ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

Decision Tree based Premature Ventricular Contractions Beat Classifier of ECG

Geeta Kanwar[1], Dr. Naveen Kumar Dewangan[2]

M.Tech Scholar, Department of ECE, Bhilai Institute of Technology, Durg, India[1]

Associate Professor, Department of ECE, Bhilai Institute of Technology, Durg, India[2]

ABSTRACT: Premature Ventricular Contraction (PVC) beats are the ectopic beats that generate from the ventricular area of the heart. PVC has a great significance in detecting and predicting life threatening ventricular abnormalities. The ECG signal is a periodic, because the amplitude of every wave varies from one cycle to another through the same recording. In this paper we classify the ECG signal on PVC and Normal using decision tree (C4.5) as a classifier. QRS peak and T wave of ECG signal have detected using Pan and Tompkins algorithm and Windowing and thresholding method. Normal and PVC signal are differentiated based on the features. Experiments were performed on ECG records in the MIT-BIH Arrhythmia Database. The work is implemented in the multipurpose tool MATLAB.

KEYWORDS: Premature ventricular contractions (PVCs), Electrocardiogram (ECG).

1. INTRODUCTION

Cardiovascular Diseases are one of the leading killer all over the world. According to world health organization 17.5 million people died from the CVDs in 2012. It is 31% of the all global deaths. On the above details 7.5 million people died only due to coronary heart diseases and 6.7 million were due to stroke. Cardiovascular diseases are many types and Arrhythmia is one of them. The abnormal rhythm of the heart is called arrhythmia, like Premature Ventricular Contraction (PVC), Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), PAC, Bradycardia, Tachycardia. Some arrhythmias are very dangerous for the patients and (PVC) is most common arrhythmia.

Among the ventricular arrhythmia premature ventricular contraction (PVC) is more importance because if its occurrence is higher than normal it increases the risk of sudden deaths in patients. Premature ventricular contraction beat is a form of irregular heartbeat in which the ventricular contract prematurely [1]. PVC results from the early depolarization of the myocardium generating in the ventricular area and is widespread form of arrhythmia in adults. It is often seen along with structural heart disease and increases the risk of sudden death. So that the classification or detection of PVC is very essential for the patients. In this context many classifications system has been proposed like support vector machine (SVM), artificial neural network (ANN), quantum neural network (QNN) and wavelet transform (WT) etc. Decision tree (DT) classifier is one of them but decision trees are simple to understand and interpret. They require minimum data and it able to handle both numerical and categorical data. [2]

In recent years, many studies have been done on automatic detection of cardiovascular disease. For classification of PVC, detection of QRS peak is very important which are presented in [3-5] by many authors. They proposed the Pan & Tompkins method for the detection of QRS peak. They used the time-frequency analysis of ECG. The most recently published work by different researchers are presented in [6-13]. Leigang Zhang et al. present a novel approach that classifies ECG signals with the combination of Wavelet transform and Decision tree classification. They used the wavelet transform to extract the ECG signals wavelet coefficients as the first features and used the combination of principal component analysis (PCA) and independent component analysis (ICA) to remove the first features relativity. They employ the ID3 algorithm which is a decision tree methods as the classifier to detect the different heartbeat arrhythmias [6]. Vaclav chudacek et al. discuss and compare a number of different approaches of clustering algorithms focusing on distinguishing PVC from the normal beats [7]. Smita L. Kasar et al. proposed an algorithm for the classification of healthy and Myocardial infarction signals. Decision tree (J48 and CART) based classifiers are implemented for the classification of ECG signals [8]. Lenka Lhotska et al. developed an algorithm to compares two different approaches for analysis of ECG signals. ECG records are pre-processed by the wavelet transform

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

and the machine learning method of decision trees and fuzzy rules induction are used for classification [9]. Omer T. Inan et al. proposed an algorithm for the classification of PVCs They combined wavelet transformed ECG waves with timing information as their feature set for classification. They used neural network as classifier [10]. Yasin Kaya et al. Proposed an algorithm for the classification of PVC. They used well-known machine learning methods, including neural networks, k-nearest neighbour, decision trees, and support vector machines [11]. Durgesh Kumar Ojha et al. proposed a method to study and analyse Electrocardiograph (ECG) waveform to detect abnormalities present with reference to P, Q, R and S peaks1[12]. Ahmat Mert et al. developed an algorithm for arrhythmia beat classification using decision tree [13].

In this paper an approach was developed for the classification of PVC and Normal beat. QRS Peak is detected using Pan-Tompkins algorithm and T wave is detected using Windowing and Thresholding method. The features of PVC i.e. QRS duration and T amplitude are used as a attributes for the Decision Tree classifier.

II.MATERIAL AND METHODS

ECG is the recording of heart electrical activity. It generates from the different part of heart as shown in figure 1. It generates due to the conduction system of heart. Figure 1 shows how ECG signal generates from different part of heart like Sinus node, Atrial muscle, A-V node, Common bundle, Bundle branches, purkinje fibers, Ventricular muscle. ECG signal comprises P, QRS, and T waves. Feature of ECG are P wave, PR- segment, PR-interval, QRS complex, ST-segment, T wave, ST-interval, RR-interval. All ECG signals incorporate Lead II from the MIT-BIH Arrhythmia Database were used in this proposed approach.

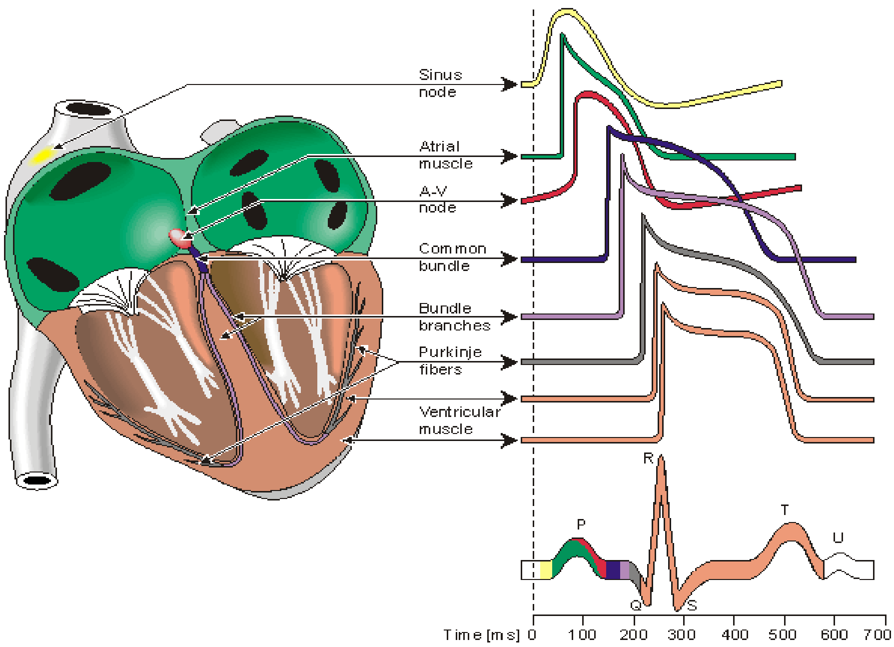

Fig 1: Generation of ECG

A. ECG DATABASE

The MIT-BIH Arrhythmia Database was used as the data source for this work. This database comprises 48 signal of 1 hour duration each, and two leads i.e. lead II and modified leads (V1, V2, V3, V4 or V5). The database signals were sampled at 360Hz [14]. We used file numbers 100, 101, 103, 105 and 220 for normal beat and 200, 203, 213, 214, 215, 221 and 223 for PVC for training and testing purpose. We take a 1 hour duration signal of these files and crop these signal on one minute duration signal to use the normal and PVC beats only. 50 signal of normal and 37 signals of PVC are used for the training the system. In our proposed work we used graphical user interface (GUI) which is a pictorial

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

interface to a program. A good GUI can make programs easier to use by providing them with a consistent appearance and with intuitive controls like pushbuttons, list boxes, sliders, menus, and so forth. The overall system structure is shown in figure 2, where we first take the MIT-BIH database as a input then after QRS Peak of the signal has detected then after T wave detection process is done. After feature selection signals are trained using C4.5 decision tree classifier and then after testing process are done. Process is described below in detailed.

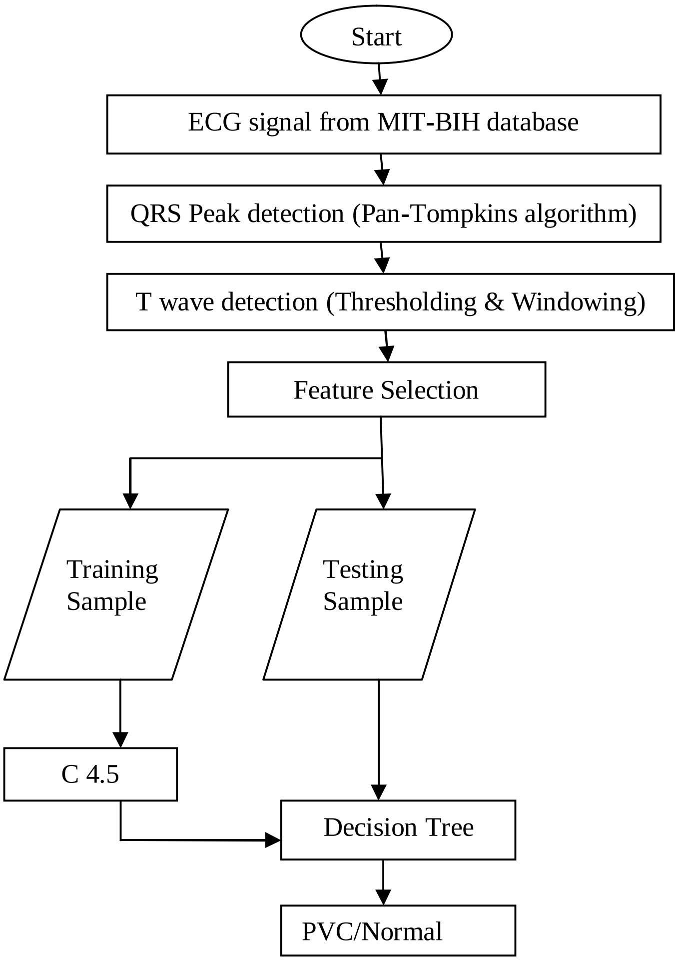

Fig 2: System Structure

B. QRS PEAK DETECTION

QRS Peak of ECG signal is detected using Pan-Tomkins algorithm. It is a most widely used algorithm. Pan and Tompkins algorithm consists of the following processing steps 1.Band-pass filtering, 2.Differentiation, 3.Squaring, 4.Moving window integration, 5.Thresholds adjustment and 6.Decision. Block diagram of Pan-Tompkins algorithm is shown in figure 3.

The first step of this algorithm is band pass filtering. Baseline of ECG signal is disrupted due external noise like muscle contraction, breathing, change in position of the electrode etc, so that the baseline of ECG has a wavy path. The baseline of ECG signal is corrected using the band-pass filter, it is a combination of low pass and high pass filter. The frequency range of band-pass filter is 5-15Hz.

After the filtering process the signal is differentiated to give the QRS Complex slope information by using the following transfer function:

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

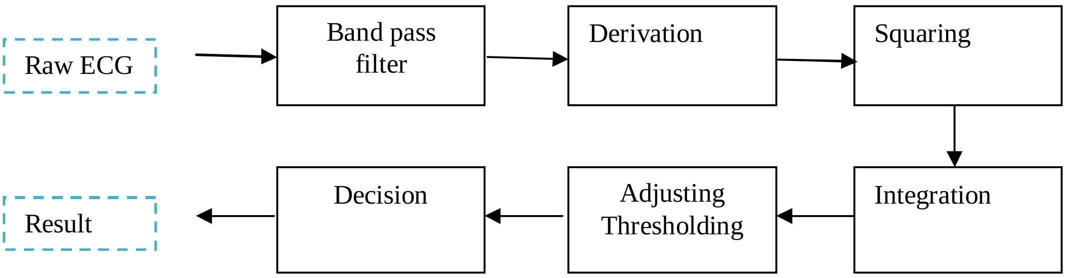

Figure 3: Pan-Tompkins Algorithm

$$(Z) = \frac{1}{8}\left[2 + Z^{-1} - Z^{-3} - 2Z^{-4}\right] \quad \ldots\ldots (1)$$

Figure 4 shows the output of band pass and derivative filter of record 221.Then after signal is squared point by point producing all the data point positive and emphasizing the higher frequency of the signal, it's given by the equation

$$X(Z) = [Y(Z)]^2 \quad \ldots\ldots\ldots (2)$$

Where X(z) and Y(Z) are respectively the Z transform of the input x(n) and output y(n) .

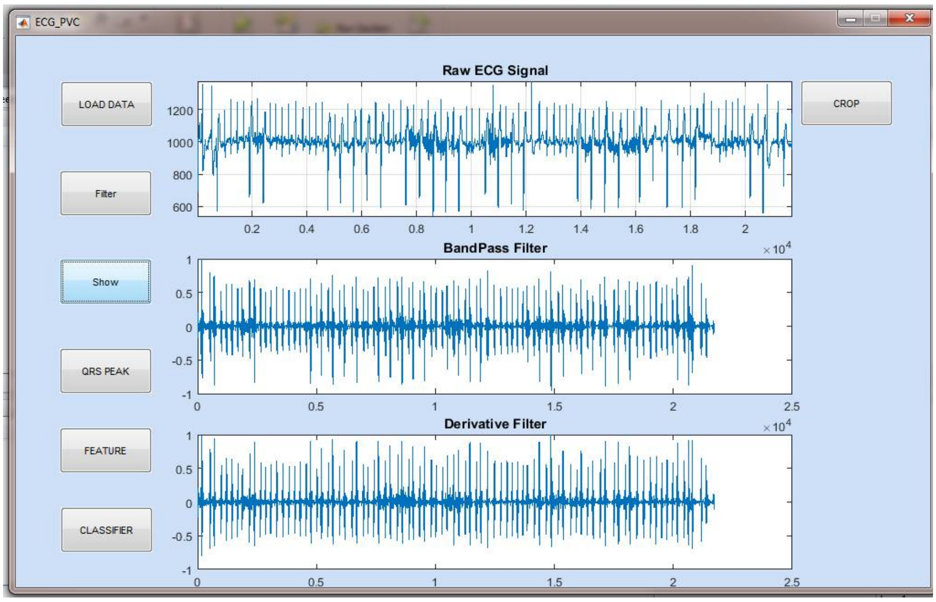

Fig 4: output of Band pass and derivative filter (Record 221)

After squaring process the signal passes along the moving-windows integrator to obtain wave-form feature information and it is calculated from:

$$H(Z) = \frac{1}{N}\left[Z^{N-1} + Z^{N-2} + \cdots \ldots \ldots Z^{N-(N-1)} + 1\right] \quad (3)$$

Where N = number of samples in the width of the integration window

In the adaptive threshold stage it adjust two threshold, first threshold to identify the peak and second is used when there is no peak detected then the algorithm go to search back for the lost peak by using a threshold lower than the first one [16]. Last step is decision stage, when a peak is detected; it is classified as a signal peak. Figure 5 shows the detected R peak of signal.

 DOI:10.15662/IJAREEIE.2016.0512035

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

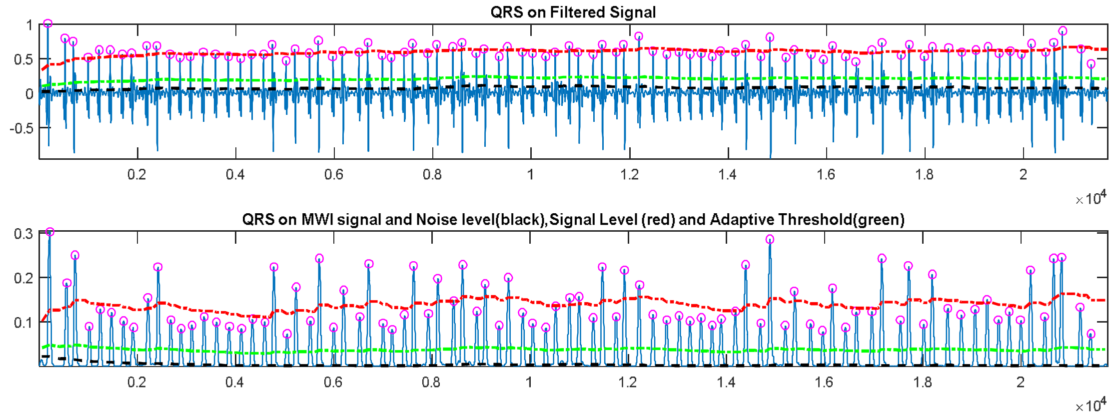

Fig 5: Detected R peak of signal (Record 221)

C. T WAVE DETECTION

Already detected QRS-complexes become a reference for detection of P and T-waves. Here we only detected the T wave of ECG signal. If we calculate S point then it is easy to calculate T points. First we set the double threshold on the ECG signal these two thresholds are based on an S wave and the baseline of the signal, named these two thresholds as Threshold 1 and threshold 2. If the signal cross the first threshold (S wave) then we considered the signal with a T wave is started and if the signal cross the second threshold (set on baseline) then it is detected as Ton. After thresholding windowing process is done, we take the window size 11 (for 11 sample) and check it for the samples after the S wave, counter is set for the next 3 sample After the s wave. When window detect the T wave then the Ti (T index) and T amplitude of the signal are saved. Figure 6 shows the detected QRS and T wave of ECG signal.

D. FEATURE SELECTION

Feature selection is also a important part of the ECG signal analysis and classification. There are many features of the ECG signal like P wave amplitude, Q wave amplitude, R wave amplitude, T wave amplitude, PR interval QRS complex duration, QTc interval, R-R interval, ST segment, R-R interval (Heart Rate) etc. For the classification of PVC and Normal beat we used the only two features of ECG i.e. QRS complex duration and T amplitude. These are the major features of PVC. QRS complex duration is find out by subtracting the Qi (Q index) from the Si (S index) and T amplitude is already detected.

E. CLASSIFICATION

There is different classification algorithm. The C 4.5 Decision Tree classification algorithm is chosen for the classification of ECG signal in to either Premature Ventricular Contractions (PVCs) or Normal signal. The features of PVC i.e. QRS duration and T wave are works as the attributes and Signal file like 100, 101 works as a class for the decision tree algorithm. A decision tree is a powerful method for classification and for facilitating decision making. Decision Tree Classifiers are used successfully in various areas such as radar signal classification, character recognition, remote sensing, medical diagnosis, expert systems.
C4.5 algorithm is the expansion of ID3 algorithm. The root node of the decision tree is on the basis of information gain of the attributes. The information gain of any attributes is calculated from the entropy. Information gain calculates the depletion in entropy (or gain in information) this give result from splitting the data into subsets based on an attribute. The information gain of example set *S* on attribute *A* is defined as,

$$\text{Gain (S, A)} = \text{Entropy(S)} - \sum \frac{|S_v|}{|S|} \text{Entropy} (S_v)$$

In the above equation, S is the number of occurrence and *Sv* is a subset of occurrence of S where A takes the value v. Entropy is a evaluation of the amount of information in an attribute. If the entropy is higher, the more the information is needed to completely describe the data. Hence, when building the decision tree, the idea is to decrease the entropy of

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

the dataset until we reach a subset that is pure (a leaf), has zero entropy and represents instances that all belong to one class. Entropy is given by,

$$\text{Entropy (S)} = -\sum p(I)\log_2 p(I)$$

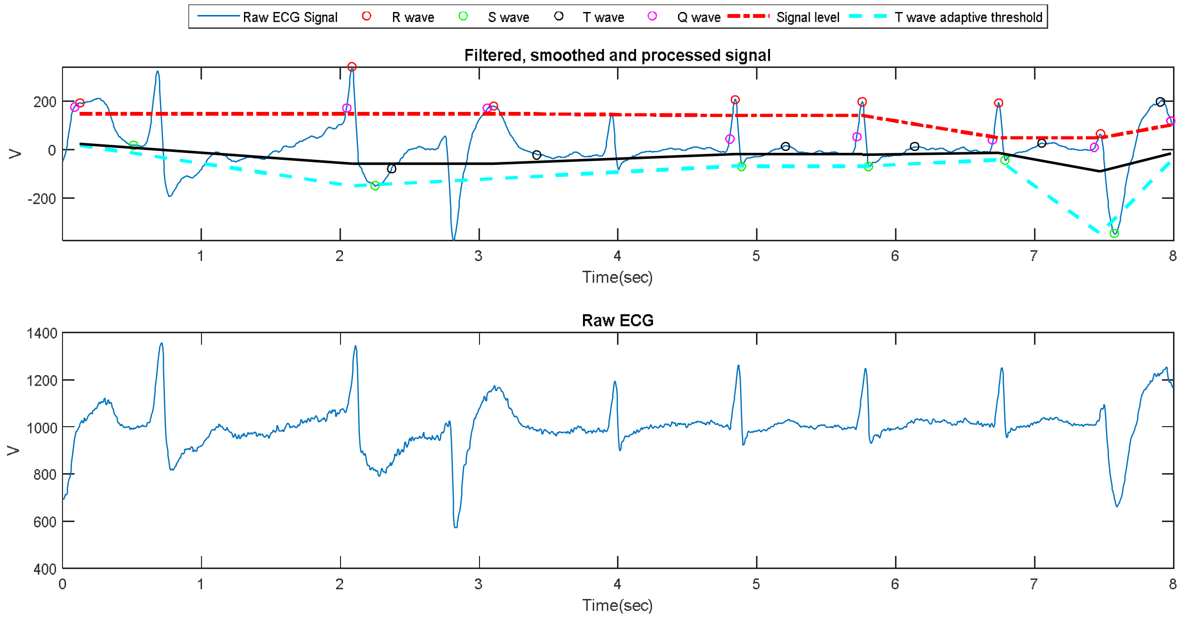

Fig 6: Resultant QRS and T wave of ECG signal

Where $p(I)$ is the proportion of S belonging to class I. It can handle both continuous and categorical attributes, whereas ID3 can handle only categorical attributes so that the C4.5 algorithm incorporates use of a statistic called the "Gain Ratio" that compensates for the number of attributes by normalizing with information encoded in the split itself [17].

$$\text{Gain Ratio} = \frac{\text{Gain(S,A)}}{\text{I(A)}}$$

In the above equation,

$$\text{I(A)} = -\sum p(I_A)\log_2 p(I_A) \quad \text{(Split Info)}$$

III. RESULT

For the analysis of proposed method total 157 signals of signal were used, which are obtained from the MIT-BIH arrhythmia database, where 87 signals have used for the testing process. Out of 87, 50 signals for normal and 37 signals for PVC, on a testing procedure. We used starting 3 beats of signals, so that total 261 beats used for the testing. 150 beats applying as normal and 111 beats as PVC. Testing results are shown below in table 1.

Table 1: Result of Classification using Decision Tree

Database	Total Beat	TP	FP	TN	FN
MIT-BIH Database	261	147	6	105	3

From the Table1 we calculate Sensitivity, Specificity, Positive Predictivity, Negative Predictivity and Accuracy of system. Sensitivity, Specificity, Positive Predictivity, Negative Predictivity and Accuracy of the system are 98%, 94.6%, 96.07%, 97.2% and 96.6% respectively.

Table 6.3 shows the comparison of previous and proposed method, it shows that presented method has better sensitivity and Positive predictivity. Sensitivity and positive predictivity of our proposed method is 98% and 96.07% respectively which is better than the previous discussed artificial neural networks optimally designed by the multi-dimensional particle swarm optimization (MD-PSO) [18] and J48 decision tree [14].

Table 6.3 Comparison of Classification

Method	Sensitivity	Positive Predictivity
Turker Ince et. al [18]	93.40%	93.30%
Vaclav Chudacek et al. [7]	91.23%	94.00%
Presented Method	**98.00%**	**96.07%**

IV. CONCLUSION

Classification of PVC based on their features is detected successfully. In recent studies accurate calculation of features play important role to classify the signal. Decision Tree (C4.5) play a very important role for normal and PVC beat classification of ECG signal. The proposed system give 96.6% accuracy for classification which is a better as compared to previous discussed ID3 classification method.

REFERENCES

[1] http://www.who.int/mediacentre/factsheets/fs317/en/
[2] http://www.heart.org/HEARTORG/Caregiver/Resources/WhatisCardiovascularDisease/What-isCardiovascularDisease_UCM_301852_Article.jsp#.V_ieKzl97IU
[1] Ribeiro, B. R., et al. "Premature ventricular beat detection by using spectral clustering methods." 2007 Computers in Cardiology. IEEE, 2007
[2] Kaya, Yasin, and Hüseyin Pehlivan. "Classification of Premature Ventricular Contraction in ECG." *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS* 6.7 (2015): 34-40.
[3] Pavlatos, Christos, et al. "Hardware implementation of Pan & Tompkins QRS detection algorithm." *Proceedings of the EMBEC05 Conference*. 2005.
[4] Debbabi, Nehla, Sadok El Asmi, and Hichem Arfa. "Correction of ECG baseline wander application to the Pan & Tompkins QRS detection algorithm." *I/V Communications and Mobile Network (ISVC), 2010 5th International Symposium on*. IEEE, 2010.
[5] Pan, Jiapu, and Willis J. Tompkins. "A real-time QRS detection algorithm." *IEEE transactions on biomedical engineering* 3 (1985): 230-236.
[6] Zhang, Leigang, Hu Peng, and Chenglong Yu. "An approach for ECG classification based on wavelet feature extraction and decision tree." *Wireless Communications and Signal Processing (WCSP), 2010 International Conference on*. IEEE, 2010.
[7] Chudáček, Václav, et al. "Comparison of methods for premature ventricular beat detection." *ITAB Conference. IEEE, Piscataway*. 2006.
[8] Kasar, Smita L., and Madhuri S. Joshi. "Analysis of Multi-Lead ECG Signals using Decision Tree Algorithms." *International Journal of Computer Applications* 134.16 (2016).
[9] Lhotska, Lenka, Jan Macek, and Daniele Peri. *Evaluation of ECG:* comparison of decision tree and fuzzy rules induction. na, 2004.
[10] Inan, Omer T., Laurent Giovangrandi, and Gregory TA Kovacs. "Robust Neural-Network-based Classification of Premature Ventricular Contractions using Wavelet Transform and Timing Interval Features." IEEE Transactions on Biomedical Engineering 53.12 (2006): 2507-2515.
[11] Kaya, Yasin, and Hüseyin Pehlivan. "Classification of Premature Ventricular Contraction in ECG." *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS* 6.7 (2015): 34-40.
[12] Ojha, Durgesh Kumar, and Monica Subashini. "Analysis of Electrocardiograph (ECG) Signal for the Detection of Abnormalities Using MATLAB." World Academy of Science, Engineering and Technology, International Journal of Medical, Health, Pharmaceutical and Biomedical Engineering 8.2 (2014): 114-117.
[13] Mert, Ahmet, Niyazi Kilic, and Aydın Akan. "ECG Signal Classification Using Ensemble Decision Tree." *J Trends Dev Mach Assoc Technol* 16.1 (2012): 179-182.
[14] Moody, George B., and Roger G. Mark. "The MIT-BIH arrhythmia database on CD-ROM and software for use with it." Computers in Cardiology 1990, Proceedings.IEEE, 1990.
[15] http://www.bem.fi/book/06/06.htm
[16] Pan, Jiapu, and Willis J. Tompkins. "A Real-Time QRS Detection Algorithm." *IEEE* transactions on biomedical engineering 3 (1985): 230-236.
[17] Quinlan, J. Ross. "Induction Of Decision Trees." Machine learning 1.1 (1986): 81-106.
[18] Ince, Turker, Serkan Kiranyaz, and Moncef Gabbouj. "Automated Patient-Specific Classification Of Premature Ventricular Contractions." 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2008.