Cardiac Arrhythmia Prediction Using Fuzzy Classifier

Prof. V.S.R.Kumari
Research Scholar,
Department of ECE,
SMCE, TummalaPalem,Guntur,A.P NDIA
Vsrk46@gmail.com

Dr. P.Rajesh Kumar
Professor in ECE
Assistant. Principal, College of Engineering
Andhra University,
Visakhapatnam, INDIA
rajeshauce@gmail.com

ABSTRACT- Heart rate variations indicate current heart disease or impending cardiac diseases. Heart rate is a non-stationary signal measured using Electrocardiogram (ECG) and is used to assess cardiac Arrhythmia. It is tedious and time consuming to study ECG reports to locate abnormalities in the collected data. It is here that computer based analytical tools for in-depth data study/classification becomes useful for diagnosis. Automatic arrhythmia assessment is easier due to the advent of image processing techniques. Many algorithms were developed for ECG signals detection/classification. This paper investigates ECG classification procedures for arrhythmic beat classification based on RR interval. The process is based on RR interval beat extraction using Symlet on ECG data. Extracted RR data is used as a classification feature and beats are classified through a boosting algorithm and Fuzzy Unordered Rule Induction Algorithm (FURIA). Classification efficiency was evaluated using MIT-BIH arrhythmia database.

Keywords- Electrocardiogram (ECG), Arrhythmia classification, MIT-BIH ECG data, RR interval, Symlet, Boosting, Fuzzy Unordered Rule Induction Algorithm (FURIA)

I. INTRODUCTION

The heart’s electrical activity is recorded by ECG and is used for heart disease diagnosis. ECG is a non-invasive technique where plot of voltage is measured by leads against time. The signal is measured on the surface of the human body. Heart rate/ rhythm disorder or morphological pattern changes can indicates cardiac arrhythmia, which is detected by analysis of a recorded ECG waveform [1, 2]. High mortality due to heart diseases has raised the need for proper detection/classification of ECG arrhythmias to ensure correct treatment. ECG’s visual inspection is tiresome but recent image processing advances has resulted in automatic arrhythmia [3] assessment research. Many algorithms were developed recently for ECG signals automatic detection/classification.

Electrical signals are measured by ECG where every heart beat is displayed as a series of electrical waves having peaks and valleys. An ECG provides two kinds of information. First, the duration of an electrical wave crossing the heart can be determined by measuring time intervals on the ECG. The ECG signal’s frequency ranges from 0.05-100 Hz and dynamic ranges from 1-10 mV. ECG signal’s are characterized by five peaks and valleys labelled by successive letters of the alphabet as P, Q, R, S and T. A good ECG performance analyzing system depends on accurate detection of QRS complex and also the T and P waves. The heart’s upper chamber activation is represented by the P wave, the atria while the QRS wave (or complex) and T wave represent ventricular (lower chambers) excitation. QRS detection is of paramount importance in automatic ECG signal analysis. Once this is identified, a detailed ECG signal examination including heart rate and ST segment are performed. A P-QRS-T wave’s amplitude and duration contains heart afflicting disease information. ECGs do not provide data on cardiac contraction/pumping function. Cardiac diseases are diagnosed through the position and magnitudes of PR interval and segment, ST interval and segment, QRS interval and QT interval. [5]. Figure 1 reveals a normal ECG pattern and its components.

Fig 1: Normal ECG Signal and its P, Q, R, S and T Peaks

ECG features are represented as statistical measures or extracted from time and frequency domain. Though impressive results are provided by this procedure in some
classification tasks, they fail to discriminate all the ECG beats. Wavelet transformation which represents signals in various translations and scales [6] is another method of feature extraction. Cardiac arrhythmias classification methods based on ECG signals were developed but were weak and inaccurate. Hence ECG features are extracted for cardiac arrhythmia classification. Many machine learning and data mining methods were taken recourse to, to improve ECG arrhythmia detection accuracy.

Automated heartbeat classification was earlier studied by other researchers [7–12] and they were based on various features representing ECG and various classification methods. Features include frequency-based features [7], ECG morphology [8, 9], and heartbeat interval features [8–11]. Backpropagation neural networks [8–10], self-organizing maps with learning vector quantization [11], and self-organizing networks [12] are some of the commonly used classification methods.

ECG classification methods for arrhythmic beat classification based on RR interval are investigated in this paper. The methodology is based on RR beat interval extraction using Symlet conversion on ECG data is used. Extracted RR data is used as classification feature. Beats are classified with a boosting algorithm and Fuzzy Unordered Rule Induction Algorithm (FURIA). Classification efficiency was evaluated using MIT-BIH arrhythmia database. The remainder of the paper is organized as follows: section 2 reviews related work in literature, section 3 details the methodology. Section 4 provides results and discussion, and section 5 concludes the paper.

II. RELATED WORK

M.G. Tsipouras et al., [13] proposed a knowledge-based method for arrhythmic beat classification and arrhythmic episode detection and classification employing only RR-interval signal extracted from ECG recordings. The arrhythmic beat classification algorithm uses a three RR-interval sliding window. Classification is performed for four categories of beats: normal, premature ventricular contractions, ventricular flutter/fibrillation and 28 heart block. Input used is beat classification of a knowledge-based deterministic automation to ensure classification and arrhythmic episode detection. The six rhythm types classified include ventricular couplet, ventricular bigeminy, ventricular tachycardia, ventricular trigeminy, ventricular flutter/fibrillation and 28 heart block. The proposed methods are evaluated through the MIT-BIH arrhythmia database revealing 94% accuracy for arrhythmic episode detection and classification and 98% accuracy for arrhythmic beat classification. The proposed method is better than other complicated techniques as RR-interval signal is used for arrhythmia beat and episode classification.

KhoreichKa [14] suggested a waveform similarity and RR interval ECG beat classification mechanism. Six heart beat types (normal beat, atrial premature beat, paced beat, premature ventricular beat, left bundle branch block beat and right bundle branch block beat) were classified using this method. Wavelet transform based techniques denoise ECG signal and extracted RR intervals were used as feature. An annotated beats training database is compiled for the classifier used for waveform comparison of unknown beats. The 46 records in the MIT/BIH arrhythmia database were used for evaluation achieving a classification rate of 97.52%.

Bashir, et al., [15] proposed a nested ensemble technique for real time classification of cardiac arrhythmia. The proposed method includes training data set and feature selection manipulation. The training dataset is manipulated for classifier learning by updating training data, and feature selection to improve performance and accuracy during classification. Experiments revealed the need to consider all ECG features for evaluation and accuracy of the proposed model.

III. METHODOLOGY

A. Symlet Wavelet

Wavelets are waveforms bound in time and frequency. Wavelet analysis splits signals into shifted and scaled versions of the mother wavelet. The Continuous Wavelet Transform (CWT) is known by the wavelet function $\psi$ adding signal times multiplied by scaled and shifted versions. Mathematically the continuous wavelet is defined by

$$C(scale, position) = \int_{-\infty}^{\infty} f(t)\psi(scale, position, t)dt$$

Many wavelet coefficients $C$, a function of scale and position, are due to CWT. The original signals constituent wavelets are obtained by multiplying every coefficient by applicable scaled and shifted wavelet. Daubechies proposed symlets which are almost symmetrical wavelets and got by modifications of the db family [16]. The two wavelet families are similar, with the difference of db wavelets having maximal phase whereas symlets have minimal phase. Symlets are compactly supported wavelets with slight asymmetry and the wavelet coefficient for it can be any positive even number and highest number of vanishing moments for a specific support width.

B. Boosting

In boosting, k classifiers set is iteratively learned and each training tuple is assigned weights [17]. The process starts with a classifier $M_i$ learning, and updating its weights. This helps the next classifier, $M_{i+1}$, to assign more weights to misclassified $M_i$ training tuples. After all k classifiers are learned iteratively, the final boosted classifier, $M^*$, sums up the votes of every classifier.

Boosting allots higher weight to the classifier’s vote with lower error rate. Hence, accurate classifiers get higher weightage. Classifier $M_i$’s vote weightage is computed as follows:
The training dataset is split into subsets conforming to the “classification power” of each feature. On feature selection, theoretic measure evaluates features and this provides the choosing a suitable attribute at every level. Information-decision. Decision trees are built in a top-down method, following a path from root to tree leaves resulting in class ECG features are classified to appropriate arrhythmia class by decision trees, evaluate significance of every feature, e.g., decision-tree built with a training dataset. J48 nodes generated for classification. Unseen data is classified based on a subtree with a leaf representing majority of examples. Reduced Error Pruning (REP) trees are a procedure training samples and hence they are pruned to reduce tree learning algorithm [19]. Usually built trees overfit training samples and to ensure that they fit other examples. Reduced Error Pruning (REP) trees are a procedure to prune decision trees as it is an easy and effective pruning method but it does tend to overprune a tree. The main disadvantage of REP trees is that it requires a separate pruning dataset. However REP is extremely powerful with a large training dataset or with boosting. REP pruning replaces a subtree with a leaf representing majority of examples. Modifications are resorted to only if it reduces error either by having equal or lower number of misclassifications.

D. Boosting with REP tree
i. Reduced Error Pruning (REP) tree

The tree created is the best fit on training data in a decision tree learning algorithm [19]. Usually built trees overfit training samples and to ensure that they fit other examples. Reduced Error Pruning (REP) trees are a procedure to prune decision trees as it is an easy and effective pruning method but it does tend to overprune a tree. The main disadvantage of REP trees is that it requires a separate pruning dataset. However REP is extremely powerful with a large training dataset or with boosting. REP pruning replaces a subtree with a leaf representing majority of examples. Modifications are resorted to only if it reduces error either by having equal or lower number of misclassifications.

E. Boosting with J48

J48 algorithm is an implementation of C4.5 decision tree learner. J48 uses greedy algorithm to prompt decision trees for classification. Unseen data is classified based on a decision-tree built with a training dataset. J48 nodes generated decision trees, evaluate significance of every feature, e.g., heart rate.

ECG features are classified to appropriate arrhythmia class by following a path from root to tree leaves resulting in class decision. Decision trees are built in a top-down method, choosing a suitable attribute at every level. An information-theoretic measure evaluates features and this provides the “classification power” of each feature. On feature selection, the training dataset is split into subsets conforming to the selected feature values. This process is repeated for all subsets, till most instances in every subset belong to one class.

F. Fuzzy Unordered Rule Induction Algorithm (FURIA)

Fuzzy Unordered Rule Induction Algorithm (FURIA) is a rule-based classification method founded on RIPPER [20]. FURIA learns fuzzy rules, - not conventional rules - learns unordered rule sets and not rule lists. Its advantage is in preserving simple and comprehensible rule sets. It also includes additionally many modifications/extensions. It uses a rule stretching method in dealing with uncovered examples. Experiments reveal that FURIA outperforms original RIPPER and classifiers like C4.5, significantly as regards classification accuracy.

FURIA particularly, learns fuzzy rules instead of conventional rules and unordered rule sets instead of rule lists. It also uses an efficient rule stretching method when handling uncovered samples. Fuzzy rules are general when compared to conventional rules but have several advantages. For example, conventional (non-fuzzy) rules produce “sharp” decision boundary models with corresponding abrupt transitions among various classes. This is both questionable and not very intuitive. Instead, one expects support for a class provided rule to lower from “full” (inside the rule core) to “zero” (near the boundary) slowly and not immediately. Fuzzy rules “soft” boundaries which is their main characteristic, but soft boundaries are converted to crisp boundaries when a classification decision is planned. But the boundaries are more flexible in fuzzy rules. For example, using suitable aggregation operators to combine fuzzy rules which are not usually axis-parallel [21].

Conventional rule learners lead to a decision list to produce which rules are learned by turn for each class starting with the smallest (in terms of relative occurrence frequency) and ending with the second largest. Also, a default rule is added for a majority class. A new query instance is classified by the first rule in the list that covers it.

This has both advantages and disadvantages. For example, it might have an unwanted bias as classes are not treated symmetrically any more. Sorting rules on priority basis compromises comprehensibility (condition part of every rule includes negated conditions of earlier rules). To overcome this FURIA learns an unordered rule set. In other words, a rule set for each class in a one-versus-rest scheme which means that the resultant model is incomplete, i.e., a new query may be uncovered by any rule (this way decision lists are less problematic).

IV. EXPERIMENTAL SETUP AND RESULTS

Bundle branch block (BBB), a form of heart block involving delay/failure of conduction in a branch in the bundle of the
heart is determined by an ECG. It may be complete/incomplete, transient, permanent, or intermittent, and is named according to involvement of left or right bundle branch. Whether a bundle branch block is complete or not cannot be found out. When linked to acute anterior wall myocardial infarction, bundle branch block helps to identify a high-risk patient. Cardiac impulses inability to be conducted down bundle branches, cause an abnormally shaped QRS complex. BBB is usually seen in high-risk, acute, anterior wall myocardial infarction, caused by ischemia or necrosis of the bundle branches, trauma (as in surgical manipulation), or mechanical branch compression by a tumor. A pacemaker may be inserted if further conduction deterioration is expected.

Bundle branch block come under a group of heart problems called intraventricular conduction defects (IVCD). Two bundle branches, right and left exist. The right bundle carries nerve impulses which contract the right ventricle (the heart’s lower chamber) and the left bundle’s nerve impulses contract the left ventricle. The two bundles initially come together at a spot called the bundle of His. Nerve impulses come through the heart’s sinus node to the bundle of His and then moves into right and left bundle branches. Bundle branch block refers to a slowing/interruption of nerve impulses. Patients with right bundle branch block (RBBB) have slowly or no conducted nerve impulses. The right ventricle receives impulse through muscle-to-muscle spread, outside regular nerve pathway. This impulse transmission mechanism is slow resulting in a delayed right ventricle contraction. Many types of left bundle branch block (LBBB) exist with each having its own characteristic failure mechanism.

A dataset created from MIT-BIH arrhythmia database which evaluates the performance of the classifiers. The evaluation dataset has 165 instance; 55 events each of Right bunch bundle block, Left bunch bundle block and Normal RR interval. Continuous wavelet transforms using symlet2 filters are applied. Figures 1 – 3 show event output on application of Continuous wavelet transforms using symlet2.

V. RESULTS AND DISCUSSION

Following are the five rules obtained by FURIA:

(a2 in [-78.404, -69.722, inf, inf]) => class=l (CF = 0.94)

(a118 in [-inf, -inf, -17.512, -12.86]) and (a213 in [-inf, -inf, 1.1767, 1.2272]) => class=N (CF = 0.97)

(a118 in [-inf, -inf, -17.512, -14.293]) and (a238 in [-inf, -inf, -0.54844, -0.35637]) => class=N (CF = 0.95)

(a233 in [-0.84513, -0.15514, inf, inf]) and (a6 in [-inf, -inf, -80.307, -76.38]) => class=N (CF = 0.92)

(a156 in [-inf, -inf, -0.75362, 2.7266]) and (a82 in [2.2714, 3.0664, inf, inf]) => class=R (CF = 0.98)

Fuzzy Unordered Rule Induction Algorithm (FURIA) and Boosting is used to classify instances with a decision stump, boosting with both J48 and REP tree. Classification accuracy experiments results are provided in the following Tables/Figures. Table 1 tabulates the results summary for various techniques. Figure 4 shows the plotted classification accuracy. Figure 5 shows the plotted Root Mean Square error.

<table>
<thead>
<tr>
<th>Name of the Technique used</th>
<th>Boosting with decision stump</th>
<th>Boosting with J48</th>
<th>Boosting with REP tree</th>
<th>FURIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification accuracy</td>
<td>92.125%</td>
<td>91.52%</td>
<td>91.52%</td>
<td>92.121%</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>0.211 (21.1%)</td>
<td>0.1947 (19.47%)</td>
<td>0.1891 (18.91%)</td>
<td>0.1897 (18.97%)</td>
</tr>
</tbody>
</table>
Table 2 tabulates the precision, recall and f-Measure. The precision and recall is computed as:

\[
\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}}
\]

\[
\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images in the Database}}
\]

\[
\text{f-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

For a good classification system the values of precision and recall should be high which is seen in the boosting with decision stump and FURIA technique.

<table>
<thead>
<tr>
<th>Technique Used</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boosting with decision stump</td>
<td>0.911</td>
<td>0.921</td>
<td>0.916</td>
</tr>
<tr>
<td>Boosting with J48</td>
<td>0.91</td>
<td>0.915</td>
<td>0.912</td>
</tr>
<tr>
<td>Boosting with REP tree</td>
<td>0.915</td>
<td>0.915</td>
<td>0.914</td>
</tr>
<tr>
<td>FURIA</td>
<td>0.911</td>
<td>0.921</td>
<td>0.916</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

ECG classification method for arrhythmic beat classification based on RR interval is investigated in this paper. The methodology is based on RR interval beat extraction using Symlet conversion on ECG data. Extracted RR data is used as a classification feature. Beats are classified using boosting algorithms and FURIA. MIT-BIH arrhythmia database is used to evaluate the classification efficiency. The instances are classified as Right bunch bundle block, Left bunch bundle block and Normal RR interval. Instances are classified using Boosting with decision stump and FURIA achieved a classification accuracy of 92.12%.

VII. REFERENCES


