International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

# Tree-Based Classification of ECG Signal Associated with DWT Features

**Soe Myat Thu**

University of Computer Studies, Yangon

**Abstract:** ***Nowadays, Electrocardiogram (ECG) is one of the most widely used techniques for diagnosing cardiac diseases. The system presents the methods to analyze electrocardiogram (ECG) signal, detect the QRS complex, and extract the morphological features and temporal features according to the different data. Firstly, the input ECG signal is often contaminated by noise. In order to extract useful information from the noisy ECG signals, the raw ECG signal has to be preprocessed. The baseline wandering is significant and can strongly affect ECG signal analysis. In the preprocessing stage, the input noisy ECG signal has been eliminated with the powerful high and low pass filters. Then, the extracted features from the ECG signals achieve using Discrete Wavelet Transform. The system classifies the heart beats types on the extracted features using Tree-based classification. Data are obtained from the records of the MIT-BIH database. The implementation of the approach is accomplished using Matlab 2016a programme software and the Experimental results for the system quality is measured the accuracy of the system.***

**Keywords:** ECG signal, Discrete Wavelet Transform, Tree-based classification.

## 1. Introduction

ECG is an important diagnostic tool in the diagnosis of cardiac as well as some metabolic problems. To learn an ECG correctly, one has to be thorough with the basic knowledge of electromechanical system of the human's heart. It also requires a lot of imaginations and logic conclusions. The function of human body is frequently associated with signals of electrical, chemical, or acoustic origin. These signals convey information which may not be immediately perceived but which is hidden in the signal's structure and reflect properties of their associated underlying biological systems.

The most important category of the biomedical signal is the signals which are originated from the heart's electrical activity. This electrical activity of the human heart, though it is quite low in amplitude (about 1 mV) can be detected on the body surface and recorded as an electrocardiogram (ECG)signal. The ECG, i.e. voltage measured as a function of time, arise because active tissues within the heart generate electrical currents, which flow most intensively within the heart muscle itself, and with lesser intensity throughout the body. The flow of current creates voltages between the sites on the body surface where the electrodes are placed. In this regard the ECG has been established as a fast and reliable tool for deciphering the current status of the heart and has been also widely used in prognosis and diagnosis of various cardiovascular diseases and abnormalities. In general the normal ECG signal consists of P, Q, R, S and T waves and in particular the QRS complex reflects the electrical activity within the heart during the ventricular contraction, the time of its occurrence as well as its shape provide much information about the current state of the heart. A typical ECG signal of a normal heartbeat (or cardiac cycle) consists of a P wave, a QRS complex and a T wave. P-wave is depolarization of atria. QRS-complex is depolarization of ventricles. T-wave is repolarization of ventricles. QRS complex detection is most important task in ECG analysis. Its detection is the first step of all kinds of automatic feature extraction. QRS detector must be able to detect a large number of different QRS morphologies. Most of the energy of the QRS complex exits between 3 Hz and 40 Hz. P and T waveform which can also provide with some important information about physiological conditions of patient suffering from heart disease. Therefore, The automatic classification of ECG signal has been gained so much importance over the few decades. Apart from saving the lives of thousands, it helps cardiologist make decisions about cardiac arrhythmias more accurately and easily.

In this proposed thesis, it proposes a method that automatically yields formal representations of six types ECG signal statistics based on Decision Tree classification to six heart beats types of classifications. The main objective of the heart beats types classification system is developed to determine the heat beats for the people based on the ECG signal and automatically classified the heat beats types of the people using Discrete Wavelet Transform the Matlab Programming software. Our approach is implemented on the following objectives are to diagnose heart attack and abnormalities of the heart, to check the health of the heart when other diseases or conditions caused and to reduce the workloads for the specialist doctors.

It believes that this recommender system simultaneously satisfy the following requirements such as Accuracy and Diversity. Accuracy is that the system detects more accuracy that patient's data is more correctly. Diversity is that the recommended input data performed by various patients. A diversity of people increases unknown heart beats types that have recorded data.

## 2. Related Works

An electrocardiogram, also called ECG or EKG, reflects the electrical activity of the heart. Every heart contraction produces an electrical impulse that is caught by electrodes placed in the skin. The heartbeat produces a series of waves with a time-variant morphology. These waves are caused by voltage variations of the cardiac cells.The cardiac cycle starts in the atria and goes down throw the ventricles, so the impulse is triggered in the atria and it precedes the heart contraction [1].

The ECG is measured at the body surface and results from electrical changes associated with activation first of the two small heart chambers, the atria, and then of the two larger heart chambers, the ventricles. A typical ECG signal is shown in Figure 1.

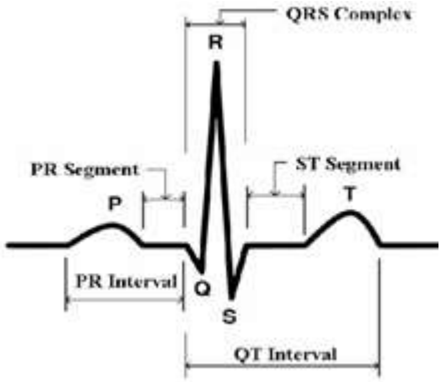



**Figure 1:** A Typical ECG signal

The contraction of the atria manifests itself as the ‘P’ wave in the ECG and contraction of the ventricles produces the feature known as the ‘QRS’ complex. The subsequent return of the ventricular mass to a rest state “repolarization” produces the ‘T’ wave. Repolarization of the atria is, however, hidden within the dominant QRS complex. Analysis of the local morphology of the ECG signal and its time varying properties has produced a variety of clinical diagnostic tools.

In this work, the two subsections are included that are Features Extraction and Classifications Methods. Many researchers proposed the feature extraction and classification methods in last decade. Mehta and M. Kumari proposed that the recorded ECG signal is often contaminated by noise. In order to extract useful information from the noisy ECG signals, the raw ECG signals have to be processed. The baseline wandering is significant and can strongly affect ECG signal analysis. The detection of QRS complexes in an ECG signal provides information about the heart rate, the conduction velocity, the condition of tissues withinthe heart as well as various abnormalities. It supplies evidence for the diagnosis of cardiac diseases. An algorithm based on wavelet transforms (WT's) has been developed for detecting ECG characteristic points. Discrete Wavelet Transform (DWT) has been used to extract relevant information from the ECG signal in order to perform classification. The QRS complex can be distinguished from high P or T waves, noise, baseline drift, and artifacts [2].Z. Zidelmala, A. Amiroua, M. Adnaneb, A. Belouchrani proposed that the use of wavelet detail coefficients for the accurate detection of different QRS morphologies in ECG. This is based on the power spectrum of QRS complexes in different energy levels since it differs from normal beats to abnormal ones. This property is used to discriminate between true beats (normal and abnormal) and false beats [3]. S.Mukhopadhyay, S.Biswas, A.B. Roy and N.Deypresented a multi-resolution wavelet transform based system for detection ‘P’, ‘Q’, ‘R’, ‘S’, ‘T’ peaks complex from original ECG signal [4]. F. Scholkmann, J. Boss and M. Wolfpresented a method for automatic detection of peaks in noisy periodic and quasi-periodic signals. The method, called automatic multiscale-based peak detection (AMPD), is based on the calculation and analysis of the local maxima scalogram, a matrix comprising the scale-dependent occurrences of local maxima [5]. J. H. Kim, S. E. Park, Gyeo-WunJeung and Kyeong-Seop Kim proposed a new method to detect R-peaks in electrocardiogram by using the prediction value from adaptive linear neuron (ADALINE) artificial neural network [6]. . S. Jayalalitha, D. Susan, S. Kumari and B. Archana proposed the KNN method to analyse the ECG signal [7]. H. M. Rai, A. Trivedi proposed ECG signal analysis for abnormalities detection using discrete wavelet transform and Back Propagation Neural Network is addressed. Proposed technique used to detect the abnormal ECG Sample and classify it into two different classes (normal and abnormal) [8]. Pramod R. Bokde1 and Choudhari N.K presented a simple method to indirectly estimate the range of certain important electrocardiogram (ECG) parameters using photoplethysmography (PPG). The proposed method, termed as PhotoECG, extracts a set of time and frequency domain features from fingertip PPG signal. A feature selection algorithm utilizing the concept of Maximal Information Coefficient (MIC) is presented to rank the PPG features according to their relevance to create training models for different ECG parameters [9].

## 3. The Proposed Tree-based Classification with DWT features System

The electrocardiogram (ECG) is a signal that records the electrical activity of the heart. The analysisof ECG signals can provide clinicians with valuable information about the patient health condition. The Electrocar- diogram (ECG) is an important and commonly used diagnostic in cardiac disease.A typical ECG signal of a normal heartbeat (or cardiac cycle) consists of a P wave, a QRS complex and a T wave. P-wave is depolarization of atria. QRS-complex is depolarization of ventricles. T-wave is repolarization of ventriclesQRS complex detection is most important task in ECG analysis. Its detection is the first step of all kinds of automatic feature extraction. QRS detector must be ableto detect a large number of different QRSmorphologies. Most of the energy of the QRS complex exits between 3 Hz and 40 Hz. P and T waveform which can also provide with some important information about physiological conditions of patient suffering from heart disease.

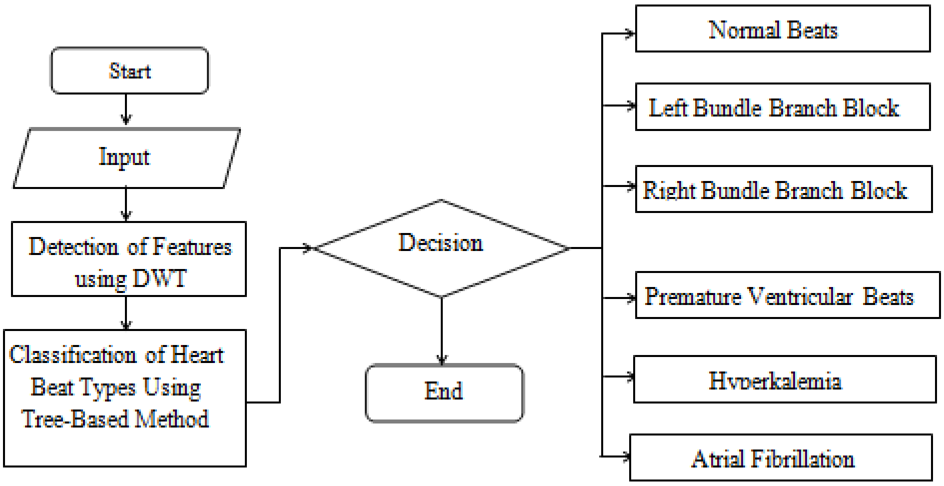



**Figure 2:** Overview of the Proposed System

The system presents the methods to analyze electrocardiogram (ECG) signal, detect the QRS complex, and extract the morphological features and temporal features according to the different data. The extracted features from the ECG signals will achieve using Discrete Wavelet Transform. The system will classify the heart beats types on the extracted features using Tree-Based Classification Method. Data are obtained from the records of the MIT-BIH database.

### 3.1. MIT/BIH Database

The ECG signal that is used in this proposed system is part of the MIT/BIH Arrhythmia Database , obtained from online.The database contains 48 data records.ECG signals (.dat files) downloaded from Physionet are first converted in to MatLab readable format (.mat file).

### 3.2 Implementation of the proposed system

The methodologies of the proposed system are:
1) Preprocessing (Noise Elimination)
2) Features Extraction using Discrete Wavelet Transform
3) Tree-based Classifications

**1) Preprocessing (Noise Elimination)**

There are several types of noise that affect the ECG such as Baseline wander (BW) and 60 Hz interference. Baseline wander is usually caused by appears respiration and as a low frequency artifact. The removal of this disturbance is an important step in ECG signal analysis, to produce a stable signal. The Baseline wander having a frequency range of (0Hz ,..., 0.5Hz). According to Nyquist's rule, if the original signal has a highest frequency $f_{max}$ , it requires a sampling frequency $f_s \geq 2f_{max}$ . Hence, at each decomposition level j, the frequency axis is recursively divided into halves at the ideal cut-off frequencies $f_j = f_{max}/2^j$. The system will propose to eliminate the baseline is based on wavelet decomposition up to level 8, which generates a set of approximation coefficients(C8), and eight sets of detail coefficients(d1,…,d8). By cancellation of approximations, the filtered signal is recovered from the details only. This is equivalent to a high-pass filter cutoff frequency $f_c = f_{max}/256$.

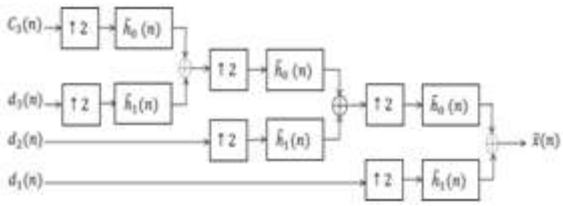


**Figure 3(a):** A three-level two-channel iterative filter bank : forward DWT

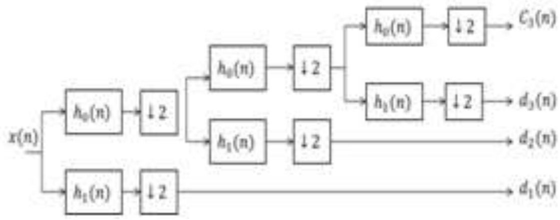


**Figure 3(b):** A three-level two-channel iterative filter bank : inverse DWT

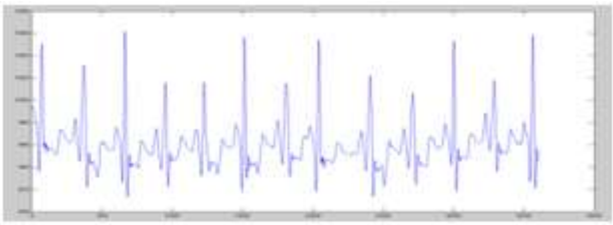


**Figure 4:** Original Signal and Denoised Signal

## 4. Features Extraction using Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) has become a powerful technique in biomedical signal processing. The system propose ECG features extraction based on wavelet detail coefficients and heartbeats types classification using Rule-Based Algorithm. After filtering, the high frequency contents of the ECG waveform are represented by DWT detail coefficients d1 to d8. DWT uses scale $s$ and position $\tau$ values based on powers of two. The values of s and τ are: $s = 2^j$ , $\tau = k*2^j$ and $(j, k) \in Z^2$ as shown in (1)

$$\varphi_{s,\tau}(t) = \frac{1}{\sqrt{2^j}}\varphi\left(\frac{t-k*2^j}{2^j}\right) \ldots\ldots(1)$$

Where, t = time

j = decomposition level
k = constant.

The key issues in DWT and inverse DWT are signal decomposition and reconstruction, respectively. The basic idea behind decomposition and reconstruction is low-pass and high-pass filtering with the use of down sampling and up sampling respectively. One can choosethe level of decomposition $j$ based on a desired cutoff frequency. To choose the best coefficient, the system compare the correlation coefficient between all the decomposed signals individually with the original ECG signal using percent of cross-correlation formula:

$$C = 100.\ \frac{\sum_{i=1}^{N} x(i).y(i)}{\sqrt{\sum_{i=1}^{N} x^2(i).\sum_{i=1}^{N} y^2(i)}} \qquad (2)$$

Where: x is the original ECG and y is the ECG reconstructed with $d_n$ coefficient.

**QRS Complex, P and T wave Detection**

The following algorithm summarizes the procedure used for detecting R peak localization through Daubechies (db4) wavelet coefficients:
Step 1: Apply Daubechies wavelet db4 to ECG signal
Step 2: $y_c$ = maximum correlation
Step 3: QRS localization: The system use the hard thresholding.
Step 4: Rejection of same QRS: if i and j are consecutive selected positions : |i - j| < 36 then i and j correspond to the same QRS , QRS duration < 100ms is standard QRS duration. The system initialize a constant Nqrs = 0 ( Nqrs is the number of QRS complexes) , which is incremented if the difference between two indexes in "th_index" is greater than 36 (100ms). Nqrs is the actual number of QRS correctly detected. R peaks are the local maximum of QRS intervals currently detected, but it must eliminate multiple detection: a peak occurring within the refractory period (200 ms) is eliminated.
Step 5: Detect Q and S points by finding the smallest value in the range R location.
Step 6: Detect T and P points by finding the highest value in the range R location.

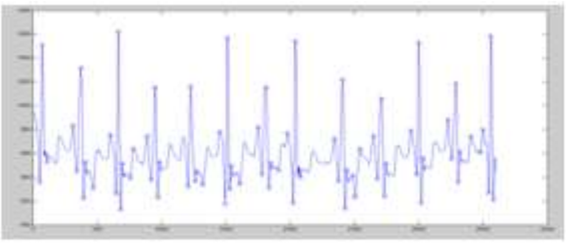


**Figure 5:** Peak value of QRS Complex features System

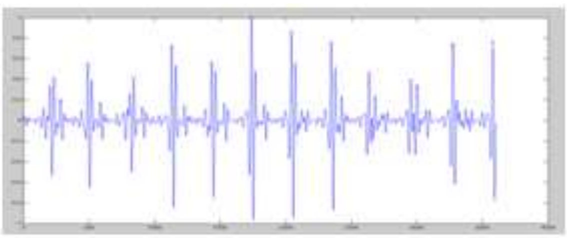


**Figure 6:** R peak value

## 5. Tree-based Classifications

Decision tree is a tree-like graph of decisions. Each branch represents the decisions to be made graphically. It is a non-parametric supervised approach. It partition input into uniform classes. This method permits the acceptation and rejection of class label at each intermediate stage. Breiman in 1996 [10]proposed Bagging or bootstrap aggregating that is a typeofensemble learning which combines classifiers in order to get maximum accuracy that a single classifier cannot provide. In BDT classifier, the training data is divided into subsets by using bootstrap resampling, and each subset is used as training data to construct each decision tree. The number of bootstrapping defines the number of constructed DT, and outputs of DTs trained by different subsets are applied to majority voting stage [11]. The proposed system will determine the heart beats types of features extracted results using Tree-Based Classification Method. The block diagram of BDT is shown in Figure 7.

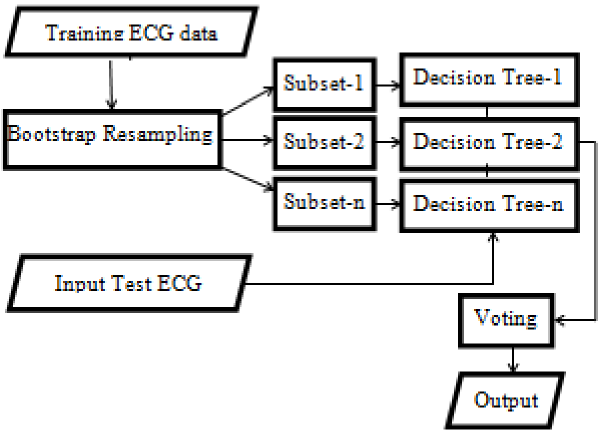



**Figure 7:** Tree-based Classifier

**Table 1:** Classification of Heart Beat Types

| NO | Classification of Beat Types |
|---|---|
| 1 | Normal Beats(N) |
| 2 | Left Bundle Branch Block(LBBB) |
| 3 | Right Bundle Branch Block(RBBB) |
| 4 | Premature Ventricular Beats(PVB) |
| 5 | Hyperkalemia(H) |
| 6 | Atrial Fibrillation(AF) |

**1) Normal Beat (N)**

Normal heart beat is generally characterized by 60 to 100 beats per minute. The regularity of the R-R interval varies slightly with the breathing cycle. Test ECG data is 100. The proposed system test the ECG data is 100 in Figure 8.

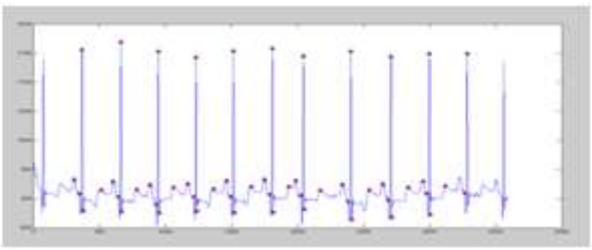
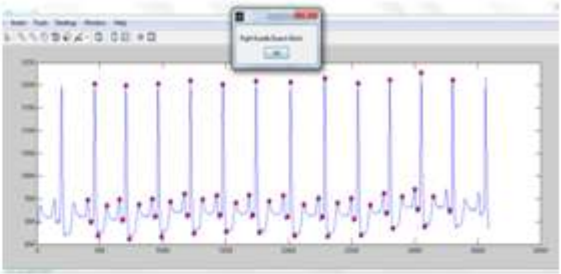


**Table 2:** ECG Sample Use For Testing

| NO | Beat Types | MIT/BIH Data |
|---|---|---|
| 1 | Normal Beats(N) | 100,101,223 |
| 2 | Left Bundle Branch Block(LBBB) | 202,207,214,219,221,233,111 |
| 3 | Right Bundle Branch Block (RBBB) | 105, 104, 234,124, 231 |
| 4 | Premature Ventricular Beats(PVB) | 112, 117, 210, 221 |
| 5 | Hyperkalemia(H) | 107, 113, 217 |
| 6 | Atrial Fibrillation(AF) | 202, 232 |

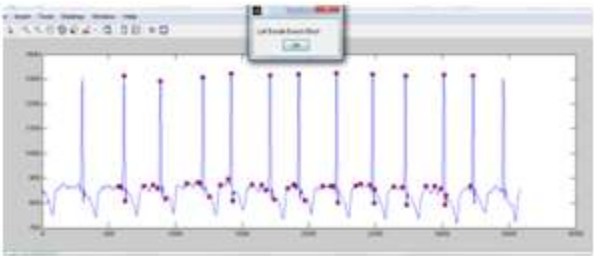
**2) Right Bundle Branch Block (RBBB)**

The proposed system test the ECG data is 101 in Figure9 . The generally significant shape of RBBB data is Tall R Wave, Wide QRS Complex, Change in shape of QRS Complex and Slurred S wave.



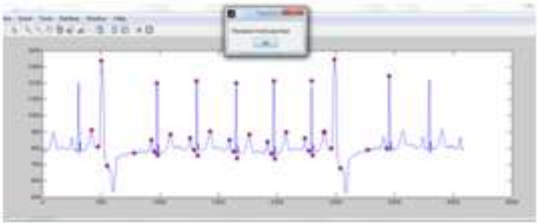
**3) Left Bundle Branch Block (LBBB)**

The proposed system test the ECG data is 124 in Figure 10 . The generally significant shape of LBBB data is Wide QRS Complex with duration of >0.12s (>3mm) and change in shape of QRS Complex.



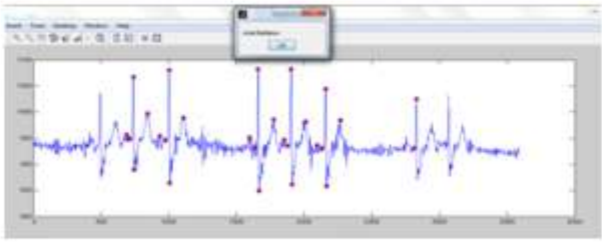
**4) Premature Ventricular Beat (VPB)**

The proposed system test the ECG data is  in Figure 11 . The generally significant shape of VPB data is Irregular Rhythm and QRS Complex duration >0.12s.



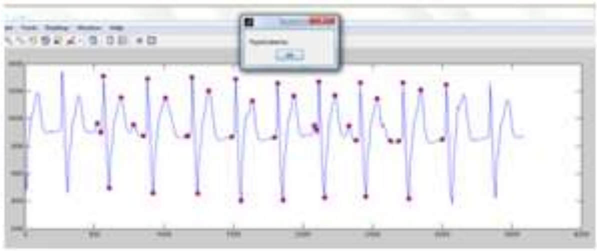
**5) Atrial Fibrillation(AF)**

The proposed system test the ECG data is  in Figure 12 . The generally significant shape of AF data is Irregular rhythmand  Absent P waves.



**6) Hyperkalemia**

The proposed system test the ECG data is in Figure13. The generally significant shape of AF data is Tall peak T wave.



## 6. Experimental Results and Further Extension

The proposed system is implemented in MATLAB 2016a environment. Figure 3 (a) and (b) presented the removal of baseline winder noise from the ECG data using high pass filter and low pass filter. Then, the proposed system illustrates successful identification ECG features such as QRS complex, R peak detection and Q, R and S features detection using Time-Frequency features with Discrete Wavelet Transform method as following figure 5 and 6. The DWT feature extraction of ECG signal algorithm firstly focus on attention on the real ECG data to detect correctly. Secondly, heart beat types classification is implemented using Tree-based classification method. The system is classified six heart beat types among the many different beats types. The experiment is conduct using MIT BIH arrhythmia database, available at 360Hz sampling rate, is used to test the performance of the proposed algorithm.

The detection rate is better than some results such as in [12]. This algorithm can be improved and extended to extract other features from ECG data, like P and T waves, PR segment, ST segment and QT interval to identify the other heart beat types.

## 7. Conclusion

The proposed system has been implemented for ECG signal classification using Discrete Wavelet Transform Technique and Decision Tree Classification method. The system is low computational overhead and good detection accuracy. It yields an average accuracy of 90%. In future, it can also try different classification methods for better result and lesser time for categorization of various kinds of abnormalities.. Cloud computing can be used to make this system available everywhere in the world.

ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

## References

[1] L. Sornmo and P. Laguna, Bioelectrical Signal Processing in Cardiac and Neurological Applications. Elsevier Academic Press, June 2005.
[2] P. Mehta and M. Kumari “Qrs Complex Detection of Ecg Signal Using Wavelet Transform”, International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11, 2012.
[3] Z. Zidelmala, A. Amiroua, M. Adnaneb, A. Belouchranib, “QRS detection based on wavelet coefficients”, computer methods and programs inbiomedicine 490–496, 2012.
[4] S.Mukhopadhyay, S.Biswas, A.B. Roy and N. Dey, “Wavelet Based QRS complex Detection”, Vol. 2, Issue 3, May-Jun 2012, pp.2361-2365.
[5] F. Scholkmann, J. Boss and M. Wolf, “An Efficient Algorithm for Automatic Peak Detection in Noisy Periodic and Quasi-Periodic Signals”, ISSN 1999-4893, 21 November 2012.
[6] J. H. Kim, S. E. Park, Gyeo-WunJeung and Kyeong-Seop Kim,” Detection of R-Peaks in ECG Signal by Adaptive Linear Neuron (ADALINE) Artificial Neural Network”, (http://creativecommons.org/licenses/by/4.0/)., MIMT,2016.
[7] S. Jayalalitha, D. Susan, S. Kumari and B. Archana, “ K-nearest Neifhbour Method of Analysing the ECG signal”, ISSN , Journal of Applied Sciences 14 (14): 1628-1632, 2014.
[8] H. M. Rai, A. Trivedi, “Classification of ECG Waveforms for Abnormalities Detection using DWT and Back Propagation Algorithm”, International Journal of Advanced Research in Computer Engineering & Technology Volume 1, Issue 4, June 2012.
[9] Pramod R. Bokde1 and Choudhari N.K, “BEAT CLASSIFICATION OF AN ECG SIGNAL USING PHOTOPLETHYSMOGRAPHYAND NEURAL NETWORK”, International Journal Of Current Medical And Pharmaceutical Research, Vol. 2, Issue, 6, pp.390-395, June, 2016.
[10] Breiman L, “Bagging predictors, Machine Learning Vol.24,pp.123-140,1996.
[11] Jerez-AragonesJ.M., Gomez-RuizJ.A. and Munoz-PerezJ, “A Combined Neural Network and Decicision Tree Model for Prognosis of Brease Cancer Telapse,Artifical Intelligence in Medicine Vol 27, pp.45-63, 2003.
[12] R. HADDADI, E. ABDELMOUNIM and A. BELAGUID, “ Discrete Wavelet Transform Based Algorithm for Recognition of QRS Complexes”, World of Computer Science and Information Technology Journal (WCSIT), ISSN: 2221-0741 Vol. 4, No. 9, 127-132, 2014.