exactness rates were gotten by the Manhattan metric.

HarjeetKaur et al. [7] analyzed ECG motion through HLZT and PCA. The primary phase of denoising has been helped out through the new approach of HLZT which yields enhanced SNR and MSE demonstrating the relevance of HLZT for denoising ECG signals. Further, location of QRS complex and R-crest is a noteworthy piece of ECG flag investigation. The identification is performed utilizing PCA and thresholding.

Sandeep Raj et al. [12] presented symmetrical morphological highlights that can be separated in time-recurrence space utilizing DOST that holds the supreme stage data which is gotten bring down dimensional space utilizing PCA and joined with the dynamic highlights, i.e. RR-interim highlights.

Ali Isin et al. [13] presented an effective exchanged profound learning based ECG order framework to bring out oblivious ECG arrhythmia diagnostics by arranging persistent ECG's into steady three diverse sort of heart circumstances; ordinary, paced or right package branch piece. After the ECG records are obtained from the online MIT-BIH arrhythmia database, they are separated from commotions and QRS waves are identified to remove R-T sections of the ECG.

UdayMaji et al. [14] propose another approach of ECG flag investigation by unearthly deterioration of flag utilizing VMD calculation. VMD calculation is a non-recursive mode deterioration show, where the modes are extricated simultaneously. Each deteriorated modes are band-restricted around an inside recurrence evaluated on-line.

ISSN 2348 – 8034
Impact Factor- 5.070

III. PROPOSED WORK

A. System Module

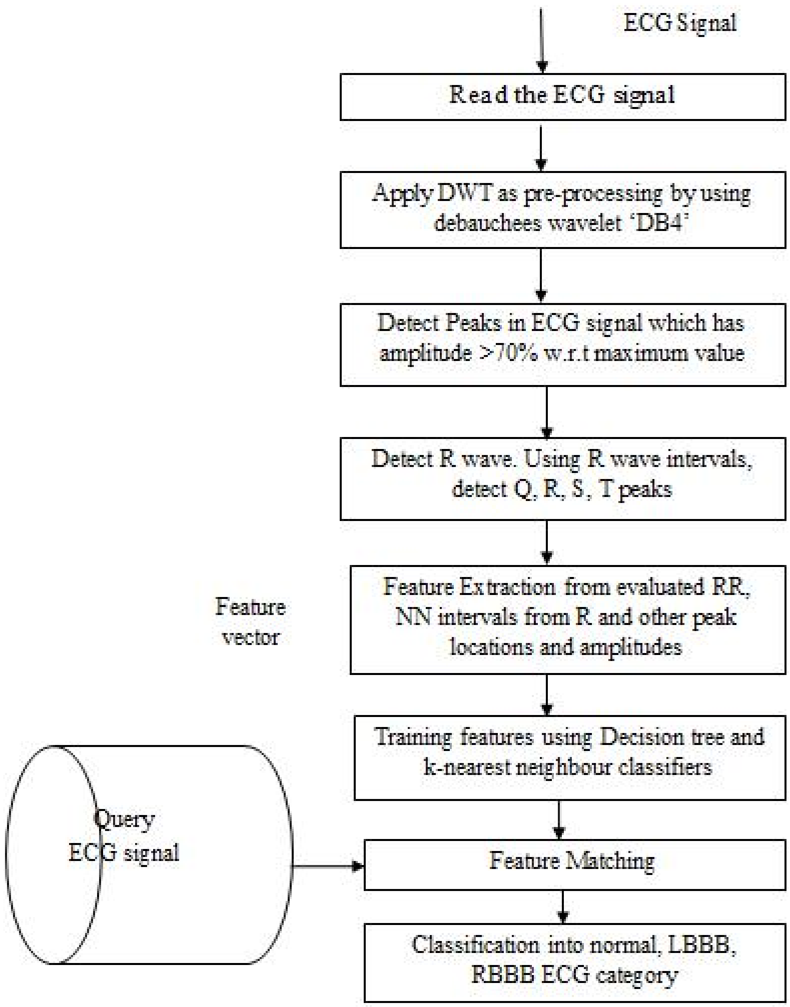

Figure 1: Flowchart for the proposed method

B. Steps in the algorithm

- ***Preprocessing***

The raw ECG signal is first preprocessed in sequence by wavelets, to remove all noise sources to reduce the artifacts in the signal. Three different signals are used for ECG analysis

- ***R peak detection***

This module is used to initialize values of parameters for the classification in detection and classification module; i.e., the running averages of R wave amplitude (Rth) and RR interval (RRth).

- ***Detection of Q,R,S,T peaks***

This module consists of the detection sub module and the classification sub module. The algorithm for detection and classification of heartbeats is used to determine Q, R and T wave peak locations of individual heartbeats and to classify heartbeats as either normal or abnormal. The detection of QRS complex is based on a well-known Pan-Tompkins QRS detector and supplemented with additionally extracted parameters (such as Q, R and T wave peak location, R wave amplitude, RR interval) to achieve reliable QRS detection and classification performance. In this

ISSN 2348 – 8034
Impact Factor- 5.070

we use mean of normal to normal intervals, standard deviation of NN intervals, square root of the mean of the sum of the squares of differences between adjacent NN intervals and covariance of NN intervals.

- ***Classification***

The most important characteristic of the classification module is its ability to distinguish between normal and LBBB and RBBB heartbeats. The classification is made based on evaluation of heart rhythm and amplitude of the R wave peak.

C. Brief of the steps used in the algorithm

1. Wavelet Based Approach

Like time-frequency representation of short-time Fourier transforms (STFT), wavelet transform (WT) of a function gives a time-scale representation. In contrast to STFT, WT uses a set of analyzing functions derived from a mother wavelet to get details of time and frequency resolution for different frequency bands. This mother wavelet is a short oscillation with zero mean as depicted in Figure 2, as an example.

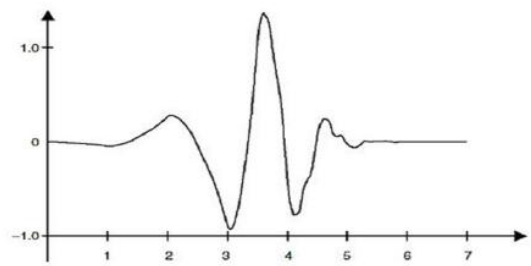

Figure 2: Example of a wavelet function (Daubechies-4 wavelet) [8]

Figure 3 clarifies the correlation between a function's local maxima in its wavelet transform and its singularity relation. The peak classification is acquired by computing the singularity degree also known as the peakiness of the wave.

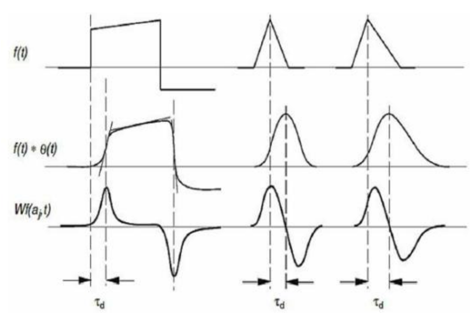

Figure 3: Example of the correlation between a function's local maxima in its wavelet transform W,f (a, t) and its singularity. The mother wavelet is the derivative of a smoothing function θ(t) [8].

Below is a general framework about the functioning of wavelet function. The resulted output is stored in C and L matrix as shown in the figure4. As we use approximation coefficients of level two we need two extract cA2 from the matrix. The 'appcoeff' function has been used for this purpose.

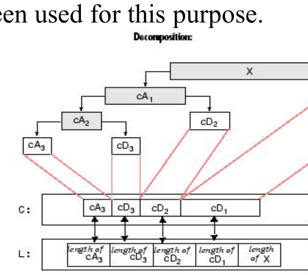

Figure 4: Decomposition using wavdec function

ISSN 2348 – 8034
Impact Factor- 5.070

2. Feature extraction phase

In our software tool, the peak-to-peak QT interval was used instead of the standard QT interval because the Q and T wave peaks can be determined more reliably than the start of Q wave and the end of T wave. The corrected ppQT interval was determined for each heartbeat as the ppQT interval divided by the square root of the corresponding RR interval. After classification and visual verification of heartbeats within the selected ECG segments, HRV analysis can be performed. HRV analysis is based on analysis of fluctuations of the so-called normal- to-normal (NN) intervals, the RR intervals originating in sinoatrial node [9].

Table 1: ECG and time domain parameters used in feature extraction phase

Parameter	Description	Unit
ECG parameters		
RR	Interval between two R waves	ms
NN	Interval between two R waves belonging to normal heartbeats; normal-to-normal interval	ms
R amplitude	Amplitude of R wave	mV
ppQT	Peak-to-peak QT interval	ms
ppQTc	Corrected ppQT interval (ppQT divided by the square root of the corresponding RR interval)	$Ms^{1/2}$
Time-domain parameters		
mean NN	Mean of normal-to-normal (NN) intervals	ms
SDNN	Standard deviation of NN intervals: estimate of overall HRV	ms
SDSD	Standard deviation of differences between adjacent NN intervals; estimate of short-term HRV; describes parasympathetic activity	ms
RMSSD	Square root of the mean of the sum of the squares of differences between adjacent NN intervals: estimate of short-term HRV; describes parasympathetic activity	ms
pNN50	Number of pairs of adjacent NN intervals differing by more than 50 ms divided by the total number of all NN intervals	%

3. Feature training using K-nearest neighbor

In KNN the classifier consider the test data into the class which has majority of votes among its k neighbors [10]. The criterion by which nearest neighbors are obtained during a test is Euclidian distance. In case of KNN all the training data is stored by the classifier and when testing data is given to the classifier it search for its k nearest data points and label the data to the new set that contains majority of its k neighbors. The values of k varies depends on the user.

4. Feature training using Decision trees

The basic philosophy in DT learning is finding the attribute or feature that best classifies the data[11]. A lot of variants of the tree learning methodology are available in the literature. Among them, the ID3, ASSISTANT, C4.5 and CART are the most popular ones. It utilizes the concepts of information gain and entropy. The information gain is primarily used as the splitting criteria to divide the nodes while building the tree. However information gain is derived from a measure called entropy which is commonly used in information theory to characterize the homogeneity of the data examples. If S is a collection of a target attribute that can take c different values, the the entropy of S relative to c labels classification is defined as: $Entropy(S) = \sum_{i=1}^{c} -pi\ log2\ pi$ where pi is the proportion of values in S belonging to the class label c.

 ISSN 2348 – 8034
Impact Factor- 5.070

5. Classifiers testing

The classifiers have been trained by training data is now tested that how much they learn to identify an exudate image this is done by giving testing data to the classifiers. The outputs are calculated for all the ECG samples in the testing data and compared with the tag values to evaluate the performance of each classifier respectively.

IV. RESULTS

Proposed algorithm has been applied on a dataset of 22 ECG signals take from Phsiotank database. The experimental setup was implemented in MATLAB software package R2015b. Out of total 8600 beats, 4200 normal beats (N), 2200 left bundle branch block beats (LBBB) and 2200 right bundle branch block beats (RBBB) were used for testing the classification performance using proposed method. Out of 48 patient records, 22 records have been used for detection of ECG beats, the record selection for different types of beats used for classification are as follows: normal ECG beat (101, 103, 105, 115, 121, 122, 123, 202, 205, 234), LBBB ECG beats (107, 109, 111, 207, 214, 217) and RBBB ECG beats (118, 119, 124, 212, 231, 233). The distribution ECG beats according to its patent record number is described in Table1. Below is the figure showing the ECG signal from the database.

Table 5.1: Distribution of the ECG record and types of beat

Beats type	MIT-BIH arrhythmia database records	Total number of beats
NORMAL	101. 103, 105, 115, 121, 122, 123, 202, 205, 234	4200
LBBB	107, 109, 111, 207, 214, 217	2200
RBBB	118, 119, 124, 212, 231, 233	2200

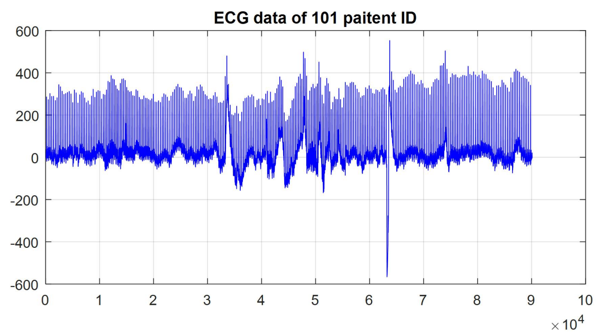

Figure 5: ECG signal taken from database

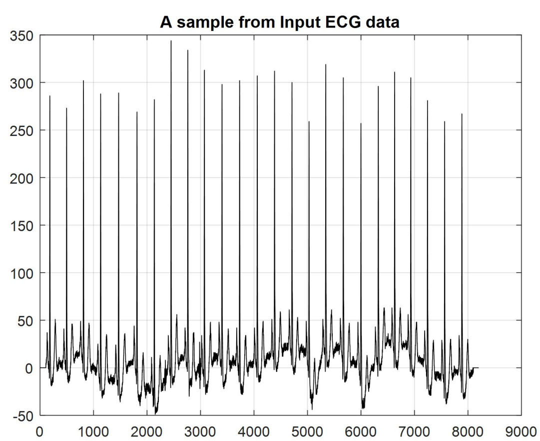

Figure 6: A sample from input ECG signal 101

[Rani, 5(7): July 2018]
DOI- 10.5281/zenodo.1305020

ISSN 2348 – 8034
Impact Factor- 5.070

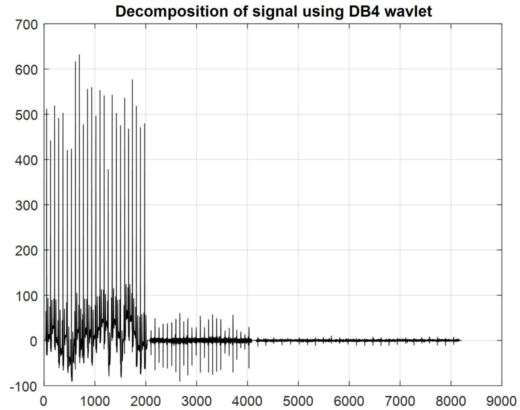

Figure 7: Decomposed signal containing high and low frequency components

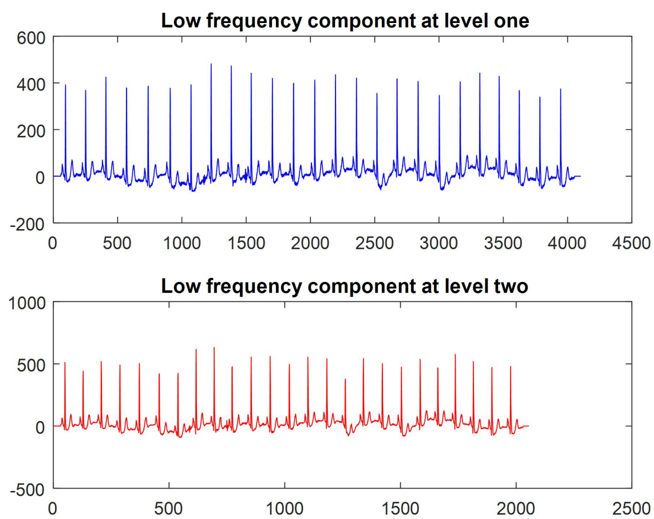

Figure 8: Low frequency components at level one and level two

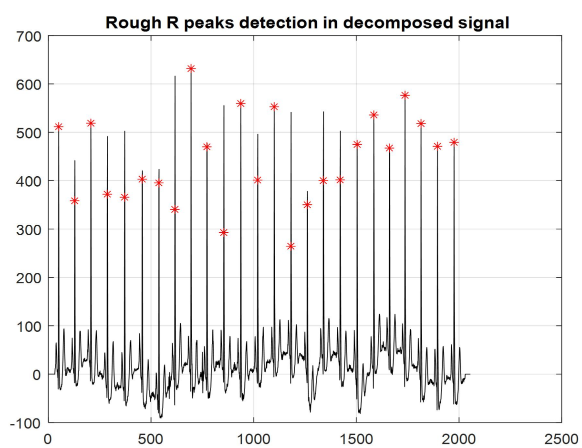

Figure 9: Detection of 'R' peaks in sample data of ECG 101

ISSN 2348 – 8034
Impact Factor- 5.070

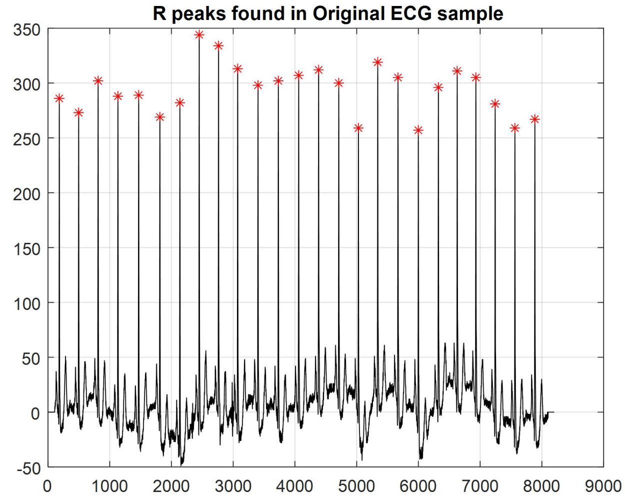

Figure 10: Sample taken from ECG signal 101

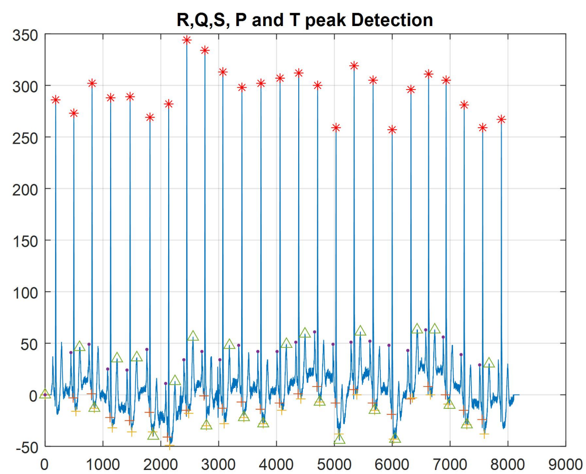

Figure 11: R,P,Q,S and T peaks in sample data of ECG 101

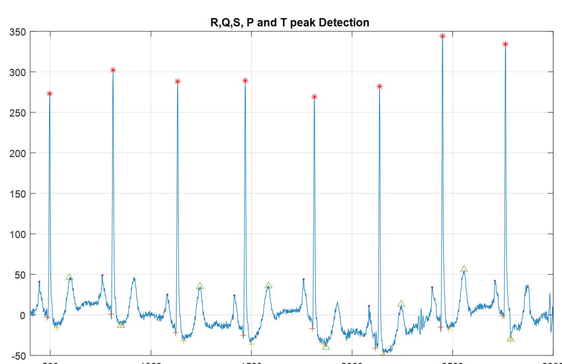

Figure 12: R, P, Q, S, T detected points zoomed view in sample data of ECG 101

ISSN 2348 – 8034
Impact Factor- 5.070

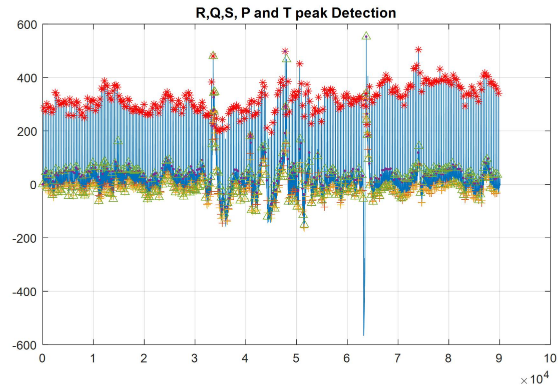

Figure 13: R, P, Q, S, T detected points in sample data of ECG 101

Table 3: Confusion matrix for testing ECG signals using decision tree classifier

Total		Predicted Class		
		NORMAL	**LBBB**	**RBBB**
Actual Class	**NORMAL**	10	0	0
	LBBB	0	6	0
	RBBB	1	0	5

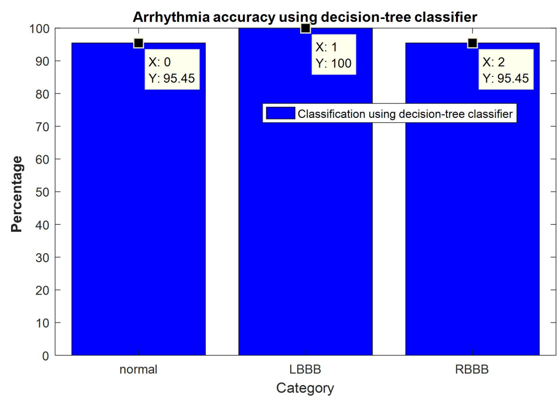

Figure 14:Arrhythmia classification accuracy using decision-tree classifier for normal, LBBB and RBBB categories

Table 4: Performance evaluation of decision classifier using accuracy, sensitivity and specificity parameters

Type of ECG signal	Accuracy	Sensitivity	Specificity
Normal	0.9545	1	0.9167
LBBB	1	1	1
RBBB	0.9545	0.83333	1

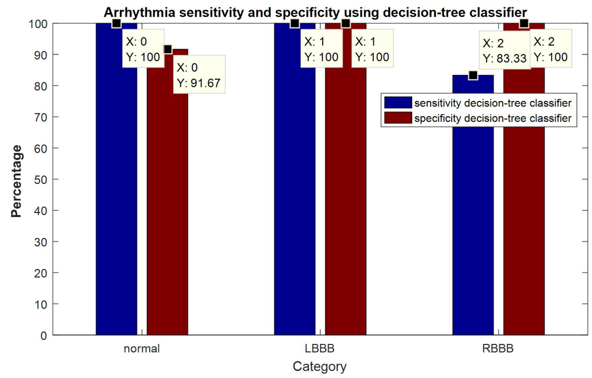

Figure 15:Arrhythmia classification sensitivity and specificity using decision-tree classifier

V. CONCLUSION

Heart rate variability has been explored in this work in order to detect abnormal arrhythmias in ECG signals. The ECG big data from MIT-BIH Arrhythmias database which includes 22 patient records are utilized to analyze and detect the two types of abnormal heartbeats along with normal heartbeat. At first R wave peaks has been detected and between two RR intervals other P, Q, S and T peaks has been detected. Then RR Interval between two R waves, NN Interval between two R waves belonging to normal heartbeats, Amplitude of R wave and (ppQT) Peak-to-peak QT interval has been find out. Then five features named as mean NN, SDNN, SDSD, RMSSD and pNN50 has been evaluated for all data samples. For classification, training and testing of features has been carried out using K-nearest neighbor and decision tree classifiers. The detection of the abnormal heart-beat from ECG big data using proposed novel feature extraction technique give exceptional accuracy performance with decision tree classifier. 22 ECG samples were tested classified into three classes: normal, LBBB and RBBB with great accuracy. The abnormal beats LBBB and RBBB are accurately classified along with the normal heartbeat using decision tree classifier which gives approx. 95% accuracy in classification of ECG samples and with kNN, it gives about 91% accuracy in classification.

REFERENCES

1. *D. Verlinde, F. Beckers, D. Ramaekers, A.E. Aubert, "Wavelet decomposition analysis of heart rate variability in aerobic athletes" Published in Autonomic Neuroscience: Basic and Clinical, 90 (1–2) (2001), pp. 138–141*
2. *AnjuMariyamZacharia, Prof. Shyjila P.A., Dr. Jubilant J. Kizhakkethottam, "Cardiac Arrhythmia Classification Using Atrial Activity Signal" Published in: Procedia Technology Volume 24, 2016, Pages 1406-1414*
3. *S. Udhaya Kumar and H. Hannah Inbarani, "Classification of ECG Cardiac Arrhythmias Using Bijective Soft Set" Published in: Big Data in Complex Systems pp 323-350*
4. *S. Senthil Kumar and H. Hannah Inbarani, "Modified Soft Rough Set Based ECG Signal Classification for Cardiac Arrhythmias" Published in: Big Data in Complex Systems pp 445-470*
5. *Sukanta Bhattacharyya and U. Snekhalatha, "Classification of Right Bundle Branch Block and Left Bundle Branch Block Cardiac Arrhythmias Based on ECG Analysis" Published in: Artificial Intelligence and Evolutionary Algorithms in Engineering Systems pp 331-341*
6. *Victor Hugo C. de Albuquerque, Thiago M. Nunes, Danillo R. Pereira, Eduardo Jose´ da S. Luz, David Menotti, Joa˜o P. Papa, Joa˜o Manuel R. S. Tavares, "Robust automated cardiac arrhythmia detection in ECG beat signals" Published in: Neural Computing and Applications February 2018, Volume 29, Issue 3, pp 679–693*
7. *HarjeetKaur, RajniRajni, "On the detection of Cardiac Arrhythmia with Principal Component Analysis" Published in: Wireless Personal Communications December 2017, Volume 97, Issue 4, pp 5495–5509*
8. *B. U. Köhler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection," IEEE Engineering in Medicine and Biology Magazine, vol. 21, no. 1, pp. 42–57, 2002*
9. *Barbara Mali, Sara Zulj, RatkoMagjarevic, DamijanMiklavcic, TomazJarm, "Matlab-based tool for ECG and HRV analysis" Biomedical Signal Processing and Control, Volume 10, 2014, Pages 108-116*

[Rani, 5(7): July 2018] ISSN 2348 – 8034
DOI- 10.5281/zenodo.1305020 Impact Factor- 5.070

10. *Mahdi Hariri and FahimeHakimi "Image-Splicing Forgery Detection Based On Improved LBP and K-Nearest Neighbors Algorithm" Electronics Information & Planning, 2015, 3(0304-9876), p. 7*
11. *S. Jhajharia, S. Verma and R. Kumar, "A cross-platform evaluation of various decision tree algorithms for prognostic analysis of breast cancer data," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, 2016, pp. 1-7*
12. *Sandeep Raj, Kailash Chandra Ray, Om Shankar, "Cardiac arrhythmia beat classification using DOST and PSO tuned SVM" Published in: Computer Methods and Programs in Biomedicine Volume 136, November 2016, Pages 163-177*
13. *Ali Isin, Selen Ozdalili, "Cardiac arrhythmia detection using deep learning" Published in: Procedia Computer Science Volume 120, 2017, Pages 268-275*
14. *Uday Maji, Madhuchhanda Mitra, Saurabh Pal, "Characterization of cardiac arrhythmias by variational mode decomposition technique" Published in: Biocybernetics and Biomedical Engineering Volume 37, Issue 3, 2017, Pages 578-589*
15. *Kaur, Ramandeep; Bhathal, Gurjit Singh; "An Extraction of Cancer Region from MR Images Using Fuzzy Clustering Means and Morphological Operations World Academy of Science, Engineering and Technology", International Journal of Computer and Information Engineering, Vol.1, No.5, 2014.*
16. *"Sachdeva, Arshika; Gupta, Gaurav; Bhathal, Gurjit Singh; "Review of Data Mining in Contrast to Modern Medical Equipments", International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM), Vol.5 , No.3,pp.18-20, 2017.*