Volume 5, Issue 1, January 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Cardiac Arrhythmias Classification using Decision Tree

Palash Mondal, Kalyani Mali
Department of CSE, University of Kalyani
West Bengal, India

Abstract— nowadays, most common reason of human death is heart disease which is cardiovascular disorder. Numerous heart diseases can be detected by analysing electrocardiograms (ECG). In this work, simple decision tree based model is used as a classification tool. The model classifies ECG signals into 5 classes, namely normal and four manifestations of heart arrhythmia: Premature Ventricular Complex (PVC), Atrial Premature Contraction (APC), Right Bundle Branch Block (RBBB), and Left Bundle Branch Block (LBBB). The total work includes three main modules: pre-processing module, feature extraction module and classification module. The ECG signals used in this work obtained from the MIT-BIH arrhythmia database. This technique is easy to implement and gives efficient result.

Keywords— Electrocardiogram (ECG); Heart Arrhythmia; Decision tree based classification; Premature Ventricular complex (PVC); Atrial Premature Contraction (APC); Right Bundle Branch Block (RBBB); Left Bundle Branch Block (LBBB).

I. INTRODUCTION

Cardiovascular diseases are the major cause of death globally. It is estimated that 23.6 million people will die from cardiovascular diseases in year 2030(WHO | Cardiovascular diseases (CVDs)) [4]. The electrocardiogram (ECG) represents the electrical activity of heart, as a waveform graph (Fig. 1 shows a schematic record of a normal heartbeat [5], in which we can observe the fiducial points P, Q, R, T). An ECG signal contains important information that can help medical diagnosis. It is the standard tool used in diagnosing heart diseases. Classification of ECG signals into different arrhythmia classes is a complex pattern recognition problem. These signals are highly nonlinear also. Therefore, different techniques such as signal processing techniques, machine learning methods were used for this purpose.

There are numerous methods to classify cardiac arrhythmias. Some most commonly used methods are Support Vector Mechanism (SVM) [8], k-NN classifier [9], Artificial Neural Networks except these there are many other methods. We get the information through literature survey [6]-[9].

The aim of the work is to build a method for detection and classification of heart diseases using ECG. Here we proposed a method for classifying normal, Premature Ventricular Complex (PVC), Atrial Premature Contraction (APC), Right Bundle Branch Block (RBBB), and Left Bundle Branch Block (LBBB) heartbeats. In this work, decision tree based classifier combined with statistical features extracted from Discrete Wavelet Transform (DWT) is used to classify ECG signals. The methodology consists of three steps. In first step, the ECG signals are pre-processed using DWT (Daubechies4 (D4)) [2] for noise reduction. In second step features are extracted from pre-processed ECG signal. In the last step, ECG signals are classified into classes using decision tree based classifier [10]-[11].

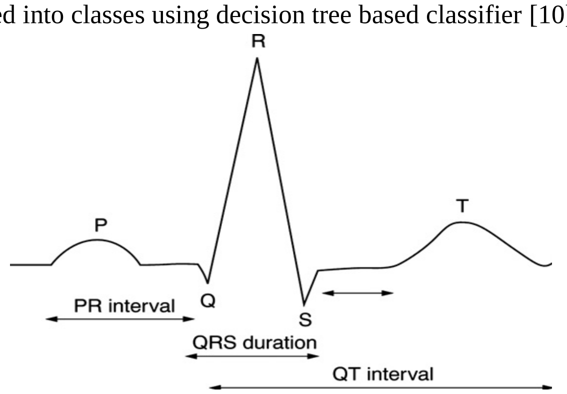

Fig. 1 A normal heartbeat ECG signal

II. ECG DATA

The prime thing for arrhythmias classification is the ECG records. In this study, the MIT-BIH-AR ECG [1] database is used for arrhythmias classification. The database consists of 48 two-lead recordings obtained from 47 subjects (recordings 201 and 202 were obtained from the same subject), each recording was measured for approximately 30 min and sampled at 360 Hz. Of the 48 recordings, 23 of them (The "100 series") were collected from routine ambulatory practice and the remaining 25 (the "200 series") were selected to include examples of uncommon but clinically important arrhythmias cases that were not well represented in the 23 100-series recordings. The ECG leads varied among the subjects due to physical limitation of electrode placement. For most of the recordings, the first channel was a modified limb lead II (MLII) (only 114 recordings used V5 as the first lead and MLII as the second lead; the leads were then swapped in the study). The second channel was usually V1 (sometimes V2, V4 or V5, depending on subjects). The database contains annotations for both QRS position and beat class information, verified by at least two independent experts. In this study, only MLII lead data are used.

III. METHODOLOGIES

The proposed method consists of three modules (shown in fig. 2). They are Pre-processing, Feature extraction and Classification. In pre-processing module noise reduction is done. Useful features are extracted from pre-processed signals in feature extraction module. Decision tree induction and testing are done in last module.

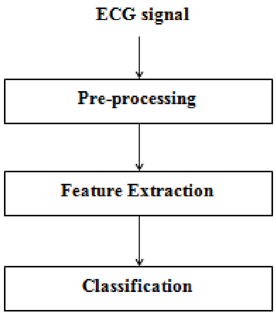

Fig.2. Steps of proposed method

A. *Pre-processing*

This module deal with noise present in the ECG signals. Presence of noise in ECG signals led classification task in wrong direction because we cannot extract correct features from the ECG due to presence of noise in ECG signals. All ECG signals from MIT-BIH [1] were collected by Holter devices. They were contaminated by baseline wandering and noises from power-line. Hence, first the ECG signals are pre-processed to remove unwanted noises from ECG. For this purpose we use Daubechies4 wavelet (D4) [3].

B. *Feature Extraction*

Feature extraction is very important task in pattern recognition problem. Classification of cardiac arrhythmia is a complex pattern recognition problem. The accuracy of pattern recognition or classification fully depends on extracted features, how discriminating they are and the correctness of them. In this work, features are extracted depending on fiducial points (P, Q, R, S and T) (shown in fig. 1). Total 15 features per beat are extracted in this work. Which are categories in three categories: intra-beat intervals, inter-beat intervals and morphological amplitudes. The intra-beat intervals are QRS duration, PR duration, RS duration, ST duration, QT duration, PT duration, PQ duration, QR duration. The inter-beat intervals are RR intervals, TP intervals. The morphological amplitudes are amplitude of R peak, Amplitude of T and P.

C. *Classification*

In this work we use greedy approach which is decision tree based classifier to classify ECG signals. It is very simple and effective method for classification. In this method first decision tree is built using known dataset which is called training data. Then the model is applied on unknown pattern to classify it. In the building phase of decision tree we analyzed the extracted features and selected best discriminating feature as splitting attribute of decision tree at every step of tree building phase. It is a greedy approach because it uses the best attribute as splitting attribute at every time of splitting. After tree building phase, the tree is tested for accuracy. Then the model is employed for classification. The sample decision tree based classifier is given below.

```
if (QRS>=0.045 and QRS<=0.065 )
        if(PQ<0.200)
              then " normal "
        else    " apc "
        end
    elseif(QRS>0.065)
        if (R amplitude<0)
             then " pvc "
             else
                  if(RS<=0.300)
                      then " apc "
                      else
                        if(PR>0.200)
                             then " lbbb "
                              else " rbbb "
                        end
                  end
        end
    end
end
```

IV. RESULT

In this study five different heartbeat classes were analysed. These are: N (normal heart beat), RBBB (Right Bundle Branch Block), LBBB (Left Bundle Branch Block), APC (Atrial Premature Contraction) and PVC (Premature Ventricular Complex). The result of the proposed method is given in table 1.

Table 1

	N	APC	PVC	RBBB	LBBB
N	128	0	0	0	0
APC	1	82	0	2	0
PVC	0	0	43	0	0
RBBB	0	0	0	97	0
LBBB	0	0	0	2	88

Here percentage of misclassification error of this model is 1.12%.

V. CONCLUSION

In this study we proposed simple decision tree based classification which classifies ECG signals into 5 Arrhythmia classes efficiently. This method is very simple to implement and time complexity is lower than other existing methods. The accuracy of the proposed method is 98.88%. This method can also classify arrhythmias using multiple lead data together by extracting discriminating features from ECG signals of multiple lead.

ACKNOWLEDGMENT

This work is done under the guidance of Asst. prof. Kalyani mali ,University of Kalyani during master degree, under the scheme : DST,PURSE scheme.

REFERENCES

[1] MIT-BIH database. Last accessed on 10/1/2015. Available at URL: http://physionet.ph.biu.ac.il/physiobank/database/
[2] Daubechies(1990), *The wavelet transform time-frequency localization and signal analysis.* IEEE Transactions on Information Theory , 5 (36), 961-1005.
[3] http://en.wikipedia.org/wiki/Daubechies_wavelet
[4] WHO | Cardiovascular diseases (CVDs). URL is: http://www.who.int/cardiovascular_diseases/en/
[5] ECGpedia, A Free Electrocardiography (ECG) Tutorial and Textbook. URL: http://en.ecgpedia.org/wiki/Main_Page.
[6] P. de Chazal, M. O'Dwyer, R. Reilly, *Automatic classification of heartbeats using ecg morphology and heartbeat interval features*, IEEE Trans. Biomed. Eng. 51(7) (2004) 1196–1206.
[7] Zhancheng Zhang, Jun Dong, Xiaoqing Luo, Kup-Sze Choi, Xiaojun Wu, *Heartbeat classification using disease-specific feature selection*, ELSEVIER. Computers in Biology and Medicine 46 (2014) 79–89.
[8] Babak Mohammadzadeh Asl, Seyed Kamaledin Setarehdan,Maryam Mohebbi. *Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal.* ELSEVIER. Artificial Intelligence in Medicine (2008) 44, 51—64.
[9] Emina Alickovic, Abdulhamit Subasi. *Medical Decision Support System for Diagnosis of Cardiovascular Diseases using DWT and k-NN.* ISSD 2012.
[10] Rokach, Lior; Maimon, O. (2008). *Data mining with decision trees: theory and applications.* World Scientific Pub Co Inc. ISBN 978-9812771711.
[11] Quinlan, J. R., (1986). *Induction of Decision Trees. Machine Learning 1: 81-106*, Kluwer Academic Publishers

