Early Detection and Continuous Monitoring of Atrial Fibrillation from ECG Signals with a Novel Beat-Wise Severity Ranking Approach

Haritha Gollakota

Follow this and additional works at: https://digitalcommons.memphis.edu/etd

Recommended Citation
https://digitalcommons.memphis.edu/etd/2009

This Thesis is brought to you for free and open access by University of Memphis Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of University of Memphis Digital Commons. For more information, please contact khggerty@memphis.edu.
EARLY DETECTION AND CONTINUOUS MONITORING OF ATRIAL FIBRILLATION FROM ECG SIGNALS WITH A NOVEL BEAT-WISE SEVERITY RANKING APPROACH

by

Haritha Gollakota

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Major: Computer Engineering

The University of Memphis

May 2019
This work is dedicated to my parents, teachers, family and friends who are all responsible for the little knowledge and wisdom if at all I have any.
Acknowledgement

My heartful gratitude to my thesis advisor, Dr. Bashir I Morshed, Principle Investigator, ESARP Lab, University of Memphis, whose insights and constant support has helped me get through in the right direction. At any hour of need, he has his office doors open for me, alongside monitoring my progress periodically which helped me in catching up my deadlines.

I would like to thank Dr. Madhusudhanan Balasubramanian and Dr. Hasan Ali of Department of Electrical and Computer Engineering, for accepting to be a part of my Thesis Committee and for their timely suggestions, both academic and non-academic, which has helped me view things in different perspective.

I would also like to add appreciation to my colleagues at ESARP lab, who has made the work and study environment so lively, with constant support and help throughout my stay. I would never forget the unconditional support I received from my parents in nitty gritty, and for the unfailing encouragement they gave me in moments of difficulty. I would also like to thank my family, friends and teachers who all backed me at the hour of need and without who I will not be where I am.
Abstract

Irregularities in heartbeats and cardiac functioning outside of clinical settings are often not available to the clinicians, and thus ignored. But monitoring these with high-risk population might assist in early detection and continuous monitoring of Atrial Fibrillation (AF). Wearable devices like smart watches and wristbands, which can collect electrocardiograph (ECG) signals, can monitor and warn users of unusual signs in a timely manner. Thus, there is a need to develop a real-time monitoring system for AF from ECG. We propose an algorithm for a simple beat-by-beat ECG signal multilevel classifier for AF detection and a quantitative severity scale (between 0 to 1) for user feedback. For this study, we used ECG recordings from MIT BIH Atrial Fibrillation, MIT-BIH Long-term Atrial Fibrillation Database. All ECG signals are preprocessed for reducing noise using filter. Preprocessed signal is analyzed for extracting 39 features including 20 of amplitude type and 19 of interval type. The feature space of all ECG recordings is considered for classification. Training and testing data include all classes of data i.e., beats to identify various episodes for severity. Feature space from the test data is fed to the classifier which determines the class label based on trained model. A class label is determined based on number of occurrences of AF and other Arrhythmia episodes, such as AB (Atrial Bigeminy), SBR (Sinus Bradycardia), SVTA (Supra ventricular Tachyarrhythmia). Accuracy of 96.7764% is attained with Random Forest algorithm. Furthermore, precision and recall are determined based on correct and incorrect classifications for each class. Precision and Recall on average for Random Forest Classifier are obtained as 0.968 and 0.968 respectively. This work provides a novel approach to enhance existing method of AF detection by identifying heartbeat class and calculates a quantitative severity metric that might help in early detection and continuous monitoring of AF.
Keywords: Atrial Fibrillation (AF), Electrocardiogram (ECG), Paroxysmal AF, Long-standing AF, Preprocessing, Feature extraction, Classification, Severity ranking of AF.
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>xi</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Electrocardiogram and Importance</td>
<td>1</td>
</tr>
<tr>
<td>Wearable devices</td>
<td>2</td>
</tr>
<tr>
<td>Atrial Fibrillation</td>
<td>3</td>
</tr>
<tr>
<td>2 Literature Review</td>
<td>5</td>
</tr>
<tr>
<td>ECG Signal Database</td>
<td>6</td>
</tr>
<tr>
<td>3 ECG Signal and Generation</td>
<td>8</td>
</tr>
<tr>
<td>Heart Anatomy</td>
<td>8</td>
</tr>
<tr>
<td>Electrocardiogram</td>
<td>9</td>
</tr>
<tr>
<td>Waves and Intervals</td>
<td>10</td>
</tr>
<tr>
<td>Electrodes and Leads</td>
<td>13</td>
</tr>
<tr>
<td>4 Pre-processing for Noise</td>
<td>17</td>
</tr>
<tr>
<td>Pre-processing</td>
<td>19</td>
</tr>
<tr>
<td>5 Features Extraction</td>
<td>21</td>
</tr>
<tr>
<td>P wave Morphology</td>
<td>22</td>
</tr>
<tr>
<td>QRS Morphology</td>
<td>23</td>
</tr>
<tr>
<td>Heart Rate Variability</td>
<td>24</td>
</tr>
<tr>
<td>Segmentation</td>
<td>26</td>
</tr>
<tr>
<td>Beat Annotation</td>
<td>27</td>
</tr>
<tr>
<td>6 Classification</td>
<td>35</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>35</td>
</tr>
<tr>
<td>Tree Bagger</td>
<td>38</td>
</tr>
<tr>
<td>Random Forest</td>
<td>39</td>
</tr>
<tr>
<td>Severity Classification</td>
<td>40</td>
</tr>
<tr>
<td>7 Results</td>
<td>41</td>
</tr>
<tr>
<td>8 Conclusion</td>
<td>47</td>
</tr>
<tr>
<td>Future Implementation</td>
<td>48</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. An artistic depiction of heart anatomy</td>
<td>8</td>
</tr>
<tr>
<td>2. ECG Formation</td>
<td>10</td>
</tr>
<tr>
<td>3. ECG wave formation with respective action of heart</td>
<td>11</td>
</tr>
<tr>
<td>4. Normal ECG signal</td>
<td>12</td>
</tr>
<tr>
<td>5. Power line interference in ECG signal</td>
<td>17</td>
</tr>
<tr>
<td>6. Baseline drift in ECG Signal</td>
<td>18</td>
</tr>
<tr>
<td>7. Feature Detection without filtering</td>
<td>20</td>
</tr>
<tr>
<td>8. Feature Detection with filtering</td>
<td>20</td>
</tr>
<tr>
<td>9. Normal vs Atrial Fibrillation of ECG</td>
<td>22</td>
</tr>
<tr>
<td>10. QRS complex in ECG</td>
<td>23</td>
</tr>
<tr>
<td>11. HRV in ECG signal</td>
<td>24</td>
</tr>
<tr>
<td>12. Atrial Bigeminy rhythm</td>
<td>28</td>
</tr>
<tr>
<td>13. Atrial Fibrillation rhythm</td>
<td>29</td>
</tr>
<tr>
<td>14. Normal rhythm</td>
<td>30</td>
</tr>
<tr>
<td>15. Sinus Bradycardia rhythm vs Normal rhythm</td>
<td>32</td>
</tr>
<tr>
<td>16. Supraventricular tachyarrhythmia vs Normal rhythm</td>
<td>33</td>
</tr>
<tr>
<td>17. Ventricular tachycardia vs Normal rhythm</td>
<td>34</td>
</tr>
<tr>
<td>18. Binary SVM</td>
<td>37</td>
</tr>
<tr>
<td>19. Multi class SVM with 3 hyperplanes</td>
<td>37</td>
</tr>
<tr>
<td>20. Tree bagging on Iris Dataset</td>
<td>38</td>
</tr>
<tr>
<td>21. Result from Pan Tompkins algorithm</td>
<td>41</td>
</tr>
<tr>
<td>No.</td>
<td>Description</td>
</tr>
<tr>
<td>-----</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>22</td>
<td>Result from feature detection</td>
</tr>
<tr>
<td>23</td>
<td>Plot for Severity of 11 patients</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Information of P, QRS, T waveforms</td>
<td>12</td>
</tr>
<tr>
<td>2. Estimated normal amplitudes and duration of waveforms</td>
<td>13</td>
</tr>
<tr>
<td>3. Lead placement across body for 12 lead electrodes</td>
<td>14</td>
</tr>
<tr>
<td>4. Parameters of time domain analysis</td>
<td>25</td>
</tr>
<tr>
<td>5. Parameters of frequency domain analysis</td>
<td>25</td>
</tr>
<tr>
<td>6. Criteria to extract P, Q, R, S, T, T' from ECG</td>
<td>26</td>
</tr>
<tr>
<td>7. Weights assigned to class coefficients</td>
<td>40</td>
</tr>
<tr>
<td>8. Accuracy result from Classifier</td>
<td>43</td>
</tr>
<tr>
<td>9. Accuracy of data on testing for %split of training</td>
<td>43</td>
</tr>
<tr>
<td>10. Confusion matrix result from Random forest classifier</td>
<td>44</td>
</tr>
<tr>
<td>11. Precision, Recall values from Random forest</td>
<td>44</td>
</tr>
<tr>
<td>12. AF Severity for 11 Patients</td>
<td>45</td>
</tr>
</tbody>
</table>
## List of Abbreviations

1. Atrial Bigeminy  
   **AB**
2. Atrial Fibrillation  
   **AF**
3. Beats per minute  
   **bpm**
4. Electrocardiogram  
   **ECG**
5. Heart Rate Variability  
   **HRV**
6. Quality of Life  
   **QoL**
7. Premature Ventricular Contraction  
   **PVC**
8. Random Forest  
   **RF**
9. Sinus Bradycardia  
   **SBR**
10. Support Vector Machine  
    **SVM**
11. Supra-ventricular Tachyarrhythmia  
    **SVTA**
12. Ventricular Tachycardia  
    **VT**
1. Introduction

A. Electrocardiogram and Importance

Even in this modern era, every year millions of people in USA die due to cardiac disorders, not only in USA but across the world. Most of the cardiac disorders show early symptoms, which are often unnoticed or mistook to be general illness. When functioning of heart deteriorates, it effects the proper functioning of body, and in turn, reduces the Quality of Life (QoL) of the individual.

Heart is muscular organ (myocardium) that is rhythmically driven to contract and expand there by driving the circulation of blood throughout the body. Before every normal heartbeat, a wave of electrical current passes through the entire heart which triggers myocardial contraction. The pattern of electrical propagation is not random, but spreads over the structure of the heart in a coordinated pattern which leads to an effective, coordinated systole. This results in a measurable change in potential difference on the body surface of the subject. The resultant amplified signal is known as an Electrocardiogram (ECG or EKG).

ECG is an important bio-signal representing the electrical activity and functioning of heart. Analyzing ECG signal to identify cardiac disorders or abnormalities in functioning, might help in monitoring early signs of AF, thus have an important impact in mobile health (mHealth) monitoring.

Different aspects affect the signal ECG, including abnormalities of cardiac conducting fibers, metabolic abnormalities (including a lack of oxygen or ischemia) of the myocardium, and macroscopic abnormalities of normal geometry of heart. ECG analysis is a routine part of
medical evaluation, due to the heart’s essential role in human health. Recording and analyzing the ECG is a non-invasive procedure, which helps in monitoring heart health relatively easily.

The shape of ECG waveform can give information about the functionality of heart. Manual observation of the ECG waveform by experts (cardiologist or healthcare professionals) are labor intensive and expensive. Computer based analysis can minimize the processing time in diagnosis of the cardiac diseases. Many techniques have been developed in order to assist physicians in the analysis of large volumes of patient data. Various signal processing methods have been utilized in extracting features from the biomedical signals and analyzes these features which have their own merits and demerits.

**B. Wearable devices**

Wearable devices have become increasingly popular and commercial wearables are taking shape of health monitoring devices including mHealth. Leading wearable technology makers like Apple, Garmin, Fitbit has designed and commercialized many wearables that can function as health monitoring devices, such as heart rate monitoring, or fall detection.

With the advent of Machine Learning in various fields, advanced machine learning algorithms can enable prediction and detection of cardiac disorders with very high accuracy and a marginal effort. Furthermore, these algorithms can be implemented in embedded systems environment like wearable devices [2], thereby enabling using of wearable technology-based sensing of bio-signals and can utilize automated machine learning based approached to analyze these bio-signals.

"Prevention is better than cure" as said by great intellect Desiderius Erasmus. Hence it is very important to identify any disease in early stages rather than identifying in later stages,
which may prevent usage of highly affective medications or minimally invasive procedures, which has lesser side effects on human body and leads to a healthier life (QoL). Typically, humans are not be able to feel or understand the complexity of different vitals of body. Hence there is a need for the automated machine learning algorithm to be robust, reliable, ability to operate in real-time, continuous monitoring (passive sensing), and relatively inexpensive.

C. Atrial Fibrillation

Arrhythmia is a heart condition where in heart rate is abnormal being either too slow or too high than normal rate. Normal heart rate varies for each human being. Atrial Fibrillation (AF) is a type of Arrhythmia which shows more often a higher heart rates, ranging from 100-300 beats per minute (bpm). AF is the most prevalent type of arrhythmia. It is estimated that around 2.7 – 6.1 million people are suffering with AF in USA, while it is estimated to rise to 12.1 million in 2030 [1].

Initial stages of Atrial Fibrillation have very less severe symptoms, and often mistook to be general illness. If not for medication at initial stages, this might lead to more severe stages of AF, thereby causing permanent disability to heart and its functioning. It may even lead to stroke and sudden death.

Many methods are in existence which enables detection of AF, almost all of them are involving physicians. More importantly, all of these existing methods do not address the important metric of how severe the disease is. Hence a critical need exists for AF detection techniques which highlights class of severity at a given time, along with reliable detection of AF.

A method to identify severity of AF as a probability between 0 to 1 has been identified and implemented using Random Forest Classifier. An accuracy of 96.77% is achieved alongside
precision and recall values being 0.968 each. Severity for few patients with samples of 60 beats are considered and calculated to demonstrate how severity is calculated.
2. Literature Review

ECG database and choosing the right one is one of the important aspects of detecting features. There are 5 widely available database for ECG signal [3]. MIT- BIH is the most popular and widely used in majority of publications. It has different databases, each with different sampling frequencies. EDB is another Database with a collection of 90 records sampled at 250 Hz of different age groups. AHA is another Database which has 155 records each with 250 Hz sampling rate as well. CU, NSD are the two other Database. MIT- BIH being widely available and used, is considered for this project. Samples in MIT- BIH are limited to 900,000 per record. Full record is downloaded using MATLAB code provided in Physionet.

Pre-Processing can be done in various ways, one such way enables pre-processing and feature detection together i.e., by using Discrete Wavelet Transform (DWT) [7]. In this method, it is assumed to have ECG information in between 0.5 Hz and 50 Hz. Hence these frequency components are removed. 8 levels of decomposition is done selecting a wavelet. Each stage, detail coefficients are eliminated and approximate coefficients are fed as input to next level of decomposition. Thus, at each level at least one peak is detected. FIR, recursive digital filters can also be used for pre-processing. Since they are computationally expensive to implement that butter worth filters with similar performance, butter worth filter is used for pre-processing [9].

Feature Detection can be done using DWT at 8 levels [7]. Also it can be done simply by detecting QRS complex using the most popular method of Pan Tomkins algorithm [4]. Using this as basis, P, Q, R, S, T, T' are calculated based on criteria [9]. This method is implemented for feature detection in this work. Different amplitudes are also calculated by keeping threshold on each feature directly. Using DWT hundreds of features can be extracted and Feature reduction
technique can also be used [14]. Decomposing each ECG into subbands with uniform bandwidths by incorporating filter bank is another method of beat wise feature detection [10].

Many Classification algorithms are used to classify Atrial Fibrillation from Normal signal. Each classification technique has its pros and cons. Support Vector Machine is the most widely used classification technique [12] for Arrhythmia detection. Artificial Neural Networks [15], Linear Prediction [5] are also used. Bayesian frame work classification [11], [19] is another probabilistic rather than statistical model. The performance of Naïve Bayes algorithm is lower than other classification methods. K-nearest neighbor is most straightforward classifier which is an unsupervised learning algorithm, which can classify unknown data [16], [17]. Tree Bagger is proven to be giving better results for multiclass classifications including arrhythmia [9]. Enhanced version of Tree bagger which is Random forest is used for implementation since, severity class is also considered. An automatic detection method is able to classify Paroxysmal AF signal from Normal signal [18], but none of the algorithms are able to predict severity thus far.

A. ECG Signal Database

MIT- BIH Atrial Fibrillation (afdb) and Long-term AF (Itafdb) are the database considered for this work. As these databases has normal episodes of ECG, Normal Sinus rhythm Database is not considered. This database is taken from MIT BIH repository, Physionet.

MIT- BIH Atrial Fibrillation Database includes 25 recordings of human subjects with atrial fibrillation affected patients mostly with Paroxysmal AF. Each record has 2 types of data, ECG 1 and ECG 2 pertaining to two different leads of collection. Only ECG 1 lead data is considered. Each record is of 10-hour duration, each sampled at 250 Hz sampling rate.
Bandwidth while recording ECG signal is considered to be 0.1 Hz to 40 Hz. Each record has rhythm annotation file (.atr) which is prepared manually indicating rhythm annotations. Beat annotation (.qrs) file is prepared using automated detector. Out of these 25 records, beats from 6 records are considered for training classifier during implementation.

Long-term AF Database has 84 long term ECG recordings of human subjects affected with paroxysmal, sustained Atrial Fibrillation. Each record is typically of length 24-25 hours. Each record is sampled at 128 Hz sampling rate. ‘qrs’, ‘atr’ annotations are available for each record created by automated detector, manual annotation respectively. 10 records were considered for testing classifier from this database.

Each record from Physionet has multiple toolbox options to select from, ‘Describe record’ option gives complete information of each record. Options are available to download record by time like 10 secs, 1 minute, 1 hour, to end. ‘to end’ data is considered for complete record.

*Download Records*

Records are available to download from Physiobank ATM. This download is restricted to 900,000 samples. Hence MATLAB code is created to download complete record from database.
3. ECG Signal and Generation

A. ECG Signal

In analyzing an ECG signal to detect a disease, Heart anatomy, different types of signals generated, signals of interest and their origin from different nodes needs to be identified.

a. Heart Anatomy

![Heart Anatomy](image)

*Figure 1: An artistic depiction of heart anatomy*

Heart is the central structure of the cardiovascular system. It contains four chambers. Upper chambers are called atria, lower chambers are called ventricles as shown in Figure 1. Blood is collected from all parts of the body at right atrium and gets pumped to the right ventricle. From the right ventricle blood is pumped to lungs, where it is oxygenated. The oxygenated blood is then pumped to left atrium and then to left ventricle. The left ventricle
supplies the blood to the entire body. Since the blood is supplied to entire body through ventricles, ventricular muscles are much larger than that of atria.

There are a group of specialized cells, which originates the action potentials required for the functionality of the heart. These specialized cells collectively called sino-atrial (SA) node (also called pacemaker) and is at top right of the right atrium. The action potentials, which are originated by the sino-atrial node propagates along the surface of the body in all directions. These are terminated at atrio-ventricular (AV) node which is at the center of the heart. At this point, some special fibers act as a “delay line” to provide proper timing between the action of the atria and ventricles. This electrical excitation rapidly spread all parts of ventricles through bundle. This bundle is divided into two branches called left bundle branch and right bundle branch, to initiate action potentials of the musculature of the two ventricles. Thus, Action potential has highs and lows formed by the expansion and contractions of heart muscle.

b. Electrocardiogram

Electrocardiography (ECG) is the process of recording the electrical activity of the heart over a period using electrodes placed on a patient's body. These electrodes detect the tiny electrical changes on the skin that arise from the heart muscle depolarizing during each heartbeat.

In a conventional 12 lead ECG, ten electrodes are placed on the patient's limbs and on the surface of the chest. The overall magnitude of the heart's electrical potential is then measured from twelve different angles ("leads") and is recorded over a period of time (usually 10 seconds). In this way, the overall magnitude and direction of the heart's electrical depolarization is
captured at each moment throughout the cardiac cycle. The graph of voltage versus time produced by this noninvasive medical procedure is referred to as an electrocardiogram (ECG).

The initial diagnosis of heart attack is usually made through observation of a combination of clinical symptoms and characteristic ECG changes. An ECG can detect areas of muscle deprived of oxygen and/or dead tissue in the heart. The study of the ECG signals, its waves, interval features plays a very important role in the detection of abnormalities of Heart.

c. Waves and Intervals

The ECG signal is characterized by five peaks and valleys labeled by the letters P, Q, R, S, T. In some cases, another peak called U is also used. The performance of ECG analyzing system depends mainly on the accurate and reliable detection of the QRS complex, as well as T- and P waves. The ECG wave formation during heart activity is shown in Figure 2.

![ECG Formation](image)

**Figure 2:** ECG Formation
Each action potential in the heart originates near the top of the right atrium at a point called the pacemaker or Sino-Atrial (SA) node. The wave generated by action potential, terminates at a point near the center of the heart, called the Atrio-Ventricular (AV) node.

The horizontal segment of this waveform preceding the P-wave is designated as the baseline or the iso-potential line. The P-wave represents depolarization of the atrial muscle. The QRS complex is the combined result of the repolarization of the atria and depolarization of the ventricles, which occur almost simultaneously. The T-wave is the wave of ventricular repolarization, the U-wave, if present is generally believed to be the result of after potentials in the ventricular muscle as shown in Figure 3. Hence, the duration amplitude and morphology of the QRS complex is useful in diagnosing cardiac arrhythmias and other disease states.

Figure 3: ECG wave Formation with respective action of heart
The Schematic representation of normal ECG waveform is shown in Figure 4 and the description each wave during heart pumping process is mentioned in Table 1.

![Normal ECG Signal](image)

**Figure 4**: Normal ECG Signal

<table>
<thead>
<tr>
<th>Waves</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘P’ Wave</td>
<td>The amplitude level is low (approximately 0.1-0.2 mV) and represent depolarization and contraction of the right and left Atria. A clear P wave before the QRS complex represents sinus rhythm.</td>
</tr>
<tr>
<td>‘QRS’ Complex</td>
<td>The QRS complex is the largest voltage deflection of approximately 1 mV but may vary in size depending on age and gender. The Voltage amplitude of QRS complex may also give information about the cardiac disease. Duration of the QRS complex indicates the time for the ventricles to depolarize.</td>
</tr>
<tr>
<td>‘T’ Wave</td>
<td>Represents ventricular repolarization</td>
</tr>
</tbody>
</table>
Table 2: Estimated normal amplitudes and duration of waveforms

<table>
<thead>
<tr>
<th>Features</th>
<th>Amplitude(mV)</th>
<th>Duration(ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P wave</td>
<td>0.1-0.2</td>
<td>60-80</td>
</tr>
<tr>
<td>PR- segment</td>
<td>-</td>
<td>50-120</td>
</tr>
<tr>
<td>PR-interval</td>
<td>-</td>
<td>120-200</td>
</tr>
<tr>
<td>QRS complex</td>
<td>1</td>
<td>80-120</td>
</tr>
<tr>
<td>ST segment</td>
<td>-</td>
<td>100-120</td>
</tr>
<tr>
<td>T-wave</td>
<td>0.1-0.3</td>
<td>120-160</td>
</tr>
<tr>
<td>ST-interval</td>
<td>-</td>
<td>320</td>
</tr>
<tr>
<td>RR-interval</td>
<td>-</td>
<td>(0.4-1.2) sec</td>
</tr>
</tbody>
</table>

Table 2 shows features of P-wave, QRS complex and T wave in maximum amplitude and its duration. The ECG for normal healthy person is termed Normal Sinus Rhythm (NSR). A Normal Sinus Rhythm starts in the Sino-atrial (SA) node and spreads down to the Atrio-ventricular (AV) node as the atria contract and force blood into the ventricles. The ventricles then contract and pump blood out of the heart as electrical signals reach ventricular muscle cells.

d. Electrodes and Leads

ECG signal can be recorded in different ways. Clinical analysis uses 12-lead ECG with 10 electrodes. Typical wearable devices which can have health monitoring techniques primarily uses single lead electrode due to usability considerations. Since all database use signals collected
by physicians, 12 lead electrode data are considered. Limb electrodes can be far down on the limbs or close to the hips/shoulders as long as they are placed symmetrically.

**Placement of the precordial electrodes:**

Ten electrodes are used for a 12-lead ECG. The electrodes usually consist of a conducting gel, embedded in the middle of a self-adhesive pad. The names and correct locations for each electrode can be given as below in Table 3.

**Table 3:** Lead placement across body for 12 Lead electrodes

<table>
<thead>
<tr>
<th>Electrode name</th>
<th>Electrode placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>On the right arm, avoiding thick muscle.</td>
</tr>
<tr>
<td>LA</td>
<td>In the same location where RA was placed, but on the left arm.</td>
</tr>
<tr>
<td>RL</td>
<td>On the right leg, lateral calf muscle.</td>
</tr>
<tr>
<td>LL</td>
<td>In the same location where RL was placed, but on the left leg.</td>
</tr>
<tr>
<td>V₁</td>
<td>In the fourth intercostal space (between ribs 4 and 5) just to the right of the sternum (breastbone).</td>
</tr>
<tr>
<td>V₂</td>
<td>In the fourth intercostal space (between ribs 4 and 5) just to the left of the sternum.</td>
</tr>
<tr>
<td>V₃</td>
<td>Between leads V₂ and V₄.</td>
</tr>
<tr>
<td>V₄</td>
<td>In the fifth intercostal space (between ribs 5 and 6) in the mid-clavicles line.</td>
</tr>
<tr>
<td>V₅</td>
<td>Horizontally even with V₄, in the left anterior axillary line.</td>
</tr>
<tr>
<td>V₆</td>
<td>Horizontally even with V₄ and V₅ in the midaxillary line.</td>
</tr>
</tbody>
</table>
The term "lead" in electrocardiography refers to the 12 different vectors along which the heart's depolarization is measured and recorded. There is a total of six limb leads and augmented limb leads arranged like spokes of a wheel in the coronal plane (vertical) and six precordial leads that lie on the perpendicular transverse plane (horizontal).

Each of these leads represents the electrical potential difference between two points. For each lead, the positive pole is one of the ten electrodes. In bipolar leads, the negative pole is a different one of the electrodes, while in unipolar leads, the negative pole is a composite pole known as Wilson's central terminal. Wilson's central terminal VW is produced by averaging the measurements from the electrodes RA, LA, and LL to give an average potential across the body. In a 12-lead ECG, all leads except the limb leads are unipolar (aVR, aVL, aVF, V1, V2, V3, V4, V5, and V6).

**Limb leads**

Leads I, II and III are called the limb leads. The electrodes that form these signals are located on the limbs—one on each arm and one on the left leg. The limb leads form the points of what is known as Einthoven's triangle.

- **Lead I** is the voltage between the (positive) left arm (LA) electrode and right arm (RA) electrode.
- **Lead II** is the voltage between the (positive) left leg (LL) electrode and the right arm (RA) electrode.
- **Lead III** is the voltage between the (positive) left leg (LL) electrode and the left arm (LA) electrode:
**Augmented limb leads**

Leads aVR, aVL, and aVF are the augmented limb leads. They are derived from the same three electrodes as leads I, II, and III, but they use Wilson's central terminal as their negative pole.

- Lead augmented vector right (aVR)' has the positive electrode on the right arm. The negative pole is a combination of the left arm electrode and the left leg electrode.
- Lead augmented vector left (aVL) has the positive electrode on the left arm. The negative pole is a combination of the right arm electrode and the left leg electrode.
- Lead augmented vector foot (aVF) has the positive electrode on the left leg. The negative pole is a combination of the right arm electrode and the left arm electrode.

Together with leads I, II, and III, augmented limb leads aVR, aVL, and aVF form the basis of the hex-axial reference system, which is used to calculate the heart's electrical axis in the frontal plane.

**Precordial leads**

The precordial leads lie in the transverse (horizontal) plane, perpendicular to the other six leads. The six precordial electrodes act as the positive poles for the six corresponding precordial leads: (V1, V2, V3, V4, V5 and V6). Wilson's central terminal is used as the negative pole.
4. Pre-Processing for Noise

ECG signal is often affected by several types of noise and artifacts. These noise and artifacts must be eliminated or at least significantly reduced before the data can be processed, otherwise the accuracy will suffer and error rate can become too high.

A. Noise and Artifacts

In medicine and biology, the term artifact refers to any component of a signal that is extraneous to the variable represented by the signal. Thus, random noise generated within the measuring instrument, electrical interference (including 60 Hz), and all other unwanted variations in the signal are considered artifacts. A major source of artifacts is movement of the subject, which in turn results in movement of the measuring device. Since many transducers are sensitive to movement, the movement of the subject often produces variations in the output signal.

B. Power line interferences

Power line interferences contains 60 Hz (North America) noise because of improper grounding. It is indicated as noisy baseline signals with frequency components at 60 Hz or its harmonics at integral multiples of the fundamental frequency as shown in Figure 5.

Figure 5: Power line interference in ECG Signal
Its frequency content is 60 Hz and its harmonics, amplitude can be up to 50 percent of peak-to-peak ECG signal amplitude. A high pass filter with cut-off frequency 60 Hz can be used to remove the power line interferences.

C. Baseline drift

Baseline drift may be caused in chest lead ECG signals by coughing or breathing with large movement of the chest or when an arm or leg is moved in the case of limb lead ECG acquisition as shown in Figure 6. Base-line drift can be sometimes caused by variations in temperature and bias in the instrumentation and amplifiers. Its frequency range generally bellows 0.5 Hz. The baseline wander is estimated by using first order zero phase low pass filter with cut-off frequency 0.5 Hz and is subtracted from the original signal.

![Figure 6: Baseline drift in ECG Signal](image)

D. Motion artifacts

Motion artifacts are transient baseline change due to electrode skin impedance with electrode motion. It can generate larger amplitude signal in ECG waveform. The peak amplitude of this artifact is 500 percent of Peak to Peak ECG amplitude and its duration is about 100 – 500ms. An adaptive filter can be used to remove the interference of motion artifacts. These noise and artifacts should be removed from the ECG by the above-mentioned pre-processing methods.
E. Pre-Processing

Several methods are used for pre-processing ECG Signals. Robustness of pre-processing procedures depends on application and how ECG signal is recorded. Single lead ECG signal obtained from wearable devices has less noise components when compared to signal obtained from electrical devices like Holter machines. Hence different types of pre-processing techniques for different applications are used. Almost all basic pre-processing techniques use filters to reduce noise components.

The simplest and most widely used noise reduction technique is recursive digital filters of the finite impulse response (FIR). This method is used when noise frequency is known beforehand, such as noise coming from electrical interference (50 Hz, 60 Hz). Since frequency is always known, this method is not helpful sometimes. High pass and Low pass filter usage can also be an option but it distorts morphology of signal, which is not good for feature detection in case of diagnosing cardiac diseases. Wavelet based multi-level decomposition technique can provide reliable results with noise reduction. But implementation is complex hence this technique is used in high level computing.

Since this work considers implementation in wearable devices, Butterworth filter of order 3 is used. Bandwidth of this bandpass filter is considered to have the cutoff frequencies of 0.5 Hz and 50 Hz [6], since most of the information of ECG signal is considered to be contained in this frequency range.
Pre-processing is an important step in detecting features of ECG signal accurately. Figures 7, 8 show the result of feature detection without and with filtering respectively.
5. Features Extraction

The feature extraction is a key stage in heart beat based AF classification. Any information taken from heartbeat which discriminates it with other type of beat is considered as a feature. Features can be taken from ECG signal morphology in time domain, frequency domain. Feature extraction/detection is different from feature selection as feature extraction is identifying different features from ECG signal whereas feature selection is selecting subset of features instead of all.

The main objective of the ECG features extraction process is to derive a set of parameters that best characterize the ECG signal. These parameters should contain maximum information about the ECG signal. Hence the selection of these parameters is an important criterion to be considered for proper classification. AF classification, therefore, involves determination of several characteristic features of the ECG signal.

Features extraction for Atrial Fibrillation

Every arrhythmia has similar features to be extracted from ECG. Figure 9 below shows missing definitive waveform ‘P’ with baseline due to atrial flutter being fired randomly without any rhythm. Duration between adjacent ‘R’ peaks i.e., RR intervals are irregular because of atrioventricular (AV node), receiving a bunch of electrical impulses from atria without rhythm, causing ventricles to contract rapidly.

Premature ventricular contraction (PVC) is another type of arrhythmia which shares similar features with Atrial Fibrillation. In PVC ventricles contracts twice as much faster than normal. It also misses ‘P’ wave an RR interval are inconsistent. But, both arrhythmias can be differentiated in frequency or time domain with interval features of other waves being considered.
Most useful methods which can differentiate features from one type of heart disease to other are HRV analysis [8], QRS complex, P wave morphology.

A. P Wave Morphology

Depolarization of atrial chamber generates P wave. Hence absence of P wave might indicate risk of AF. Parameters in P wave morphology which are important to detect AF are:

i. Amplitude of P-wave

ii. Width of P-wave

iii. Time distance between P peak and R peak.

iv. Amplitude between P and R peak.

Detecting P wave is challenging as the amplitude of P wave is relatively smaller than R wave and can be overshadowed by noise. Hence pre-processing plays an important role in detecting P wave.
B. QRS Morphology

In ECG signal, QRS complex has the highest peak amplitude (shown in Figure 10) and is easily detectable. Most popular QRS detection algorithm is Pan and Tompkins often called as Pan Tompkins algorithm [4]. This algorithm can detect QRS with 99.30% of accuracy. Implementation is simple as four filters are needed for detection.

i. Band pass filter

ii. Derivative filter

iii. Smoothing function

iv. Moving window integration

As in case of P wave morphology 3 parameters need to be extracted from QRS complex.

i. Amplitude of R peak

ii. Width of QRS complex

Figure 10: QRS Complex in ECG
QRS complex are important to identify arrhythmia in addition to it calculating RR intervals which is total duration between R peak and adjacent R peak is also an important feature.

C. Heart Rate Variability (HRV)

HRV Analysis gives information about cardiac condition, which cannot be visualized during manual inspection of ECG morphology. HRV analysis is one of the most important features in any type of arrhythmia detection. HRV analysis is no single valued feature, but it has several methods, like time domain analysis, frequency domain analysis and non-linear method of analysis.

Time domain analysis (Table 4) use RR intervals to get parameters for analysis. This is useful in case of PVC, AF. NN interval is time duration between one beat and another beat which can also be given by RR intervals (shown in Figure 11).

Figure 11: HRV in
Frequency domain analysis parameters (Table 5) are used to detect RR variation. Other feature extraction techniques are Power spectral density (PSD), Burg method, Higher order spectra (HOS), Short time Fourier transform (STFT), Continuous wavelet transform (CWT).

Non-linear method describes processes like, Recurrence plots, Sample entropy, Correlation Dimension analysis etc.

**Table 4:** Parameters of time domain analysis

<table>
<thead>
<tr>
<th><strong>Index</strong></th>
<th><strong>Definition</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>AVNN</td>
<td>Average of NN intervals</td>
</tr>
<tr>
<td>SDNN</td>
<td>Standard Deviation of NN intervals</td>
</tr>
<tr>
<td>SDANN</td>
<td>Standard Deviation of average of NN intervals</td>
</tr>
<tr>
<td>RMSSD</td>
<td>Root Mean Square standard deviation</td>
</tr>
<tr>
<td>NN50</td>
<td>Number of pairs of successive NN intervals differ more than 50ms</td>
</tr>
<tr>
<td>pNN50</td>
<td>Proportion of NN50 divided by total number of NNs</td>
</tr>
<tr>
<td>TINN</td>
<td>Triangular interpolation of NN interval histogram</td>
</tr>
</tbody>
</table>

**Table 5:** Parameters of frequency domain analysis

<table>
<thead>
<tr>
<th><strong>Index</strong></th>
<th><strong>Definition</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>ULF</td>
<td>0-0.0033 Hz</td>
</tr>
<tr>
<td>VLF</td>
<td>0.0033-0.04 Hz</td>
</tr>
<tr>
<td>LF</td>
<td>0.04-0.15 Hz</td>
</tr>
<tr>
<td>HF</td>
<td>0.15-0.40 Hz</td>
</tr>
<tr>
<td>LF/HF ratio</td>
<td>Ratio of low to high frequency power</td>
</tr>
</tbody>
</table>
D. Segmentation

Each ECG signal is segmented beat wise. Automatic methods are available for beat wise detection [13]. The main criteria for segmentation and identify the beat is ‘R’ peak. Since R peak detection is done reliably through Pan Tompkins algorithm, with proven accuracy of around 99% [4], R peak is taken as base for detecting other waves in ECG Signal.

Table 6: Criteria to extract P, Q, R, S, T, T’ from ECG

<table>
<thead>
<tr>
<th>Point</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>Minima in between 1/8 of RR interval before each R point to R point</td>
</tr>
<tr>
<td>S</td>
<td>Minima in between each R point to ¼ of RR ahead of each R point</td>
</tr>
<tr>
<td>P</td>
<td>Maxima from 3/8 of RR before each R point to Q point</td>
</tr>
<tr>
<td>T</td>
<td>Maxima after ¼ of RR interval of each R point to the 3/8 of RR of each R point</td>
</tr>
<tr>
<td>T’</td>
<td>Minima after ¼ of RR interval of each R point to the 3/8 of RR of each R point</td>
</tr>
</tbody>
</table>

Following steps are followed for Feature Extraction:

i. ‘R’ peak is detected using Pan Tompkins algorithm. Its amplitude and interval is recorded.

ii. Distance between adjacent R peaks is calculated and it termed as RR interval.

iii. Using Criteria given for S, Q waves, S and Q waves are identified, their respective indices are noted. By this time, QRS complex is detected.

iv. Using Criteria (Table 6) given for P, T waves, P and T waves are identified, their respective indices are noted. P wave morphology is identified by this time.

v. T’ is identified as given in criteria, its index is noted as well. T’ index is marked to be the end of each beat.
vi. Heart Rate Variability is calculated based on RR interval for each beat.

Intervals between each wave combination is also calculated. Amplitudes between each wave combination is also calculated. Features calculated are:

**Amplitude features:**

- P, Q, R, S, T, T', PQ, PS, PR, PT, PT', QR, QS, QT, QT', RS, RT, RT', ST, ST'.

**Interval features:**

- P, Q, R, S, T, T, PQ, PS, PR, PT, PT', QR, QS, QT, RS, RT, ST, TT', HRV.

20 amplitude and 18 interval features, HRV are calculated. In total 39 features are calculated and feature space with 39 elements for each row(beat) is formed.

**E. Beat Annotation**

ECG is a continuously varying signal and several factors determines how the shape of waveforms in ECG would be. An athlete will generally have higher heart rate, than normal individual whilst resting. Heart rates are higher while doing exercise and relatively slow while resting and sleeping. Hence any heart disease cannot be determined with in short time.

Heart diseases often show varying symptoms which might appear as other disease. Atrial Fibrillation when diagnosed at later stages show signs of severe damage in heart functioning which in turn shows sparse episodes of Ventricular flutter. Irregular heart beat shows episodes of Atrial Bigeminy. Slower heart rates show episodes of Sinus Bradycardia. These conditions are common in individual, but occurring with Atrial Fibrillation episodes, these conditions add severity to Atrial Fibrillation.
6 different types of beats are identified which add severity to Atrial Fibrillation. Each beat detected is put into any one of these beats which will be identified by classifier trained to differentiate between these beats.

i. **Atrial Bigeminy**

Bigeminy is a rhythm abnormality (Figure 12), where in rhythms continuously differs from long to short duration. Bigeminy occurs due to ectopic beats i.e., each beat occurs after sinus rhythm. 2 beats occur like twins. A normal sinus rhythm occurring with another premature beat. Premature beat when originated from ventricles chamber of heart is called Premature Ventricular contraction (PVC). When originated from Atrial chamber, it is called Atrial Bigeminy or Premature Atrial Contraction (PAC). For most of Atrial Fibrillation episodes, since it is irregular rhythm problem, Atrial Bigeminy succeeds or precedes Atrial Fibrillation episodes.

![Figure 12: Atrial Bigeminy rhythm](image)

Beat next to PVC experiences a brief pause which develops Bigeminy. A PVC wave front comes across a resistive AV node that does not conduct the wave front. Atrium not getting depolarized, does not reset the sinus node. The interval from PVC to p wave is prolonged which is almost equal to p-p interval.
When atrial rhythm is irregular, Bigeminy occurrence depends on length of p-p interval and happens more frequently with intervals of longer duration. In people without heart diseases and symptoms leading to heart diseases, Bigeminy, when appeared in ECG signal, does not need any treatment. Beta blockers can be used to suppress ventricular ectopy if it is symptomatic. Atrial Bigeminy, when occurred with other episodes like VT can turn more serious disorder.

ii. **Atrial Fibrillation**

To detect severity of Atrial Fibrillation it is definitive that Atrial Fibrillation beats are to be detected. It is the most common rhythm problem which starts at atrial chamber of heart. Different from the usual Sinus node directing the electrical rhythm, different impulses fires at once rapidly, causing irregular rhythm at atria. Due to the irregular rhythm atria cannot contract and squeeze blood efficiently to ventricles.

![Figure 13: Atrial fibrillation rhythm](image)

Figures 13, 14 show the difference between the rhythms in Atrial Fibrillation and Normal ECG signal.
Since atria is beating rapidly with irregular rhythm, blood does not flow through smoothly. This might cause a blood clot. A clot when pumped out of heart, travels to brain, eventually resulting in a stroke. People with Atrial Fibrillation are 5 to 7 times more prone to heart strokes than a normal person. Clots can also travel to other parts of body resulting in damage.

A part from causing clots, irregular heart rhythm may cause heart to function less efficiently, which eventually decreases heart pumping ability. This when occurred for a long time can significantly weaken heart functioning and can lead to heart failure.

There is no specific cause of occurrence of Atrial Fibrillation beats. It may occur randomly and can vanish by itself. But when Atrial Fibrillation is lasted for a longer time, it should be treated. Some of the most common causes of Atrial Fibrillation are after heart surgery, congenital heart disease, heart disease, hypertension, excessive caffeine intake, stress levels. Heart palpitations, lack of energy, chest discomfort, shortness of breath are some of the symptoms of Atrial fibrillation.

\[ Figure 14: \] Normal rhythm
iii. Normal rhythm

While detecting irregular heart rhythm, it is equally important to detect normal heart rhythm as well. Though not much weightage is given to normal heart rhythm in case of severity detection, it is important to distinguish between normal and abnormal beats.

Heart pumps blood to the rest of body, atrial chambers (upper) contract and followed by ventricular chambers (lower). These actions, when happened in synchronization, allows in efficient pumping of blood and in turn functioning of heart. Timing of contractions is guided by electrical system of heart.

Electrical impulse begins at SA node which is at the right atrium. This node adjusts the impulse depending on persons activity, like SA node increases rate of impulses during exercise and reduces it during sleep. When an impulse occurs, it spreads from right to left atria, causing them to contract and allow flow of blood into ventricles. This impulse travels to atrioventricular chamber. AV node is the only node which allows the impulse to travel from atria to ventricles. This impulse in turn travels through the walls of ventricles causing them to contract and pump blood out of heart. Right ventricle pumps blood to lungs and left ventricle to the body.

SA node directing electrical impulse in heart is called Normal sinus rhythm. Normal heart rate is 60-100 times per minute. Figure 14 shows Normal Sinus rhythm.

iv. Sinus Bradycardia

Bradycardia is a type of slow heartbeat. SA node, which fires electrical impulse 60-100 times at normal rhythm (Figure 15), fires the same at less than 60 times at bradycardia. Adults and children will have sinus bradycardia which does not show any symptoms. These cases, it is not harmful and does not mean there is a heart problem. Sinus bradycardia is very common in
young people, athletes, and during sleep. This is called physiologic sinus bradycardia. Heart may skip or drop beats in this case.

Sinus bradycardia is a case of concern when associated with other medical condition. Problems with SA node, advanced age, heart conditions existing by birth, obstructive sleep apnea, hyperthyroidism, drugs that effect SA node are few of the causes for Sinus Bradycardia.

Figure 15: Sinus Bradycardia rhythm vs Normal rhythm

Severe symptoms of Sinus bradycardia are dizziness, fainting. Using medicines that include beta blockers, calcium channel blockers also causes Sinus Bradycardia. Reducing such medication usage might reduce the condition as well. Sinus bradycardia when not having any symptoms, or not associated with heart diseases is not a case of concern but when associated with other heart diseases, should be looked for. Hence it is considered a case of concern when detected with atrial fibrillation.
v. *Supraventricular tachyarrhythmia*

Supra ventricular tachyarrhythmia (SVT) is an abnormal fast rhythm, arising from irregular electrical activity in upper part of heart (Figure 16). It starts from SA node as does normal heart beat.

Symptoms for SVT can arise and vanish by itself. Stress, exercise can result in increase in heart rate. Episodes typically last from days to weeks, and sometimes till getting treated. Heart rate may go up to 150-270 beats per minute.

SVT can be diagnosed by detecting QRS complex. Most SVT episodes have narrow QRS complex. Due to the abnormalities in electrical conduction, QRS complex sometimes will be wide, misguided to be ventricular tachycardia. Due to the irregularity in heart rate sometimes SVT is misguided to be Atrial fibrillation as well. In which case p wave morphology and QRS complex.

*Figure 16: Supraventricular tachyarrhythmia vs Normal rhythm*
vi. **Ventricular Tachycardia**

Ventricular tachycardia is a heart rhythm disorder caused by abnormal electrical signals arising from ventricular chambers (Figure 17). This causes the electrical impulse out of sync with atrial chamber and heart rate more than 100 beats per minute. When this happens, heart will not be able to pump enough blood to body and lungs.

Ventricular tachycardia is brief and typically last for few seconds. It is a very dangerous condition, when episodes last for few seconds may lead to the death of person. In some cases, it starts functioning of heart and thereby leading to sudden cardiac arrest, without showing any early symptoms. This severe condition mostly occurs in people with preexisting heart condition or have a heart attack previously.

*Figure 17: Ventricular tachycardia vs Normal rhythm*
6. Classification

Once feature extraction and beat annotation is done, a classifier model needs to be developed, which learns how to classify in between different beats and does the classification on test data. For such classifier, 3 classification algorithms are considered.

- Support Vector Machine
- Tree Bagger
- Random Forest

i. **Support Vector Machine (SVM)**

Support Vector Machine (SVM) is a Supervised machine learning algorithm. SVMs analyze data, recognize patterns and is used for classification and regression analysis. Given a set of training examples with 2 classes, SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

A Support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which is be used for classification, regression. An ideal separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the
generalization error of the classifier. Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function selected to suit the problem.

The hyper planes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is constant. The vectors defining the hyperplanes can be chosen to be linear combinations with parameters of images of feature vectors that occur in the data base. If becomes small as grows further away from, each term in the sum measures the degree of closeness of the test point to the corresponding data base point. In this way, the sum of kernels above can be used to measure the relative nearness of each test point to the data points originating in one or the other of the sets to be discriminated. Note the fact that the set of points mapped into any hyperplane can be quite convoluted as a result, allowing much more complex discrimination between sets which are not convex at all in the original space.

**Binary SVM**

The training data if linearly separable as shown in Figure 18, two hyperplanes can be selected in a way that they separate the data and there are no points between them, and then try to maximize their distance. The region bounded by them is called "the margin". These hyperplanes can be described by the equations and using geometry. Distance between these two hyperplanes is determined, such that it should be minimized to prevent data points from falling into the margin.
Various strategies used in multiclass classification are one-against-all (OAA), one-against-one (OAO) and the DAGSVM. The most popular ones are the one-against-all (OAA) and the one-against-one (OAO) strategies, good choice of Kernel helps in SVM becoming a multi class Classifier (Figure 19). RBF Kernel is used to test SVM in this work.

**Multi class SVM**

![Binary SVM](image18.png)

*Figure 18: Binary SVM*

![Multiclass SVM with 3 hyperplanes](image19.png)

*Figure 19: Multiclass SVM with 3 hyperplanes*
ii. **Tree Bagger**

Boot strap aggregating, in short, Bagging is an ensemble meta algorithm designed to improve stability and accuracy of machine learning algorithms. Usually Bagging is applied on weak learners to improve efficiency. It also reduces variance and helps avoiding overfitting.

Decision tree is one of the basic classification algorithms that uses tree like model for decisions on possible outcomes of an event. It is one way of displaying an algorithm that contains conditional control statements. Despite being simple to understand, implement and interpret, they are relatively unstable, inaccurate as many other predictors perform better with similar data.

Hence Bagging is applied on decision tree, which can be used for multi-level classification as by the name Decision stumps.

![Figure 20: Tree Bagging on Iris Dataset.](image-url)
Figure 20 is an example of implementation of tree bagging on Iris Dataset. Iris is a popular multi variate dataset often used as example in linear discriminant analysis. It has 3 different types of species differentiated by 4 different features.

**iii. Random Forest**

Random forests operate by constructing multiple decision trees during training and outputs the class which is the mode of classes as classifier, mean prediction as regression of individual trees. Random forest is much enhanced version of tree bagger thereby reducing problem of overfitting caused by tree bagging.

The main advantage of random forest in this project lies in its unique feature of randomly selecting features. As an example, for 100 training samples each with 20 features, instead of training the model with 100*20 feature space, for each training space, random forest creates a subspace of features randomly and train the model with that feature space. For the next iteration it again selects the feature space randomly and train the model with it. As it selects subspace randomly creating a forest like random feature spaces it is called Random Forest. Instead of searching for most important feature for splitting a node, best feature among the randomly selected subset of features is used to split the tree by creating node.

Random forest can also be used to rank the importance variables or features in regression or classifier. This is useful in case there are lot of variables for feature reduction. Since 40 features space is considered, feature reduction is not implemented in this project. Implementing Random forest is far less computationally expensive when compared to Multi SVM, which gives similar performance as Random forest.
**Severity Classification**

The Classifier model is trained based on labeled data, and classifies given ECG signal with respective classes. Here we propose a novel severity classification, determined on a scale of 0 to 1, using the below equation and weights as given in Table 7.

\[
P(S) = \frac{(w_{no} \times B_{no} + w_{af} \times B_{af} + w_{ab} \times B_{ab} + w_{sb} \times B_{sb} + w_{sv} \times B_{sv} + w_{vt} \times B_{vt})}{B_{total}}
\]

- \(P(S)\) : Probability of Severity
- \(w\) : Weights of beats
- \(B_x\) : Number of beats of label ‘x’
- \(B_{total}\) : Total number of beats

**Table 7**: Weights assigned to class coefficients

<table>
<thead>
<tr>
<th>Beats</th>
<th>Normal((w_{no}))</th>
<th>AF((w_{af}))</th>
<th>AB((w_{ab}))</th>
<th>SBR((w_{sb}))</th>
<th>SVTA((w_{sv}))</th>
<th>VT((w_{vt}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>0</td>
<td>0.5</td>
<td>0.25</td>
<td>0.25</td>
<td>0.5</td>
<td>1</td>
</tr>
</tbody>
</table>
7. Results

MIT BIH Atrial Fibrillation Database which has mostly Normal and Paroxysmal Atrial Fibrillation records and Long-Term Atrial Fibrillation Database which has both Paroxysmal and Sustained Atrial Fibrillation records are considered for downloading records. Total 16 records are downloaded from these Databases and are considered for testing.

Each record is pre-processed for noise reduction using Butterworth filter of order 3. Plots of signal before pre-processing is taken as raw signal. Plots from band pass filter, derivative filter, smoothing filter, moving average filter is taken to see how the signal is pre-processes at different stages. This is given in Figure 21.

![Figure 21: Result from Pan Tompkins algorithm](image)

After Pre-processing stage, Pan Tompkins algorithm is implemented to detect QRS complex, ‘R’ peak and respective index is identified for each beat. ‘R’ peaks and indices are basis to detect other features of the signal. Using criteria given in Table 6, P, Q, S, T, T’
amplitudes and indices are identified for each beat. These detected features are plotted in Figure 22.

![Figure 22: Result from feature detection](image)

Amplitude features, Interval features are determined and HRV feature is also calculated and each beat will have 39 features in total. 1458 beats from all the 16 records are considered and a feature space with 39 features is created for all the beats and is written to CSV file, with each row representing one beat of data. The CSV file will have $1458 \times 39$ feature space readied to go through classification.

Each beat (a row of CSV file) is labeled with 6 possible classes, (Normal, Atrial Bigeminy, Atrial Fibrillation, Sinus Bradycardia, Supraventricular tachyarrhythmia, Ventricular Tachycardia) from corresponding annotations given in Physionet, so that each row has a label which can be used to train the classifier. Labeled data is then given to Weka explorer to create models for each classifier SVM, Tree Bagging, Random Forest. Trained model is tested using
10-fold cross validation. Cross validation creates subsets of samples in numbers of n-fold, in this case 10 subsets are created and some subsets are used for training where as the remaining are used for testing the model created. The results with 10-fold cross validation are obtained as below.

**Table 9: Accuracy of data on testing for %split of training**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>SVM</th>
<th>Tree Bagging</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>84.84%</td>
<td>92.66%</td>
<td>96.78%</td>
</tr>
</tbody>
</table>

Results of accuracy from 10-fold cross validation suggests that the data is well classified by using Random Forest Algorithm. Hence Random Forest is used as Classification model. 70%, 80%, 90% of data is considered for training and the model is tested for on remaining data. Random Forest Classifier gave results of the same as in Table 9.

**Table 9: Accuracy of data on testing for %split of training**

<table>
<thead>
<tr>
<th>%split</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>94.9657</td>
<td>94.1781</td>
<td>94.5205</td>
</tr>
</tbody>
</table>

Once the classifier is trained, Weka gives a detailed result of data accurately classified against data mis classified. This result is taken as Confusion Matrix metric and is obtained for Random Classifier as given in Table 10. It can be observed that episodes having similar symptoms, are more often misclassified than other episodes (SVTA has same features as AB).
Table 10: Confusion matrix result from Random forest classifier

<table>
<thead>
<tr>
<th></th>
<th>AB</th>
<th>AF</th>
<th>N</th>
<th>SBR</th>
<th>SVTA</th>
<th>VT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>157</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>AF</td>
<td>0</td>
<td>307</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>0</td>
<td>1</td>
<td>284</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>SBR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>460</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SVTA</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>162</td>
<td>1</td>
</tr>
<tr>
<td>VT</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>41</td>
</tr>
</tbody>
</table>

Table 11: Precision, Recall values from Random forest classifier

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>0.963</td>
<td>0.908</td>
</tr>
<tr>
<td>AF</td>
<td>0.975</td>
<td>0.984</td>
</tr>
<tr>
<td>N</td>
<td>0.956</td>
<td>0.983</td>
</tr>
<tr>
<td>SBR</td>
<td>0.998</td>
<td>1</td>
</tr>
<tr>
<td>SVTA</td>
<td>0.905</td>
<td>0.915</td>
</tr>
<tr>
<td>VT</td>
<td>0.953</td>
<td>0.872</td>
</tr>
</tbody>
</table>

Metrics like precision, recall values gives information on sensitivity and specificity of Classifier model. Average precision and recall values are obtained to be 0.968 and 0.968
respectively. Class wise precision and recall values for Random Forest Classifier are obtained as given in Table 1.

Sample data is created by considering 60 beats per signal, and different episodes are assumed to be present in each signal as given in Table 12. Severity is calculated based on proposed model and the results are tabulated in Severity column of Table 12. A plot as shown in Figure 23, is taken with X-axis taking data of Patients name i.e., Patient 1, Patient 2 to Patient 11, and Y-axis taking the data of patient’s respective severity calculated in Table 12. This plot shows that patient with a greater number of VT beats has highest severity i.e., Patient 11 and patient with a greater number of normal beats has the lowest severity i.e., Patient 10.

**Table 12: AF Severity for 11 Patients**

<table>
<thead>
<tr>
<th></th>
<th>AB</th>
<th>AF</th>
<th>N</th>
<th>SBR</th>
<th>SVTA</th>
<th>VT</th>
<th>Severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person 1</td>
<td>5</td>
<td>20</td>
<td>0</td>
<td>10</td>
<td>20</td>
<td>5</td>
<td>0.48</td>
</tr>
<tr>
<td>Person 2</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>5</td>
<td>0.541</td>
</tr>
<tr>
<td>Person 3</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>Person 4</td>
<td>10</td>
<td>20</td>
<td>0</td>
<td>8</td>
<td>20</td>
<td>2</td>
<td>0.45</td>
</tr>
<tr>
<td>Person 5</td>
<td>10</td>
<td>20</td>
<td>0</td>
<td>17</td>
<td>10</td>
<td>3</td>
<td>0.425</td>
</tr>
<tr>
<td>Person 6</td>
<td>13</td>
<td>15</td>
<td>0</td>
<td>15</td>
<td>15</td>
<td>2</td>
<td>0.40833</td>
</tr>
<tr>
<td>Person 7</td>
<td>14</td>
<td>10</td>
<td>10</td>
<td>15</td>
<td>10</td>
<td>1</td>
<td>0.30833</td>
</tr>
<tr>
<td>Person 8</td>
<td>9</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>0.2667</td>
</tr>
<tr>
<td>Person 9</td>
<td>10</td>
<td>10</td>
<td>30</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0.1875</td>
</tr>
<tr>
<td>Person 10</td>
<td>5</td>
<td>2</td>
<td>50</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.054</td>
</tr>
<tr>
<td>Person 11</td>
<td>0</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0.625</td>
</tr>
</tbody>
</table>
Figure 23: Plot for Severity of 11 patients
8. Conclusion

A. Key Findings

Atrial Fibrillation is the most common heart disease. People with age more than 60 are more prone to this disease. Short term Atrial Fibrillation is not a concern for Quality of Life (QoL) of patients. However, long term Atrial Fibrillation and Atrial Fibrillation associated with other heart diseases effect QoL of patients very badly. Heart functioning deteriorates without proper medication at early stages. This may eventually reduce life span of patient. Hence there is a need for early detection.

Wearable devices are most common in modern world and health monitoring functions like, step tracking, HRV, Heart rate detection, fall detection are being imbibed into wearables to enable consumer aware of some important aspects of health. Recently release Apple watch series 4 (Nov. 2018) has ability to detect Atrial Fibrillation using a single lead ECG sensor located on bottom of the dial connecting to wrist. It can report anomaly directly to physician as well. Such health monitoring avoids delay in detecting diseases despite not very often hospital visits.

We proposed a novel, easy to implement method of detecting severity class of Atrial fibrillation, which can be adaptive for wearable devices. This method is computationally less expensive to implement as well. This prevents patient from having huge medical costs at later stages of disease, which can be cured by medication at an earlier stage. This method can be used to implement not only in wearable device but also at places where detection of ECG signal is possible with computational ability. In short, results are encouraging with 96.7% accuracy obtained from classifier and equally better precision and recall values, being at 0.968.
The work done as part of this project is accepted to be published in IEEE EIT 2019 Conference and below is the citation for the same.


B. Future implementation

In proposing novel approach for severity classification, we considered a linear first order equation to calculate probability of severity as a first attempt. This can be improved by considering more complex equation as in exponential variable, equation of order 2 or higher.

As of now 96.7% of accuracy has been achieved which is a modest result to implement in wearable devices. This can be improved by taking more precise data, annotations happening under thorough supervision. Also, data needs to be considered from all forms of age group, segregated by other preexisting conditions if any.

Severity class can be determined more precisely, when frequency of classes is also determined and is incorporated into the equation to determine severity. Frequency of class as in based on a certain time window, how many times each class of beat if occurring. As more frequent episodes of Atrial fibrillation, for example, can be more severe when compared to episodes occurred in length for once and vanished.

This model can be further improved by taking real time ECG signal samples, calculating the severity and validating the data under supervision of physician by verifying each beat being correctly classified and severity matches the physician prediction. As there are no existing
methods to validate the calculated severity, validation by physician can give better insight of working model of algorithm to avoid false positives.
References:


Appendix

The main Program that runs the other functions:

clc, close all, clear all;

%%Loading ECG Signal
ecg1=load('afdb_08219.mat');
%ecg1=load('08405_10sm.mat');
ecg2=struct2cell(ecg1);
ecg3=cell2mat(ecg2);
%For Changing Columns data to rows Data
ecg=ecg3. ;
ecg_signal=ecg(1,:);
%fs=input('Sampling Frequency ');
fs=250;
gr=0;

%Perform Feature extraction
[amp_features, int_features, ecg_h, rr_int, r_ind]=All_Features(ecg_signal ,fs, gr);

%Writing Data to CSV
features=[amp_features, int_features];
features1=array2table(features);
%Delete file
%delete ECGData.csv
%First time write
writetable(features1, 'afdb_08219.csv')
%dlmwrite('array_write_203.csv',features1)
%Append Data after first time
%dlmwrite('ECGData.csv',features,'-append')
Program to perform Feature Extraction:

function [amp_features, int_features, ecg_n, rr_int, r_ind] = All_Features(ecg_signal, fs, gr)

ecg = ecg_signal;
NoS = length(ecg_signal) % Number of Samples

%% function [qrs_amp_raw, qrs_i_raw, delay] = pan_tompkin(ecg, fs)
% Complete implementation of Pan-Tompkins algorithm

%% Inputs
% ecg : raw ecg vector signal 1d signal
% fs : sampling frequency e.g. 200Hz, 400Hz and etc
% gr : flag to plot or not plot (set it 1 to have a plot or set it zero not
% to see any plots
%% Outputs
% qrs_amp_raw : amplitude of R waves amplitudes
% qrs_i_raw : index of R waves
% delay : number of samples which the signal is delayed due to the filtering
%% Method
% See Ref and supporting documents on researchgate.


%% References :

%% ---------------- License -----------------------------
%%
%% THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
if ~isvector(ecg)
    error('ecg must be a row or column vector');
end
if nargin < 3
    gr = 1;  % on default the function always plots
end
ecg = ecg(:); % vectorize

% ============ Noise cancelation(Filtering)( 5-15 Hz) ===============

% It has come to my attention the original filter doesnt achieve 12 Hz
% b = [1 0 0 0 0 0 -2 0 0 0 0 0 1];
% a = [1 -2 1];
% ecg_l = filter(b,a,ecg);
% delay = 6;

if fs == 200
    %------------------ remove the mean of Signal -----------------------
    ecg = ecg - mean(ecg);
    %== Low Pass Filter  H(z) = ((1 - z^(-6))^2)/(1 - z^(-1))^2 ==
    %It has come to my attention the original filter doesnt achieve 12 Hz
    % b = [1 0 0 0 0 0 -2 0 0 0 0 0 1];
    % a = [1 -2 1];
    % ecg_l = filter(b,a,ecg);
    % delay = 6;
    Wn = 12*2/fs;
    N = 3;
    % order of 3 less processing
    [a,b] = butter(N,Wn,'low');
    % bandpass filtering
    ecg_l = filtfilt(a,b,ecg);
    ecg_l = ecg_l/ max(abs(ecg_l));
% ============ Noise cancelation(Filtering)( 5-15 Hz) ===============
ax(1) = subplot(321); plot(ecg); axis tight; title('Raw signal');
ax(2) = subplot(322); plot(ecg_l); axis tight; title('Low pass filtered');
end

%% ===== High Pass filter H(z) = (-1+32z^(-16)+z^(-32))/(1+z^(-1)) =====
%%
%% It has come to my attention the original filter doesn achieve 5 Hz
%%
%% b = zeros(1,33);
%% b(1) = -1; b(17) = 32; b(33) = 1;
%% a = [1 1];
%% ecg_h = filter(b,a,ecg_l); % Without Delay
%% delay = delay + 16;

Wn = 5*2/fs;
N = 3;
% order of 3 less processing
[a,b] = butter(N,Wn,'high');
% bandpass filtering
ecg_h = filtfilt(a,b,ecg_l);
ecg_h = ecg_h/ max(abs(ecg_h));
if gr
    ax(3)=subplot(323); plot(ecg_h); axis tight; title('High Pass Filtered');
end
else
% bandpass filter for Noise cancelation of other sampling frequencies(Filtering)
f1=5;
% cutoff low frequency to get rid of baseline wander
f2=15;
% cutoff frequency to discard high frequency noise
Wn=[f1 f2]*2/fs;
% cutt off based on fs
N = 3;
% order of 3 less processing
[a,b] = butter(N,Wn);
% bandpass filtering
ecg_h = filtfilt(a,b,ecg);
ecg_h = ecg_h/ max( abs(ecg_h));
if gr
    ax(1) = subplot(3,2,[1 2]); plot(ecg); axis tight; title('Raw Signal');
    ax(3)=subplot(323); plot(ecg_h); axis tight; title('Band Pass Filtered');
end
end

% ------------------------ derivative filter ------------------------
%
% H(z) = (1/8T)(-z^(-2) - 2z^(-1) + 2z + z^2) %
if fs ~= 200
    int_c = (5-1)/(fs*1/40);
b = interp1(1:5,[1 2 0 -2 -1].*(1/8)*fs,1:int_c:5);
else
\[ b = [1 2 0 -2 -1].*(1/8)*fs; \]

\[ \text{ecg}_d = \text{filtfilt}(b,1,\text{ecg}_h); \]
\[ \text{ecg}_d = \text{ecg}_d/\max(\text{ecg}_d); \]

\% if gr
\% \hspace{1em} \text{ax}(4) = \text{subplot}(324); \text{plot}(\text{ecg}_d);
\% \hspace{1em} \text{axis} \text{ tight};
\% \hspace{1em} \text{title}('Filtered with the derivative filter');
\% end

\% \hspace{1em} \text{% Squaring nonlinearly enhance the dominant peaks}
\% \hspace{1em} \text{ecg}_s = \text{ecg}_d.^2;
\% if gr
\% \hspace{1em} \text{ax}(5) = \text{subplot}(325);
\% \hspace{1em} \text{plot}(\text{ecg}_s);
\% \hspace{1em} \text{axis} \text{ tight};
\% \hspace{1em} \text{title}('Squared');
\% end

\% \hspace{1em} \text{% Moving average}
\% \hspace{1em} \text{ecg}_m = \text{conv}(\text{ecg}_s, \text{ones}(1, \text{round}(0.150*fs))/\text{round}(0.150*fs));
\% \hspace{1em} \text{delay} = \text{delay} + \text{round}(0.150*fs)/2;

\% if gr
\% \hspace{1em} \text{ax}(6) = \text{subplot}(326); \text{plot}(\text{ecg}_m);
\% \hspace{1em} \text{axis} \text{ tight};
\% \hspace{1em} \text{title}('Averaged with 30 samples length, Black noise, Green Adaptive Threshold, RED Sig Level, Red circles QRS adaptive threshold');
\% \hspace{1em} \text{axis} \text{ tight};
\% end

\% \hspace{1em} \text{% Fiducial Marks}
\% \hspace{1em} \text{% Note: a minimum distance of 40 samples is considered between each R wave}
\% \hspace{1em} \text{% since in physiological point of view no RR wave can occur in less than}
\% \hspace{1em} \text{% 200 msec distance}
\% \hspace{1em} \text{[pks,locs] = findpeaks(\text{ecg}_m, 'MINPEAKDISTANCE', \text{round}(0.2*fs));}
\% \hspace{1em} \text{% Initialize Some Other Parameters}
\% \hspace{1em} \text{\text{LLp} = length(pks);}
\% \hspace{1em} \text{% Stores QRS wrt Sig and Filtered Sig}
\% \hspace{1em} \text{qrs_c = \text{zeros}(1, \text{LLp}); \text{ \% amplitude of R}}
\% \hspace{1em} \text{qrs_i = \text{zeros}(1, \text{LLp}); \text{ \% index}}
\% \hspace{1em} \text{qrs_i_raw = \text{zeros}(1, \text{LLp}); \text{ \% amplitude of R}}
\% \hspace{1em} \text{qrs_amp_raw = \text{zeros}(1, \text{LLp}); \text{ \% Index}
% ------------------- Noise Buffers -------------------------------
nois_c = zeros(1,LLp);
nois_i = zeros(1,LLp);
% ------------------- Buffers for Signal and Noise -------------------
SIGL_buf = zeros(1,LLp);
NOISL_buf = zeros(1,LLp);
SIGL_buf1 = zeros(1,LLp);
NOISL_buf1 = zeros(1,LLp);
THRS_buf1 = zeros(1,LLp);
THRS_buf = zeros(1,LLp);

%% initialize the training phase (2 seconds of the signal) to
determine the THR_SIG and THR_NOISE
THR_SIG = max(ecg_m(1:2*fs))*1/3;
% 0.25 of the max amplitude
THR_NOISE = mean(ecg_m(1:2*fs))*1/2;
% 0.5 of the mean signal is considered to be noise
SIG_LEV= THR_SIG;
NOISE_LEV = THR_NOISE;

%% Initialize bandpath filter threshold(2 seconds of the bandpass
signal)
THR_SIG1 = max(ecg_h(1:2*fs))*1/3;
% 0.25 of the max amplitude
THR_NOISE1 = mean(ecg_h(1:2*fs))*1/2;
SIG_LEV1 = THR_SIG1;
NOISE_LEV1 = THR_NOISE1;

%% ============ Thresholding and decision rule ============= %
Beat_C = 0;
% Raw Beats
Beat_C1 = 0;
% Filtered Beats
Noise_Count = 0;
% Noise Counter
for i = 1 : LLp
    % locate the corresponding peak in the filtered signal ===
    if locs(i)-round(0.150*fs)>= 1 && locs(i)<= length(ecg_h)
        [y_i,x_i] = max(ecg_h(locs(i)-round(0.150*fs):locs(i)));
    else
        if i == 1
            [y_i,x_i] = max(ecg_h(1:locs(i)));
            ser_back = 1;
        elseif locs(i)>= length(ecg_h)
            [y_i,x_i] = max(ecg_h(locs(i)-round(0.150*fs):end));
        end
    end
    % update the heart_rate ============= %
if Beat_C >= 9
    diffRR = diff(qrs_i(Beat_C-8:Beat_C));
% calculate RR interval
    mean_RR = mean(diffRR);
% calculate the mean of 8 previous R waves interval
    comp = qrs_i(Beat_C) - qrs_i(Beat_C-1);
% latest RR
    if comp <= 0.92*mean_RR || comp >= 1.16*mean_RR
        % ------ lower down thresholds to detect better in MVI ------ %
        THR_SIG = 0.5*(THR_SIG);
        THR_SIG1 = 0.5*(THR_SIG1);
    else
        m_selected_RR = mean_RR;
% The latest regular beats mean
    end
end

%% == calculate the mean last 8 R waves to ensure that QRS is not
==== %
if m_selected_RR
    test_m = m_selected_RR;
% if the regular RR available use it
elseif mean_RR && m_selected_RR == 0
    test_m = mean_RR;
else
    test_m = 0;
end
if test_m
    if (locs(i) - qrs_i(Beat_C)) >= round(1.66*test_m)
% it shows a QRS is missed
        [pks_temp,locs_temp] = max(ecg_m(qrs_i(Beat_C)+
round(0.200*fs):locs(i)-round(0.200*fs))); % search back and locate
        the max in this interval
        locs_temp = qrs_i(Beat_C)+ round(0.200*fs) + locs_temp -
1; % location

        if pks_temp > THR_NOISE
            Beat_C = Beat_C + 1;
            qrs_c(Beat_C) = pks_temp;
            qrs_i(Beat_C) = locs_temp;
            % ------------------ Locate in Filtered Sig ------------------ %
            if locs_temp <= length(ecg_h)
                [y_i_t,x_i_t] = max(ecg_h(locs_temp-
round(0.150*fs):locs_temp));
            else
                [y_i_t,x_i_t] = max(ecg_h(locs_temp-
round(0.150*fs):end));
            end
        end
    end
% ---------------- Band pass Sig Threshold -------------------

if y_i_t > THR_NOISE1
    Beat_C1 = Beat_C1 + 1;
    qrs_i_raw(Beat_C1) = locs_temp-round(0.150*fs)+(x_i_t-1);% save index of bandpass
    qrs_amp_raw(Beat_C1) = y_i_t;
    % save amplitude of bandpass
    SIG_LEV1 = 0.25*y_i_t + 0.75*SIG_LEV1;
    % when found with the second thres
end

not_nois = 1;
SIG_LEV = 0.25*pks_temp + 0.75*SIG_LEV;
% when found with the second threshold
else
    not_nois = 0;
end

%%% ================ find noise and QRS peaks
================== %

if pks(i) >= THR_SIG
    % ------- if No QRS in 360ms of the previous QRS See if T wave --
    %
    if Beat_C >= 3
        if (locs(i)-qrs_i(Beat_C)) <= round(0.3600*fs)
            Slope1 = mean(diff(ecg_m(locs(i)-
            round(0.075*fs):locs(i)))); % mean slope of the waveform at that position
        else
            Slope1 = mean(diff(ecg_m(qrs_i(Beat_C)-
            round(0.075*fs):qrs_i(Beat_C)))); % mean slope of previous R wave
        end
        % slope less then 0.5 of previous R
        Noise_Count = Noise_Count + 1;
        nois_c(Noise_Count) = pks(i);
        nois_i(Noise_Count) = locs(i);
        skip = 1;
        % T wave identification
        % ------- adjust noise levels ------- %
        NOISE_LEV1 = 0.125*y_i + 0.875*NOISE_LEV1;
        NOISE_LEV = 0.125*pks(i) + 0.875*NOISE_LEV;
        else
            skip = 0;
    end
end

%%% =========== skip is 1 when a T wave is detected =========

if skip == 0
Beat_C = Beat_C + 1;
qrs_c(Beat_C) = pks(i);
qrs_i(Beat_C) = locs(i);

%------------------- bandpass filter check threshold ------------------
-- %
if y_i >= THR_SIG1
    Beat_C1 = Beat_C1 + 1;
    if ser_back
        qrs_i_raw(Beat_C1) = x_i;
    end
else
    qrs_i_raw(Beat_C1) = locs(i)-round(0.150*fs)+(x_i - 1); % save index of bandpass
end
qrs_amp_raw(Beat_C1) = y_i;
% save amplitude of bandpass
SIG_LEV1 = 0.125*y_i + 0.875*SIG_LEV1;
% adjust threshold for bandpass filtered sig
end
SIG_LEV = 0.125*pks(i) + 0.875*SIG_LEV;
% adjust Signal level
end

elseif (THR_NOISE <= pks(i)) && (pks(i) < THR_SIG)
    NOISE_LEV1 = 0.125*y_i + 0.875*NOISE_LEV1;
    % adjust Noise level in filtered sig
    NOISE_LEV = 0.125*pks(i) + 0.875*NOISE_LEV;
    % adjust Noise level in MVI
elseif pks(i) < THR_NOISE
    Noise_Count = Noise_Count + 1;
    nois_c(Noise_Count) = pks(i);
    nois_i(Noise_Count) = locs(i);
    NOISE_LEV1 = 0.125*y_i + 0.875*NOISE_LEV1;
    % noise level in filtered signal
    NOISE_LEV = 0.125*pks(i) + 0.875*NOISE_LEV;
    % adjust Noise level in MVI
end

%% ------------------- adjust the threshold with SNR ----------------------
%%
if NOISE_LEV ~ 0 || SIG_LEV ~ 0
    THR_SIG = NOISE_LEV = 0.25*(abs(SIG_LEV - NOISE_LEV));
    THR_NOISE = 0.5*(THR_SIG);
end

%------ adjust the threshold with SNR for bandpassed signal ------
-- %
if NOISE_LEV1 ~ 0 || SIG_LEV1 ~ 0
    THR_SIG1 = NOISE_LEV1 = 0.25*(abs(SIG_LEV1 - NOISE_LEV1));
    THR_NOISE1 = 0.5*(THR_SIG1);
end

61
% take a track of thresholds of smoothed signal
SIGL_buf(i) = SIG_LEV;
NOISL_buf(i) = NOISE_LEV;
THRS_buf(i) = THR_SIG;

% take a track of thresholds of filtered signal
SIGL_buf1(i) = SIG_LEV1;
NOISL_buf1(i) = NOISE_LEV1;
THRS_buf1(i) = THR_SIG1;

% reset parameters
skip = 0;
not_nois = 0;
ser_back = 0;
end

%% Adjust Lengths
qrs_i_raw = qrs_i_raw(1:Beat_C1);
qrs_amp_raw = qrs_amp_raw(1:Beat_C1);
qrs_c = qrs_c(1:Beat_C);
qrs_i = qrs_i(1:Beat_C);

% Other Amplitudes
r_ind=qrs_i_raw;
r_amp=qrs_amp_raw;

count=length(r_ind);
for i=2:count
    rr_int(i)=r_ind(i)-r_ind(i-1);
end
    rr_int(1)=round(mean(rr_int(2:end)));

% Sample Entropy of rr interval

% rr_std=std(r_ind);
% for i=1:count
%     saen(i) = SampEn(2, 0.2*rr_std,rr_int, 1);
% end
%saen(1)=mean(saen(2:3));
rr_1by8(i)=round(0.125*rr_int(i));
rr_1by4(i)=round(0.25*rr_int(i));
rr_3by8(i)=round(0.375*rr_int(i));
end

% rr_1by8
% rr_1by4
% rr_3by8

for i=1:count
    for_q(i)=(r_ind(i)-rr_1by8(i));
    for_s(i)=(r_ind(i)+rr_1by4(i));
end

for_q(for_q<0)=1;
for_s(for_s>NoS)=NoS;

for i=1:count
    [q_amp(i), q_i(i)]=min(ecg_h(for_q(i):r_ind(i)));
    [s_amp(i), s_i(i)]=min(ecg_h(r_ind(i):for_s(i)));
end

for i=1:count
    q_ind(i)=abs(for_q(i)+q_i(i));
    s_ind(i)=abs(r_ind(i)+s_i(i));
end

% q_ind
% s_ind

%-------For P and T-------

for i=1:count
    for_p(i)=(r_ind(i)-rr_3by8(i));
    for_t(i)=(s_ind(i)+rr_3by8(i));
    left_t1(i)=(r_ind(i)+rr_1by4(i));
    right_t1(i)=(r_ind(i)+rr_3by8(i));
end

% for_t
for_p(for_p<0)=1;
for_t(for_t>NoS)=NoS;
s_ind(s_ind>NoS)=NoS;
left_t1(left_t1>NoS)=NoS;
right_t1(right_t1>NoS)=NoS;

for i=1:count
    [p_amp(i), p_i(i)]=max(ecg_h(for_p(i):q_ind(i)));
    [t_amp(i), t_i(i)]=max(ecg_h(s_ind(i):for_t(i)));
% For T1
[t1_amp(i), t1_i(i)] = min(ecg_h(left_t1(i):right_t1(i)));
end
for i=1:count
  p_ind(i)=abs(for_p(i)+p_i(i));
  t_ind(i)=abs(s_ind(i)+t_i(i));
  t1_ind(i)=abs(left_t1(i)+t1_i(i));
end

t_ind(t_ind>NoS)=NoS;
t1_ind(t1_ind>NoS)=NoS;
s_ind(s_ind>NoS)=NoS;

% p_t=p_ind/fs
% q_t=q_ind/fs;
% r_t=r_ind/fs;
% s_t=s_ind/fs;
% t_t=t_ind/fs;
% t1_t=t1_ind/fs;

% Amplitudes between adjacent Peaks
for i=1:count
  pq_amp(i)= p_amp(i)+q_amp(i);
  pr_amp(i)= p_amp(i)+r_amp(i);
  ps_amp(i)= p_amp(i)+s_amp(i);
  pt_amp(i)= p_amp(i)+t_amp(i);
  qr_amp(i)= q_amp(i)+r_amp(i);
  qs_amp(i)= q_amp(i)+s_amp(i);
  qt_amp(i)= q_amp(i)+t_amp(i);
  rs_amp(i)= r_amp(i)+s_amp(i);
  rt_amp(i)= r_amp(i)+t_amp(i);
  st_amp(i)= s_amp(i)+t_amp(i);
  pt1_amp(i)= p_amp(i)+t1_amp(i);
  qt1_amp(i)= q_amp(i)+t1_amp(i);
  rt1_amp(i)= r_amp(i)+t1_amp(i);
  st1_amp(i)= s_amp(i)+t1_amp(i);
end

% % intervals between adjacent peaks
for i=1:count
  pq_ind(i)= q_ind(i)-p_ind(i);
  pr_ind(i)= r_ind(i)-p_ind(i);
  ps_ind(i)= s_ind(i)-p_ind(i);
  pt_ind(i)= t_ind(i)-p_ind(i);
  qr_ind(i)= r_ind(i)-q_ind(i);
  qs_ind(i)= s_ind(i)-q_ind(i);
qt_ind(i)= t_ind(i)-q_ind(i);
rs_ind(i)= s_ind(i)-r_ind(i);
rt_ind(i)= t_ind(i)-r_ind(i);
st_ind(i)= t_ind(i)-s_ind(i);
ttl1_ind(i)=t1_ind(i)-t_ind(i);
ptl1_ind(i)=t1_ind(i)-p_ind(i); %Duration of total Beat from P-T1

end

% p_ind
% q_ind
% r_ind
% s_ind
% t_ind
% t1_ind

%Vectorization
p_ind=p_ind(:); q_ind=q_ind(:); r_ind=r_ind(:); s_ind=s_ind(:);
t_ind=t_ind(:); pq_ind=pq_ind(:); pr_ind=pr_ind(:); ps_ind=ps_ind(:);
pt_ind=pt_ind(:);
qr_ind=qt_ind(:);qs_ind=qs_ind(:);qt_ind=qt_ind(:);rs_ind=rs_ind(:);
rt_ind=rt_ind(:); st_ind=st_ind(:); t1_ind=t1_ind(:);
ttl1_ind=tt1_ind(:); ptl1_ind=ptl1_ind(:);

p_a=p_amp(:); q_a=q_amp(:); r_a=r_amp(:); s_a=s_amp(:); t_a=t_amp(:);
tl_a=t1_amp(:); pq_a=pq_amp(:); pr_a=pr_amp(:);
ps_a=ps_amp(:); pt_a=pt_amp(:); qr_a=qr_amp(:); qs_a=qs_amp(:);
qt_a=qt_amp(:); rs_a=rs_amp(:); rt_a=rt_amp(:); st_a=st_amp(:);
ptl_a=ptl1_amp(:); qt1_a= qt1_amp(:); rt1_a= rt1_amp(:);
stl_a=stl1_amp(:);

%rr_time=[rr_int]/fs;
hrv1 = diff(rr_int);
hrv(2:length(rr_int))=hrv1;
hrv(1)= hrv1(1);
hrv=hrv(:);

amp_features=[p_a, q_a, r_a, s_a, t_a, tl_a, pq_a, pr_a, ps_a, pt_a,
qr_a, qs_a, qt_a, rs_a, rt_a, st_a, ptl_a, qt1_a, rt1_a, stl_a];
int_features=[p_ind, q_ind, r_ind, s_ind, t_ind, t1_ind, pq_ind,
pr_ind, ps_ind, pt_ind, qr_ind, qs_ind, qt_ind, rs_ind, rt_ind,
st_ind, ttl1_ind, ptl1_ind, hrv];

%features=[amp_features, int_features];

t=(1:1:NoS)/fs;
s=1:NoS;
figure;
plot(s,ecg_h,p_ind,p_amp,'*g',q_ind,q_amp,'^k',r_ind,r_amp,'ob',s_ind,
s_amp,'+b',t_ind,t_amp,'+r',t1_ind,t1_amp,'*r');
delay

%%-------------------End of Program-------------------%
end