Classification of ECG Beats Using Machine Learning Technique

A Dissertation submitted towards the partial fulfilment of the requirement for the award of degree of

Master of Technology
in
Signal Processing & Digital Design

Submitted by
Nishant Kumar Nirala
2K15/SPD/10

Under the supervision of

Sh. M. S. Choudhry
(Associate Professor, Department of ECE)

Department of Electronics & Communication Engineering
Delhi Technological University
(Formerly Delhi College of Engineering)
Delhi-110042

2015-2017

DELHI TECHNOLOGICAL UNIVERSITY

Established by Govt. Of Delhi vide Act 6 of 2009

(Formerly Delhi College of Engineering)

SHAHBAD DAULATPUR, BAWANA ROAD, DELHI-110042

CERTIFICATE

This is to certify that the dissertation title "**Classification of ECG beats using machine learning Technique**" submitted by **Mr. NISHANT KUMAR NIRALA, Roll. No. 2K15/SPD/10**, in partial fulfilment for the award of degree of Master of Technology in "**Signal Processing and Digital Design (SPDD)",** run by Department of Electronics & Communication Engineering in Delhi Technological University during the year 2015-2017, is a bonafide record of student's own work carried out by him under my supervision and guidance in the academic session 2016-17. To the best of my belief and knowledge the matter embodied in dissertation has not been submitted for the award of any other degree or certificate in this or any other university or institute.

Dr. S. Indu

Head of Department

Electronics and Communication Dept.

Delhi Technological University

Delhi-110042

Sh. M.S.CHOUDRY

Supervisor

Associate Professor (ECE)

Delhi Technological University

Delhi-110042

DECLARATION

I hereby declare that all the information in this document has been obtained and presented in accordance with academic rules and ethical conduct. This report is my own work to the best of my belief and knowledge. I have fully cited all material by others which I have used in my work. It is being submitted for the degree of Master of Technology in Signal Processing & Digital Design at the Delhi Technological University. To the best of my belief and knowledge it has not been submitted before for any degree or examination in any other university.

Nishant Kumar Nirala

M. Tech. (SPDD)

2K15/SPD/10

Date: ____JULY, 2017

Place: Delhi Technological University, Delhi

ACKNOWLEDGEMENT

I owe my gratitude to all the people who have helped me in this dissertation work and who have made my postgraduate college experience one of the most special periods of my life.

Firstly, I would like to express my deepest gratitude to my supervisor **Sh. M. S. Choudhry**, Associate Professor (ECE) for his invaluable support, guidance, motivation and encouragement throughout the period during which this work was carried out.

I also wish to express my heart full thanks to my classmates as well as staff at Department of Electronics & Communication Engineering of Delhi Technological University for their goodwill and support that helped me a lot in successful completion of this project.

Finally, I want to thank my parents, family and friends for always believing in my abilities and showering their invaluable love and support.

Nishant Kumar Nirala

M. Tech. (SPDD)

2K15/SPD/10

ABSTRACT

According to the World Health Organization, cardiovascular diseases (CVD) are the main cause of death worldwide. An Estimated 17.5 million people died from CVD in 2012, 31% of all Representing Global deaths. The electrocardiogram (ECG) is a core tool for the pre-diagnosis of heart diseases. Many advances on ECG arrhythmia classification have Been developed in the last century; however, there is still research to identify malignant ECG waveforms on beats. The SVC complexes are known to be associated ventricular arrhythmias with malignant and in sudden cardiac death (SCD) cases. This Kind of detecting arrhythmia has been crucial in clinical applications. In this work, we extracted from 108,653 , 80 different features of the ECG beats classified MIT-BIH database in order to classify the Normal, SVC and other kind of ECG beats.

The main aim of this was to classify the ECG beats as 'Normal', 'SVC' and 'other' with the help of supervised machine learning technique. In particular, we used supervised learning technique like Logistic Regression, Neural Network, KNN, Support vector machine, Random forest, Decision Tree. With the help of proposed method, we easily classify the ECG beat.

From these algorithm Decision tree has the highest accuracy as 86.60%.

INDEX

LIST OF FIGURES

LIST OF TABLE

Chapter 1

Introduction

ECG stands for Electrocardiogram; it represents the cardiac points in the cardiac cycle of an individual. ECG is a method which is used to analyses the cardiovascular disease and results of this method helps in the diagnosis of the patient. ECG signal is a one-dimensional signal which is frequently used in the treatment of various heart disorders.

It provides the information about the functioning of the heart. The process of recording the ECG is simple, we simply connect the electrodes on the chest of the patient and record the electrical activity of the heart. If there is any abnormality is found in ECG, then it is termed as arrhythmia.

There are number of researchers who have detected the heart beat patterns based on the features extracted from ECG signal. The features representation is done in the frequency or time domain. Contingent upon the components, the grouping is permitted to perceive between classes. Presently a-days, the programmed ECG flag examination confronts a troublesome issue because of a huge variety in morphological and fleeting qualities of the ECG waveforms of various patients and similar patients. At various circumstances, the ECG waveforms may contrast for a similar patient to such a degree that they are not at all like each other and in the meantime indistinguishable for various sorts of beats. Inferable from this, the beat classifiers perform well on the preparation information however give poor execution on the ECG waveforms of various.

1.1 Activities of human heart

Heart is an organ found in the bodies of living creatures that pumps the blood throughout the body. The blood which is being circulated is the blood consisting of oxygen. The main function of the heart is the intake of the impure blood through the veins and supplying it to the lungs. Lungs hence do the job of purification. The heart of humans consists of four chambers namely left and right atrium and left and right ventricle. Other parts involve atrioventricular and Sino ventricular nodes. The flow diagram of the heart in shown in figure 1.1.

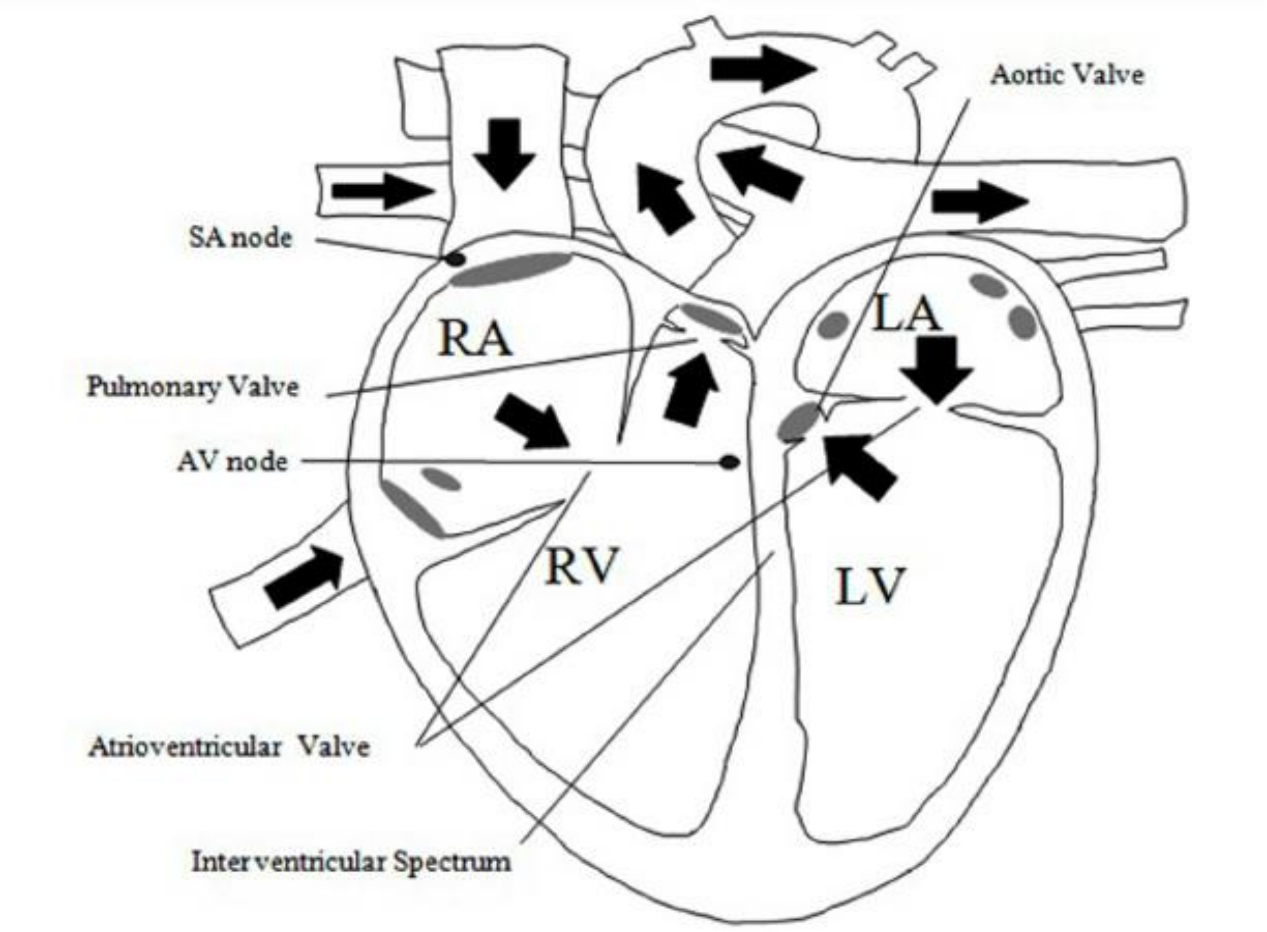

Fig. 1.1: Blood flow diagram of the human heart.

The right atria and the left atria are in the upper portion of the heart while the left and right ventricles are in the lower portion of the heart. For the matter of keeping both of the chambers isolated electrically, the atrias and the ventricles are connected by some tissue that is fibrous in nature and is non-conductive. The blood that is deoxygenated comes from the right atrium. Then the contraction of the right atrium takes place which results in the transfer of the blood to the right ventricle. This enables stretching of the ventricle and at this moment the pump action of the heart is at its peak. From the right ventricle, the blood is then transferred to the lungs where the refinement of the blood takes place. Here comes the duty for the left atrium. It collects the pure blood from the lungs. At this moment, the blood goes to left ventricle from the left atrium and again there is pumping in the heart. Contraction of the left ventricle takes place and the pure blood goes down the whole body. The pumping of the heart can be referred as function of systole and diastole. Duration for the contraction of muscle is termed as systole while the dilation of the heart muscle is called diastole. The function of the heart can be summarised as follows. The deoxygenated blood is first supplied to the lungs via heart then the purified or oxygenated blood is coming back in the heart. The heart then supplies the oxygenated or pure to the rest of the organs. Hence heart work in rhythm. And thus, the waveform of the heart beat waveform repeats itself after certain amount of time. The rhythm of the heart is regulated by the electrical signals generated by the heart pacemaker.

1.2 ECG (Electro Cardio Gram) Signal

The pumping activity of the human heart is recorded with the help of bio medical technique called as Electrocardiography. The rhythmic activity of the heart is calculated by applying mathematics on the beats of the heart also called as heart

beats which generally takes unit beats per minute. The heart beats are found with the help of the waveform generated by the ECG signal. [9] Gives detailed discussion of the calculation of the heart beats. A person having heart disease might have irregularities or the abnormality in the rhythmic activity of the heart. When the pattern of the heart's electrical activity shows difference compared to the regular morphological statistics of the heart, then a person is said to have arrhythmia. Result of electrocardiography on a person with a healthy heart is shown in figure 1.2. This shows a normal ECG signal. Figure 1.2 rather denotes a zoomed-out version of a typical ECG waveform.

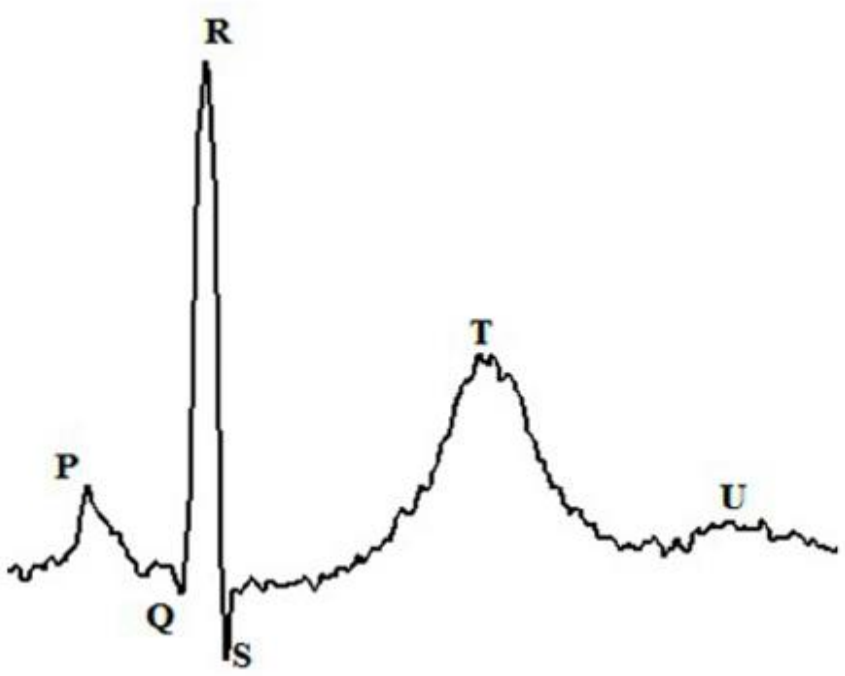

Fig. 1.2: The typical ECG waveform

The waveform of the ECG has several components that can be termed as P wave, T wave, U wave along with the QRS wave which is also called as the QRS complex. With the abnormality in the heart's functions; these waves are significantly affected by the disease. Hence the type of disease is generally determined taking the details

about the waves mentioned above. The wave P is formed when there is the transfer of blood to ventricles from the atria's. This is also called as depolarization. P wave is the result of the depolarization of the atriums. The QRS complex results when there is the depolarization of the ventricles.

PQ component of the ECG waveform is the segment of time needed for the settlement of the time taken while migration from atrias to ventricles. When the repolarization of the atrias takes place, then it is quite not visible because it is overlapped by the QRS complex. It can be seen from the above figure that the T wave is formed when the ventricles are repolarised. There is also another wave similar to the T wave which is also called as U wave. This U wave is due to the delayed polarization in the ventricles segment. U wave is again not visible [4]. A detailed diagram regarding the entire wave segment is shown in the figure below. This figure mentions all the time duration and corresponding ratio of amplitudes off the waves namely QRS, P, T etc. A brief overview of all the waves are mentioned below.

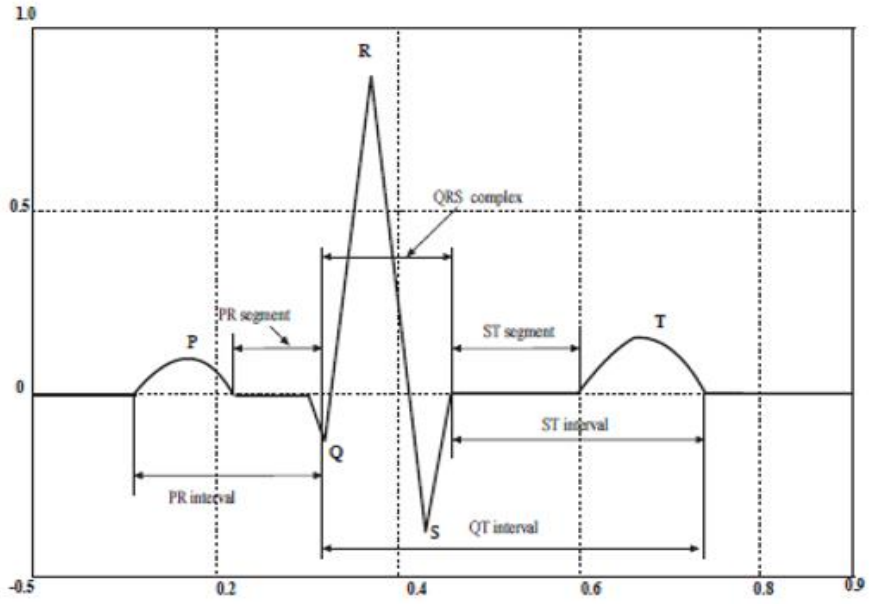

Fig. 1.3: Schematic representation of normal ECG signal.

1.3 ECG signal acquisition

The electrical activity if the hearts cells combines to form the electrical activity of the heart which results in the generation of the ECG waveform. The electrical activity of the heart muscles has influence on the surface of the body and these heart activity is the rapid change in the voltage signals generated due to the electrical activity of the heart. The stricture of a cell is such that it contains ionic conductors. The cell is then covered by the cell membrane. This cell membrane is considered as the selective ionic filter. There are ionic fluids in the body which also acts as carrier for the bio medical signals. The cell membrane alters its composition when the cells are stimulated by external currents. These external stimulus forces the sodium ion from the body fluid to enter into the cell which results in the effect of avalanche that also alters the composition of the cell. Since the movement of the potassium ion is slower as compared to that of sodium ion hence they are not able to migrate out of the cell while the cells of sodium ion, because of the avalanche effect, enter in the abundance inside the cell. When the increment of the sodium cells reaches equilibrium, a potential is released which represents the voltage difference of about +20 milli volts. This is also called as depolarisation of the cells. It takes quite some time to the repolarisation of the cell. In the repolarisation of the cells since the potassium ion is also involved in the process, hence we could suggest that the process of the repolarisation is similar to that of depolarization. The result of this process (depolarization and repolarisation) is the formation of the Electro Cardiogram signal. There are several leads which are placed in different parts of the body that helps in the acquisition of the ECG signal. The reference of the electrode is taken to be the right leg of the human body. 3 other leads I, II, III are placed in the left arm, right arm, left leg respectively. These three leads combine to form a triangle also known as the WILSON'S CENTRAL TERMINAL shown in figure 1.4. aVL stands for augmented limbs lead for the left arm, and aVR stands for the segmented limb lead

for the Right arm while aVF stands for the augmented limb lead for the foot.

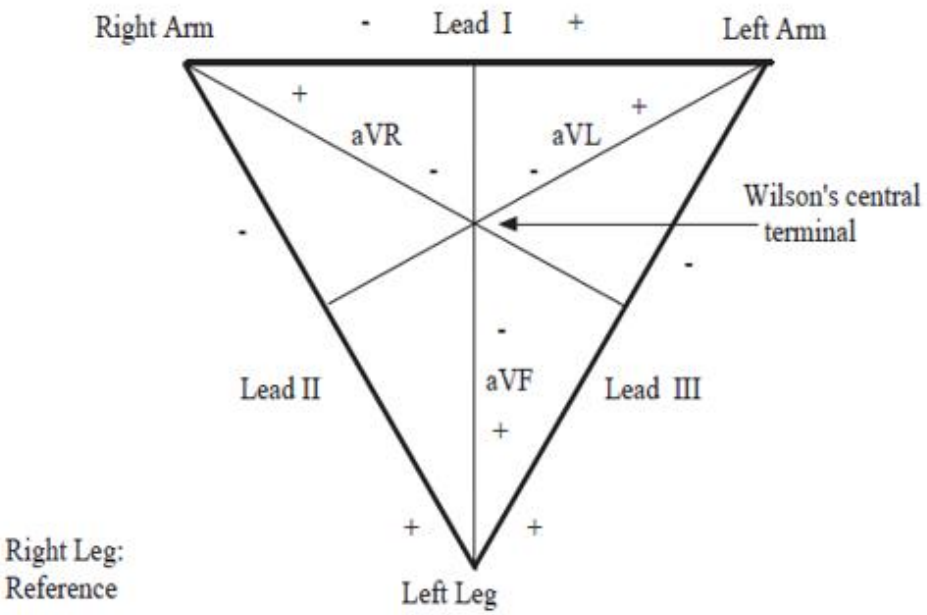

Fig. 1.4: Einthoven's triangle and the axes of the six ECG leads formed by using four limb leads.

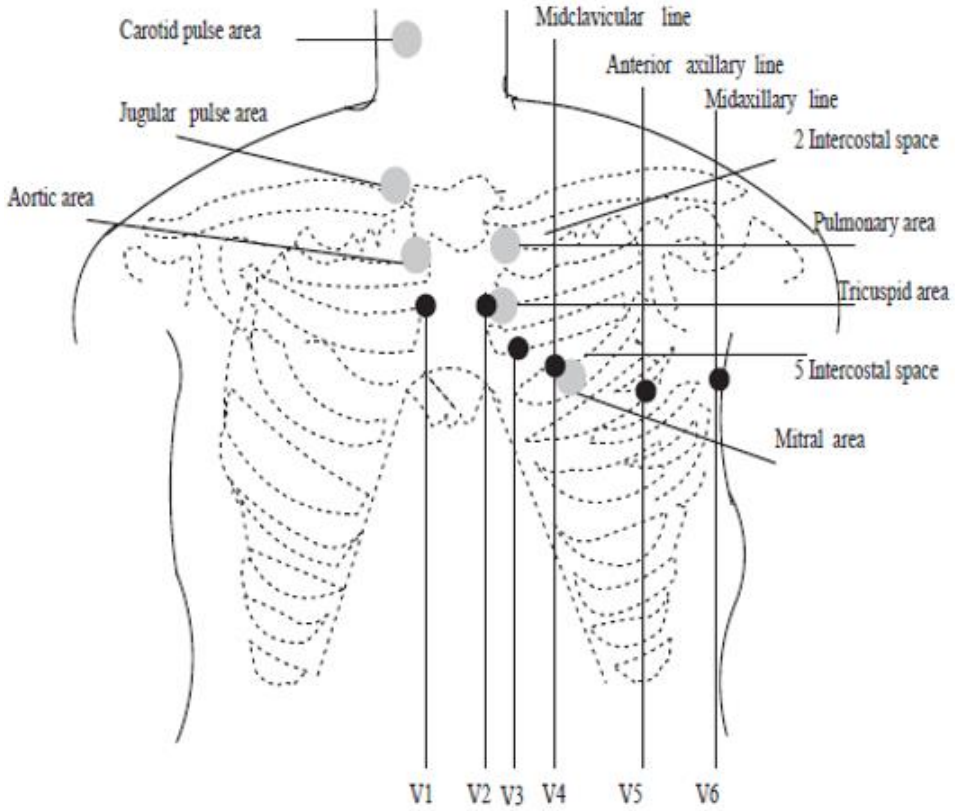

Fig. 1.5: Positions for placement of the leads V1-V6 for acquisition of ECG signal.

1.4 Different types of Heart Arrhythmia

When the heart of human is functionally acts normal or when there are no abbretions or disturbance or variations of the ECG signal compared to the morphologically defined ECG signal , then such activity of the heart is called Normal Sinus Rhythm (NSR). The rate of heart pumping of a normal person is about 60-100 beats/minute. RR interval is also effected by the cycle of breathing. When the heart beat of a human heart increases to a value of 200 beats/minute, then this type is called as sinus tachycardia. Whereas if the heart rate of a person is found below the normal range of heart beat, then it is called as sinus bradycardia. The bradycardia is found to be more dangerous than that of tachycardia because a slow heart rate effects the main organs of the body. The impropoer circulation of blood is the main cause behind the variuos heart arrhythmia. Some of the common types of heart arrythmia are discussed below:

1.4.1 Sinus Node Arrhythmia

The sino-atrial node of the heart is responsible for the arrhythmia known as sinus node arrhythmia. The pacemaker produces the electrical impulses.

Charactristics:

- 120ms or more variations in the interval in the wave P-P.
- The P-P interval slowly lengthens and compresses in a cyclical fashion, generally, corresponding to the phases of the respiratory cycle.
- Unchanged P-R interval

1.4.2 Premature Atrial Contraction

It is occurred when the P-wave is abnormal while the T wave and the QRS complex is completely normal. This arrhythmia occurs when the pacemaker fires before the

Sinus atrial node. The PAC could occur in pair or may be thrice sometimes. When the PAC occurs thrice or more then it is called as atrial tachycardia.

Pattern followed in the PACs are:

- **Bigeminy** — every other beat is a PAC.
- **Trigeminy** — every third beat is a PAC.
- **Quadrigeminy** — every fourth beat is a PAC.
- **Couplet** – two consecutive PACs.
- **Triplet** — three consecutive PACs.

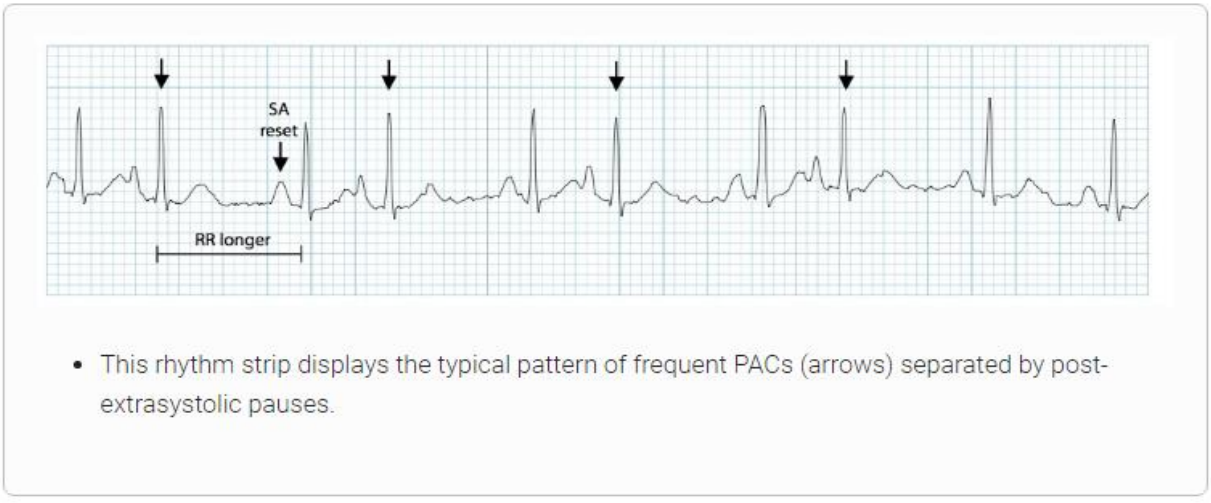

Fig 1.6: Rhythm strip displays of typical pattern of frequent PACs

1.4.3 Junctional Arrhythmia

It arises in the form of impulses arriving in the P-V node and its bundles. It generates in the atrio-ventricular junction. This type of arrhythmia occurs when the polarity of the P wave is opposite as that of normal P-wave.

1.4.4 Premature Ventricular Contraction

This type of arrhythmia generates from the ventricles. Thus, in this type of arrhythmia, there is no depolarization in the sinus atrial node. Thus, the morphology

of the P wave remains the same. The Pre-mature ventricular contraction may occur at any time of the heartbeat.

The PVC can be classified as:

1. **Unifocal** : Arising from a single ectopic focus; each PVC is similar.
2. **Multifocal** : Arising from two or more than two ectopic ; multiple QRS morphologies.
3. PVCs arising from the *right* ventricle have a left bundle branch block morphology (dominant S wave in V1).
4. PVCs arising from the *left* ventricle have a right bundle branch block morphology (dominant R wave in V1).

1.4.5 Junctional Escape

In the arrhythmia cases having the junction escape, the width of the QRS complex is generally less than 120ms. Thus, the rhythm rate of the heart is below that is in the range 40 to 60 beats per minute.

Causes of Junctional Escape:

1. Severe sinus bradycardia
2. Sinus arrest
3. Sino-atrial exit block
4. High-grade second-degree AV block
5. Third degree AV block
6. Hyperkalaemia
7. Drugs: beta-blocker, calcium-channel blocker or digoxin poisoning

Chapter 2

Literature Review

There have been a significant number of studies [1-15] using mathematical tools to track the presence of arrhythmias in the ECG. Pre-processing includes classical digital filters to remove baseline noise, noise at high frequencies and power line. Work will focus on finding unique features to find the pattern of beats in a more efficient manner. Common features involve the RR interval or heart rate, heart rate normalization, statistical characteristics such as mean, standard deviation, maximum, minimum, kurtosis, skewness, transformations not lines, DWT (Discrete Wavelet Transform) coefficients, the presence of waves and characteristic heartbeat intervals as the P wave or QRS duration known widely in the medical field. They have been used most common supervised classification algorithms. In particular, SVM (Support vector machine) with the kernel of RBF (Radial basis function) and different types of Neural Networks (NN) have been used extensively to have the ability to create non-linear decision functions between classes. Technically unsupervised as (PCA) Principal Component Analysis, (ICA) Independent Component Analysis, Factorial, Kohonen Maps self-organizing analysis have also been used in order to reduce the dimensionality of the features. Other works also include analysis of agglomerates as K-means Gaussian mixture.

One of the most important points is the way to approach the problem. There are two acceptable ways to address the problem, intra and inter methodologies patients [16], apart from the way of classifying beats.

(AAMI) The Association for the Advancement of medical instrument provides guidelines for classifying beats of an ECG signal in normal beats, supraventricular

beats, ventricular beats, fusion beats and beats not rated [17, 18]. However, the presence of the latter two types of beats is almost nil comparing them with normal, supraventricular and ventricular beats. In recent work, usually not taken as a kind of beat, but the fusion beats are included in ventricular beats and patients with a large number of beats unsorted and / or their heartbeats are removed for testing, working around normal, supraventricular and ventricular beats.

Reported work using inter-patients [17-20] proposal. The yield is much lower than [13-17] because there is no information from the same patient between sets training, validation and testing. However, these works reflect the real difficulty because the results involve the inevitable variability among patients.

For these reasons, there are few published works this way. **Chazal et al.** [17] work extracting features in the time domain as heart rate and descriptors wave of the two signals of each patient recordings database MITBIH available to subsequently split signals patients training and testing, unmixed same patient information between sets. Bayesian classifier used the Linear Discriminant doing better than their predecessors. **Llamedo et al.** [18] presented another overall classifier using the same patient's database MITBIH in the training set and test. **Chazal et al.** [17] but included in their work two databases have more to get more results and better results of its implementation. He used heart rate and learned characteristics of the discrete wavelet transform. Moreover, he tested different sets of features and two different classifiers Bayesian staying with Linear Discriminant sorting. **Llamedo et al.** [22] last modified work to include more than two leads, using the same features and classifier. The database was required INCART containing 12 standards. Using all leads, he could improve his previous work, unfortunately it is the only database available with this number of referrals and labelled. **Ye et al.** [23] work in classifying all types of arrhythmias in the database and apart MITBIH sorting according to

AAMI. He used morphological characteristics of the DWT coefficients and implemented ICA each beat. Given the high dimensionality matrices which reduced the PCA algorithm. Once reduced, he introduced the characteristic heart rate. SVM classifier was used and combined the two branches of MITBIH database. Once the features were extracted, he first mixed patient information in both the training and test sets of obtaining results very high, close to 100%. Then tested their methodology training and testing as in **Chazal et al.** [17] and **Llamedo et al.** [18, 22] significantly lowering the performance variability between patients. The results are comparable with those of **Chazal et al.** [17] and **Llamedo et al.** [18, 22]. Mar et al. [20] mainly work in the selection of characteristics to classify ECG signals. He proved a considerable number of features time domain wave morphology, statistical characteristics and time-frequency space using discrete wavelet transform. Search algorithm Sequential Floating Forward was used to find out which features are relevant in conjunction with linear discriminant Bayesian classifier. Once the optimum characteristics were removed, a perceptron multilayer classifier was used. The results were compared with the works of Chazal et al. [17] Llamedo et al. [18, 22] and Ye et al. [19] but with selected characteristics.

In [16], it is mentioned that models general classifiers are few reliable and are not widely used in practice. An alternative is the inter-patient methodology, since the pattern of a single patient is much easier than the combined pattern of several patients. These jobs require a certain number of beats previously labelled by a medical expert to train the algorithm. Papers presented in [17 - 21] results obtained above 95% in most cases, some fences 100% positive predictive value and sensitivity when it comes to classify beats according to the AAMI standard. Only **Llamedo et al.** [23] and **Al Rahhal et al.** [24] presented results without training the same patient or medical help. **Llamedo et al.** [23] modifies his previous work [18,

22] a general classifier, including the K-means algorithm, a model unsupervised. This modification improves the performance of their previous work and allows an expert to contribute to the result of the algorithm. The end result is about 95% with medical help. **Al Rahhal et al.** [24] uses active sorting using a deep learning approach. They are using the algorithm "Auto encoder" to represent the signal in a minor dimension in conjunction with the heart rhythm to be part of the characteristics. Finally, a "Normalized exponential or softmax " function is trained in the NN with the same patients as in **Chazal et al.** [28] **Llamedo et al.** [17,22,23], **Ye et al. [19]** and **Mar et al.** [20]. Their results, helps with medical expertise, is close to 100% with a significant number of previously tagged the patient's heartbeat. Wiens et al. [21] compares the result of its algorithm developed based on active learning with software HAMILTON, this software is implemented based on rules and has the advantage that does not require prior information or beats previously labelled the same patient, but only recognizes beats ventricular premature. The algorithm **Wiens et al.** [21] it is being better than HAMILTON with a number of beats labelled the same patient.

In short, it seems that so far it has not been possible to overcome the problem with the variability among patient to train classifiers general heartbeats. The expert guidance of a physician is required to develop a model that fits the patient. This means that in the future, other doctors have the need to relabel other beats to keep the algorithm working and updated as the cardiac signal may change over time as and heart rate.

We present a new methodology which takes a closer unsupervised to sort between beats of supraventricular origin (SVO) formed by beating normal rate and supraventricular arrhythmia, and ventricular origin (VO), formed by the types of

arrhythmia and ventricular fusion. This approach does not require medical care, as it tries to capture the signal pattern and is developed based on observations made when looking for a general pattern signals. According to our experiments, this method works best when there are at least 40 beats of ventricular origin. This study is a first approximation to the normal beats consequently separating and supraventricular arrhythmias origin in future work.

Chapter 3

Proposed Methodology

Our proposed method includes the following steps:

Fig 3.1: Block diagram for the process of proposed methodology

3.1 Input ECG signal

Input ECG signals were taken from the MIT-BIH Arrhythmia Database. The MIT-BIH Arrhythmia Database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979. Twenty-three recordings were chosen at random from a set of 4000 24-hour ambulatory ECG recordings collected from a mixed population of inpatients (about 60%) and outpatients (about 40%) at Boston's Beth Israel Hospital; the remaining 25 recordings were selected from the same set to include less common but clinically significant arrhythmias that would not be well-represented in a small random sample.

The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. Two or more cardiologists independently annotated each record; disagreements were resolved to obtain the computer-readable reference annotations for each beat (approximately 110,000 annotations in all) included with the database.

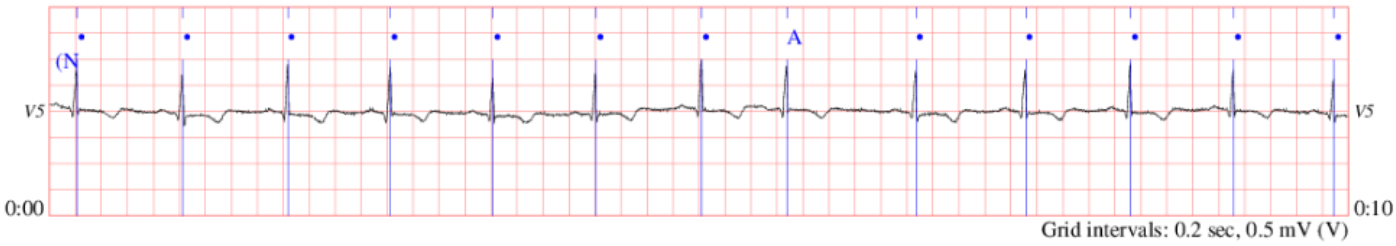

Fig 3.2: Sample ECG signal from MIT-BIH arrhythmia database

As it can be seen from the figure above, each beat is annotated so that it can be identified as per the required application. Annotation symbols used for different type of beats can be listed as in the table below:

Symbol	Meaning
. or N	Normal beat
L	Left bundle branch block beat
R	Right bundle branch block beat
A	Atrial premature beat
A	Aberrated atrial premature beat
J	Nodal (junctional) premature beat
S	Supraventricular premature beat
V	Premature ventricular contraction
F	Fusion of ventricular and normal beat
[	Start of ventricular flutter/fibrillation
!	Ventricular flutter wave
]	End of ventricular flutter/fibrillation
E	Atrial escape beat
J	Nodal (junctional) escape beat
E	Ventricular escape beat
/	Paced beat
F	Fusion of paced and normal beat
X	Non-conducted P-wave (blocked APB)
Q	Unclassifiable beat
|	Isolated QRS-like artifact

Table 3.1: Different type of beat annotations for MIT-BIH arrhythmia database

3.2 Beat Extraction

With the help of annotation file provided with each of the ECG signal in MIT-BIH arrhythmia database, we can easily extract a beat from the ECG signal with the help of following steps:

1. Detect R peak with the help of annotation file (annotation file provides the approximate location of R peak, with the help of MATLAB we can see the maximum value around that time or sample value).
2. Extract part of ECG signal around the R peak using window of predefined width. In our work, we defined the window using the standard value for different parts of the ECG signal. Further explanation about defining the window size is given below.

Defining the window size

After detecting the R peak, we used two windows to extract the ECG signal. One window is for the extraction of ECG signal from starting to R peak and another one is for the extraction of ECG signal from R peak to the ending. Before discussing further about the window size, lets take a closer look at the different parts of the ECG signal and their range of duration. An ECG signal is shown below with its constituent parts.

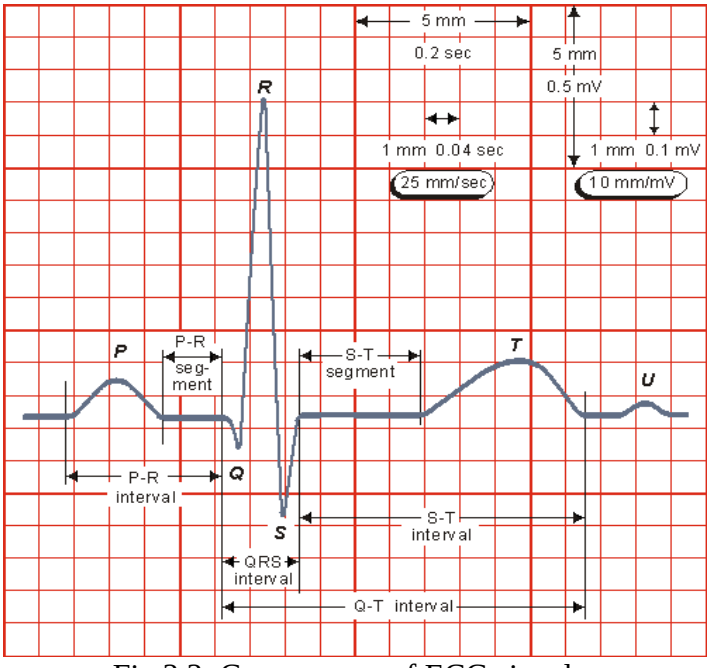

Fig 3.3: Components of ECG signal

Duration of the different components of ECG signal varies according to the type of arrhythmia in objective signal. For example, in case of PVC, width of the QRS complex increases while for SVC it decreases. There exists a range for the time duration of these segments or intervals. These are:

Segment or Interval	***Time duration (ms)***
PR interval	120-200
QRS complex	<100
ST segment	80-120
T wave	100-250

Table 3.2: Time duration of different ECG components

These are the main components of a single beat of any ECG signal which contains the maximum amount of information. As the R peak lies almost in between the QRS complex, thus R peak can be taken as the midpoint of QRS complex. Now even after considering the maximum values for the constituting parts of ECG signal, the length of the window before and after the R peak can be defined as:

Before R peak: 200 ms for PR interval + 50 ms for half of the QRS complex = 250ms

After R peak: 50 ms for remaining half of the QRS complex + 120 ms for ST segment + 250 ms for T wave = 420 ms.

Thus, the total duration of a single beat will be: 250 + 420 = 670 ms.

MIT-BIH dataset is sample at 360 Hz. So, the sampling duration will be:

$$Sampling\ Duration = \frac{1}{360} = 2.7777\ ms \tag{3.1}$$

Above procedure can be easily visualized as shown in the figure below:

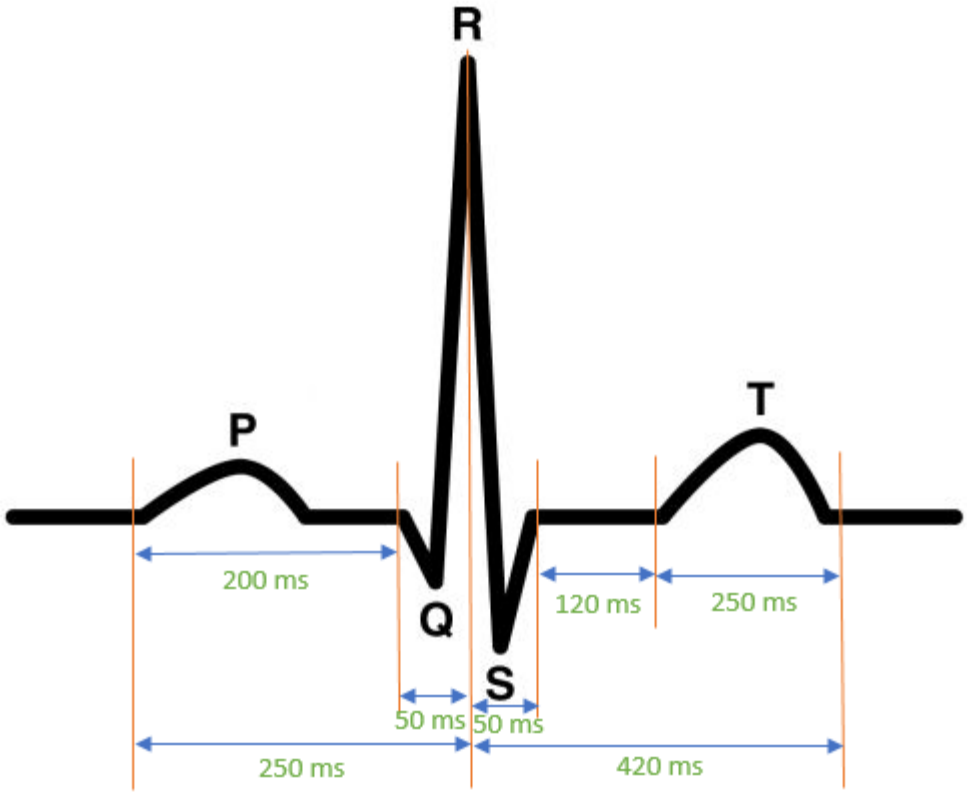

Fig 3.4: Window size for extraction of beats

Total number of samples in a single beat will be:

$$round\left(\frac{670}{2.7777}\right) = 241. \tag{3.2}$$

3.3 Feature Extraction

Accuracy of any machine learning classification task depends a lot on the quality of features. Features should have the following properties:

1. Good representation of the data.
2. Dimensionality should be proper with respect to computational resources.
3. Features should not be redundant.

For extraction of the features we have following choices:

1. Feed all the sample values directly to the machine learning classifier.
2. Feed equally spaced samples to the machine learning classifier.
3. Calculate some statistical parameters and use these values as features.

First option have dimensionality problem. There is a total of 241 features in every beat. Processing that much amount of data will need high computational resources. Besides this, it doesn't represent data very well thus the classification accuracy will low. Second option may have appropriate dimension according to the choice of sampling interval but doesn't represent data very well. Third option will represent data efficiently as compared to the above choices but it will have less number of features thus will underfit the data.

Thus, a proper choice for feature extraction will be to use hybrid technique *i.e.* a combination of statistical and signal value parameters. Here we used ***mean***, ***standard deviation***, ***maxima*** and ***minima*** as the statistical parameters and six sample values are taken as signal value features.

For better representation of ECG beat, this feature extraction process was applied after separating every beat into its mainly constituent parts, in the time domain. For frequency domain, we divided the whole signal into four equal parts. Each ECG beat in time domain is separated into four parts that contains the maximum amount of information *i.e.*

1. P wave
2. QRS complex (including PR segment and some part of ST segment)
3. ST segment
4. T wave

Note:

1. 2nd portion of the beat contains portion of PR segment to its left and some part of ST segment (till J node, approx.).

   This is because QRS complex is the most important part of the ECG signal to discriminate most of the arrhythmias. Thus, it is a good practice to take maximum possible width for QRS complex so that we can cover it even in the case of arrhythmias where width of the QRS complex changes significantly. For example, in the case of PVC where width of the QRS complex increases.

2. As the width of every constituent part (P wave, QRS complex, St Segment and T wave) varies from time to time and for each patient also, so there is no general rule or algorithm to separate these parts from ECG beat. Thus, partition of the ECG beats in four parts was done on the abstract basis. There will be overlap for some parts but this is where machine learning algorithms outperforms the common statistical algorithms.

   ECG beats classification is a kind of pattern classification problem where ML classifier will learn the input pattern and gives the output accordingly. So even after the overlap or extension of any part of ECG signal in the other part of the signal, output label will also change accordingly and ML classifier learn these changes and gives accurate results.

Now feature can be extracted for both the time and frequency domain in following steps:

1. Partition the ECG beat into its four constituent parts for time domain and for frequency domain divide the whole signal into four equal parts.
2. For each part extract four statistical and six sample features. These six characteristic samples are: **in case of time domain**, the beginning (S1) and the ending (S6) of each signal, also four samples (S2,S3,S4,S5) equally spaced in between the signal and **in case of frequency** as we divided it in quarterly so the beginning (S1) and the ending (S6) of the quartered vector, also four samples (S2,S3,S4,S5) equally spaced into the quartered vector as we can see in the following fig 3.5. Statistical Features can be computed using following formulae:

$$floor(n \times \frac{N}{5}) \quad (3.3)$$

Where,

n = 1 to 5, and

N = Total number of samples

For Example:

For P wave: N= 54(as discussed earlier)

S1 = 1

S2 = $floor(1 \times \frac{54}{5})$ = floor (10.8) = 10.

S3 = $floor(2 \times \frac{54}{5})$ = floor (21.6) = 21

S4 = $floor(3 \times \frac{54}{5})$ = floor (32.4) = 32

S5 = $floor(4 \times \frac{54}{5})$ = floor (43.2) = 43

S6= $floor(5 \times \frac{54}{5})$ = floor (54.0) = 54

Similarly we can find for others also.

Fig 3.5: Feature extraction process from ECG beats

From the above figure, it is clear that for every ECG beat a total of 80 features are extracted. 40 from the time domain signal and 40 from the frequency domain signal. Feature extraction process is same for both the domains.

Partition process is further explained below.

3.3.1 Partition process

Time domain signal is divided into four parts as P wave, QRS complex, ST segment and T wave. As mentioned above, a single beat extracted from the ECG signal is of 670 ms time duration and consists of 241 samples. Now to divide this signal into its four constituents, we took the time duration for these parts as below:

ECG beat component	***Time duration***
P wave	150 ms
QRS complex	170 ms
ST segment	100 ms
T wave	250 ms

Table:3.3: ECG beat component and their time duration

Maximum time duration for PR interval is 200 ms. It includes P wave and PR segment. Thus, to extract only the P wave and to give larger space for QRS complex we took it as 150 ms. QRS complex lasts for 100 ms at maximum, after including the approximate PR segment of 50 ms from the left and 20 ms of time duration from right (from ST segment), total width of QRS complex will be 170 ms. Now after including the 20 ms part of ST segment into QRS partition, remaining width of the ST segment will be 100 ms. And for the fourth part of T wave, its already mentioned that the maximum duration of T wave is 250 ms. To divide the signal in the MATLAB software we need to know the number of samples in each part. This can

be computed by dividing the time duration for each part with sampling time duration *i.e.* ***2.7777 ms.***

After following the above computation, we got the results as shown below:

ECG beat component	***Number of samples***
P wave	54
QRS complex	61
ST segment	36
T wave	90

Table:3.4: ECG beat component and their Number of samples

For frequency domain, we divided the signal into four equal parts as the frequency axis is symmetric for whole the signal.

Note:

For better performance of the machine learning classifiers, we need to normalize the input signal. In particular, normalization reduces the convergence time for classifier. Random initialization can cause longer convergence time, normalization overcomes that problem. Most commonly used normalization technique is the mean normalization which can be done using following formula:

$$x_i = \frac{x_s - \mu_{ECG}}{maxima_{ECG} - minima_{ECG}} \quad (3.4)$$

Where, x_s is the sample value of ECG, μ_{ECG} is the mean value of ECG and $maxima_{ECG}, minima_{ECG}$ are the maximum and minimum value of ECG signal.

3.4 Classification

After extracting the features from ECG beats, our next task is to classify them. We extracted three types of beats from the MIT-BIH database *i.e.* Normal, SVC and Other type of beats. Class labels are assigned to these type of beats as shown below:

Type of beat	***Class label***
Normal	1
Other	2
SVC	3

Table 3.5: Class label for different type of beats

We used six different classifiers for the classification tasks. These are:

1. Logistic Regression
2. Neural Network
3. Support Vector Machine
4. Decision Tree
5. Random Forest
6. K-Nearest Neighbour

Performance analysis and comparison will be discussed in Results chapter. A brief overview of these classifiers is discussed below.

3.4.1 Logistic Regression

Logistic Regression belongs to the family of generalized linear models. It is a binary classification algorithm used when the response variable is dichotomous (1 or 0). Inherently, it returns the set of probabilities of target class. But, we can also obtain response labels using a probability threshold value. Following are the assumptions made by Logistic Regression:

1. The response variable must follow a binomial distribution.
2. Logistic Regression assumes a linear relationship between the independent variables and the link function (logit).
3. The dependent variable should have mutually exclusive and exhaustive categories.

Logistic regression is named for the function used at the core of the method, the logistic function. The logistic function, also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

Hypothesis representation for Logistic regression

Consider the general form of hypothesis for linear regression, *i.e.*

$$h_\Theta(x) = \theta^T x \tag{3.5}$$

now this function can be modified as shown below:

$$h_\Theta(x) = g(\theta^T x) \tag{3.6}$$

$$g(z) = \frac{1}{1+e^{-z}} \tag{3.7}$$

Thus,

$$h_\Theta(x) = \frac{1}{1+e^{-\Theta^T x}} \tag{3.8}$$

This function is called sigmoid or logistic function.

Now logistic function can be trained by defining a cost function which can be minimized with the help of gradient descent algorithm. Cost function and gradient descent functions can be defined as shown below:

Cost function

At every step, this cost function will be calculated until it reaches at predefined level. Gradient descent algorithm will change the value of weights after every iteration by differentiating the cost function.

$$J(\theta) = \left[\frac{1}{m}\sum_{i=1}^{m} y^i \log\left(h_\theta\left(x^i\right)\right) + \left(1 - y^i\right)\log(1 - h_\theta(x^i))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n} \Theta_j^2 \quad (3.9)$$

Gradient descent

Weight update should be simultaneous here *i.e.* first of all calculate the derivative and then update all the weights at the same time.

$$\theta_0 := \theta_0 - \alpha\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^i) - y^i\right) x_0^i \quad (3.10)$$

$$\theta_j := \theta_j - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^i) - y^i\right) x_j^i\right] + \frac{\lambda}{m}\theta_j \quad ; \quad j = 1,2 \ldots\ldots n \quad (3.11)$$

3.4.2 Neural Network

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning

process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well.

Neural networks are typically organized in layers. Layers are made up of a number of interconnected 'nodes' which contain an 'activation function'. Patterns are presented to the network via the 'input layer', which communicates to one or more 'hidden layers' where the actual processing is done via a system of weighted 'connections'. The hidden layers then link to an 'output layer' where the answer is output as shown in the figure below.

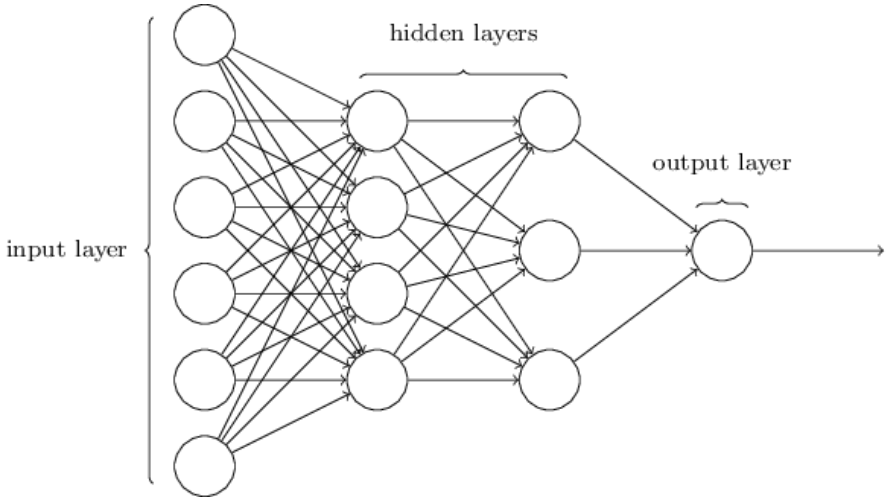

Fig 3.6: Neural Network Diagram

Most ANNs contain some form of 'learning rule' which modifies the weights of the connections according to the input patterns that it is presented with. In a sense, ANNs learn by example as do their biological counterparts; a child learns to recognize dogs from examples of dogs.

Although there are many different kinds of learning rules used by neural networks, the delta rule is often utilized by the most common class of ANNs called 'backpropagational neural networks' (BPNNs). Backpropagation is an abbreviation for the backwards propagation of error.

With the delta rule, as with other types of backpropagation, 'learning' is a supervised process that occurs with each cycle or 'epoch' (i.e. each time the network is presented with a new input pattern) through a forward activation flow of outputs, and the backwards error propagation of weight adjustments. More simply, when a neural network is initially presented with a pattern it makes a random 'guess' as to what it might be. It then sees how far its answer was from the actual one and makes an appropriate adjustment to its connection weights.

Note also, that within each hidden layer node is a sigmoidal activation function which polarizes network activity and helps it to stabilize.

Backpropagation performs a gradient descent within the solution's vector space towards a 'global minimum' along the steepest vector of the error surface. The global minimum is that theoretical solution with the lowest possible error. The error surface itself is a hyper paraboloid but is seldom 'smooth'.

Sigmoid function is used most commonly as the activation function for ANNs. Sigmoid function can be defined as:

$$S(x) = \frac{1}{1+e^{-x}} \qquad (3.12)$$

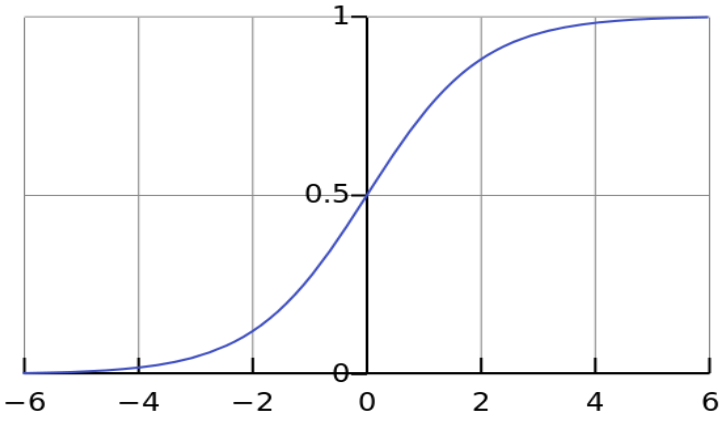

Fig 3.7: Sigmoidal function

3.4.3 Support Vector Machine

A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be employed for both classification and regression purposes. SVMs are more commonly used in classification problems. SVMs are based on the idea of finding a hyperplane that best divides a dataset into two classes, as shown in the image below.

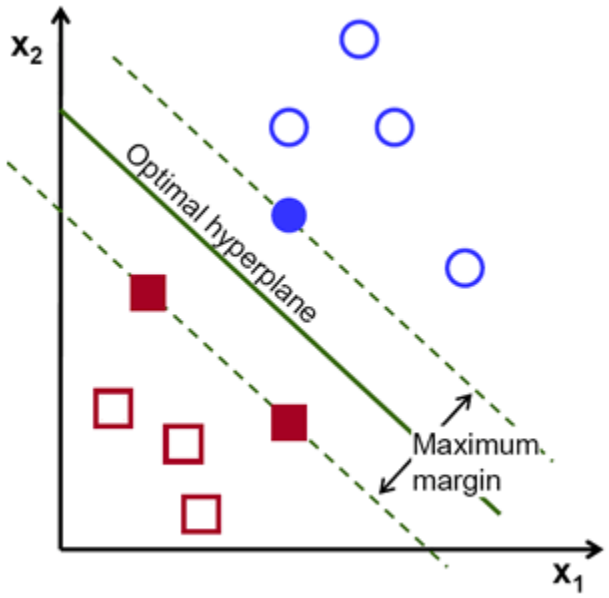

Fig 3.8: Support vector machine diagram

Support Vectors

Support vectors are the data points nearest to the hyperplane, the points of a data set that, if removed, would alter the position of the dividing hyperplane. Because of this, they can be considered the critical elements of a data set.

Hyperplane

As a simple example, for a classification task with only two features (like the image above), we can think of a hyperplane as a line that linearly separates and classifies a set of data.

Intuitively, the further from the hyperplane our data points lie, the more confident we are that they have been correctly classified. We therefore want our data points to be as far away from the hyperplane as possible, while still being on the correct side of it.

So, when new testing data is added, whatever side of the hyperplane it lands will decide the class that we assign to it.

Finding the best hyperplane

The distance between the hyperplane and the nearest data point from either set is known as the margin. The goal is to choose a hyperplane with the greatest possible margin between the hyperplane and any point within the training set, giving a greater chance of new data being classified correctly.

For multiclass classification, SVM uses one-vs-all technique in which the class under consideration is taken as positive and all the other classes are taken as negative.

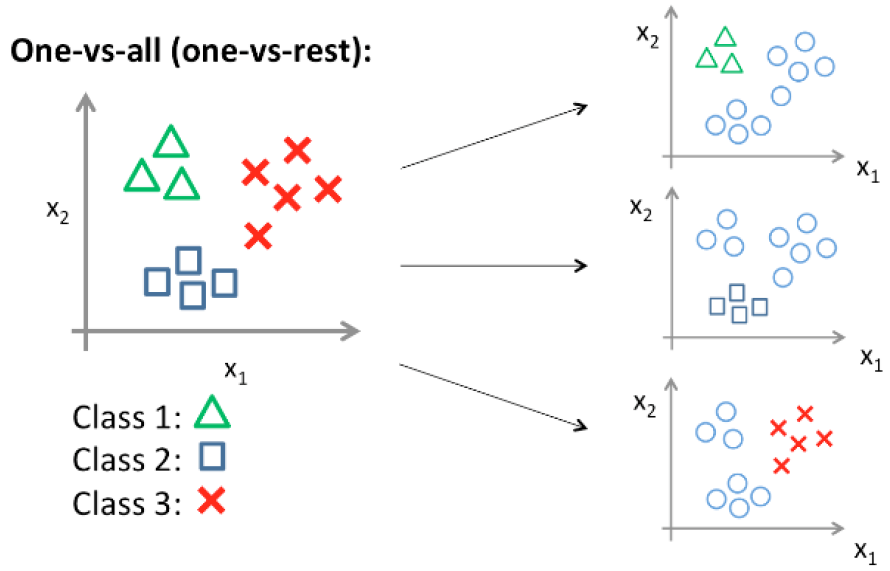

Fig 3.9: One-vs-all classification

The objective function for SVM involves minimization of the error function and can be defined as below:

$$\frac{1}{2}w^T w + C \sum_{i=1}^{N} \zeta_i \quad (3.13)$$

Subject to the constraints:

$$y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i \quad (3.14)$$

Where, C is the capacity constant, w is the vector of coefficients, b is a constant, and ζ_i represents parameters for handling nonseparable data (inputs). The index i labels the N training cases.

Note that $y \epsilon \pm 1$ represents the class labels and x_i represents the independent variables. The kernel ϕ is used transform data from the input (independent) to the feature space. It should be noted that the larger the C, the more the error is penalized. Thus, C should be chosen with care to avoid overfitting.

As mentioned above SVM uses kernel trick to classify nonlinear data. There exist many types of kernel functions for SVM. Most commonly used kernels are:

- Polynomial kernel: $(x^T x_i + \Theta)^d$
- Gaussian kernel: $e^{-\frac{1}{2\sigma^2}||x-x_i||^2}$
- Sigmoid kernel: $\tanh(\eta x x_i + \theta)$

3.4.4 Decision Tree

Tree based learning algorithms are considered to be one of the best and mostly used supervised learning methods. Tree based methods empower predictive models with high accuracy, stability and ease of interpretation. Unlike linear models, they map

non-linear relationships quite well. They are adaptable at solving any kind of problem at hand (classification or regression).

Decision tree is a type of supervised learning algorithm (having a pre-defined target variable) that is mostly used in classification problems. It works for both categorical and continuous input and output variables. In this technique, we split the population or sample into two or more homogeneous sets (or sub-populations) based on most significant splitter / differentiator in input variables.

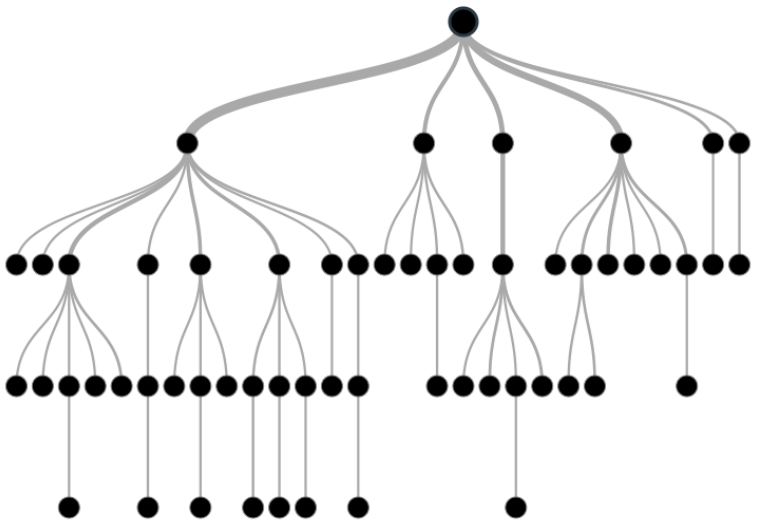

Fig 3.10: Example decision tree

3.4.4.1 Types of Decision Trees

Types of decision tree is based on the type of target variable we have. It can be of two types:

1. **Categorical Variable Decision Tree:** Decision Tree which has categorical target variable then it called as categorical variable decision tree.
2. **Continuous Variable Decision Tree:** Decision Tree has continuous target variable then it is called as Continuous Variable Decision Tree.

3.4.4.2 Important Terminology related to Decision Tree

1. **Root Node:** It represents entire population or sample and this further gets divided into two or more homogeneous sets.
2. **Splitting:** It is a process of dividing a node into two or more sub-nodes.
3. **Decision Node:** When a sub-node splits into further sub-nodes, then it is called decision node.
4. **Leaf/ Terminal Node:** Nodes do not split is called Leaf or Terminal node.

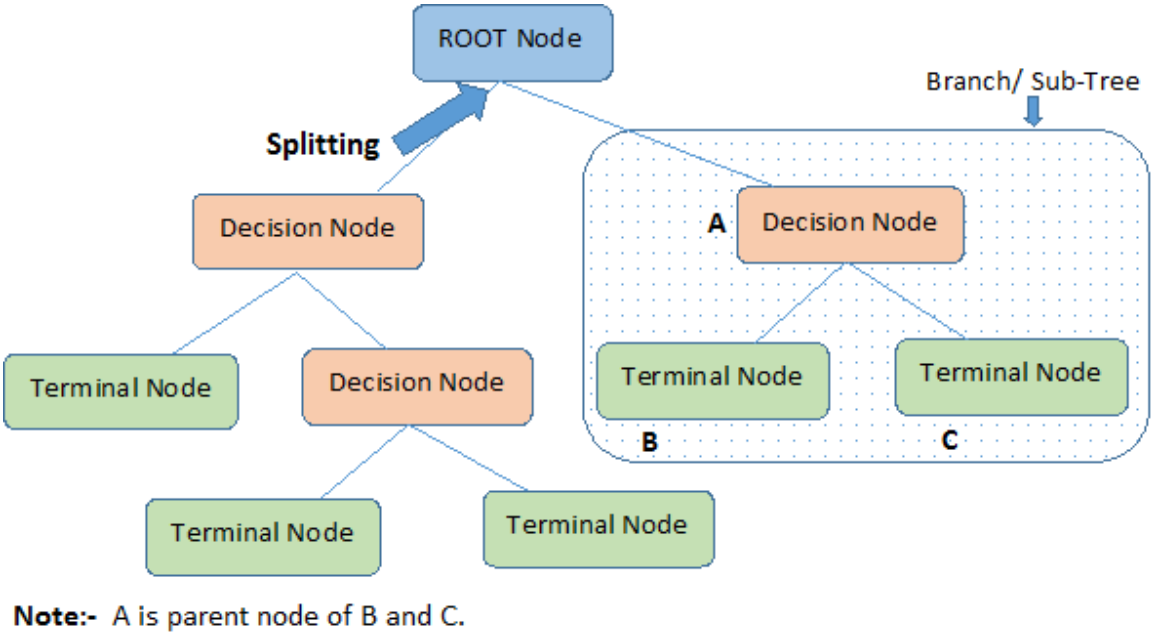

Fig 3.11: Terminologies related to decision tree

5. **Pruning:** When we remove sub-nodes of a decision node, this process is called pruning. You can say opposite process of splitting.
6. **Branch / Sub-Tree:** A sub section of entire tree is called branch or sub-tree.
7. **Parent and Child Node:** A node, which is divided into sub-nodes is called parent node of sub-nodes whereas sub-nodes are the child of parent node.

3.4.4.3 Algorithms for Decision tree splitting

Decision trees use multiple algorithms to decide to split a node in two or more sub-nodes. The creation of sub-nodes increases the homogeneity of resultant sub-nodes. In other words, we can say that purity of the node increases with respect to the target variable. Decision tree splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes. Most commonly used algorithms are:

1. Gini index
2. Chi-Square
3. Information Gain
4. Reduction in variance

We used the information gain algorithm for splitting of the decision tree. This is briefly explained below.

3.4.4.4 Information Gain algorithm

In the image shown below, if we look at each node and think which node can be described easily. The answer will be C because it requires less information as all values are similar. On the other hand, B requires more information to describe it and A requires the maximum information. In other words, we can say that C is a Pure node, B is less Impure and A is more impure.

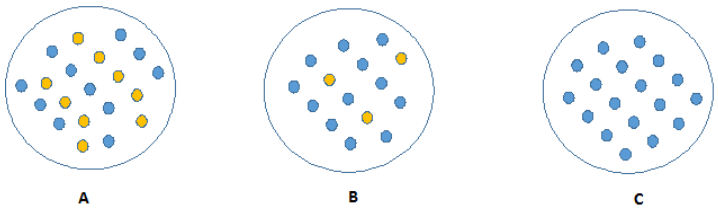

Fig.3.12: Information gain

Now, we can build a conclusion that less impure node requires less information to describe it. And, more impure node requires more information. Information theory is a measure to define this degree of disorganization in a system known as Entropy. If the sample is completely homogeneous, then the entropy is zero and if the sample is an equally divided (50% – 50%), it has entropy of one.

Entropy can be calculated using formula:

$$Entropy = -p\ log_2 p - q \log_2 q \tag{3.15}$$

Here p and q is probability of success and failure respectively in that node. Entropy is also used with categorical target variable. It chooses the split which has lowest entropy compared to parent node and other splits. The lesser the entropy, the better it is.

3.4.4.5 Steps to calculate entropy for a split:

1. Calculate entropy of parent node
2. Calculate entropy of each individual node of split and calculate weighted average of all sub-nodes available in split.

3.4.5 Random Forest

Random forest is a tree-based algorithm which involves building several trees (decision trees), then combining their output to improve generalization ability of the model. The method of combining trees is known as an ensemble method. Ensembling is nothing but a combination of weak learners (individual trees) to produce a strong learner. In a normal decision tree, one decision tree is built and in

a random forest algorithm number of decision trees are built during the process. A vote from each of the decision trees is considered in deciding the final class of a case or an object, this is called ensemble process. This is a democratic process. Since, many decision trees are built and used in a process of Random Forest algorithm, it is called a forest.

Random Forest can be used to solve regression and classification problems. In regression problems, the dependent variable is continuous. In classification problems, the dependent variable is categorical.

Random Forest uses Gini Index based impurity measures for building decision tree. Gini Index is also used for building Classification and Regression Tree (CART).

3.4.5.1 Gini Index

Gini index says, if we select two items from a population at random then they must be of same class and probability for this is 1 if population is pure.

1. It works with categorical target variable "Success" or "Failure".
2. It performs only Binary splits
3. Higher the value of Gini higher the homogeneity.
4. CART (Classification and Regression Tree) uses Gini method to create binary splits.

Steps to Calculate Gini for a split

1. Calculate Gini for sub-nodes, using formula sum of square of probability for success and failure ($p^2 + q^2$).
2. Calculate Gini for split using weighted Gini score of each node of that split.

3.4.6 k-Nearest Neighbour

Neighbors-based classification is a type of *instance-based learning* or *non-generalizing learning*: it does not attempt to construct a general internal model, but simply stores instances of the training data. Classification is computed from a simple majority vote of the nearest neighbors of each point: a query point is assigned the data class which has the most representatives within the nearest neighbors of the point.

he principle behind nearest neighbor methods is to find a predefined number of training samples closest in distance to the new point, and predict the label from these. The number of samples can be a user-defined constant (k-nearest neighbor learning), or vary based on the local density of points (radius-based neighbor learning). The distance can, in general, be any metric measure: standard Euclidean distance is the most common choice. Neighbors-based methods are known as *non-generalizing* machine learning methods, since they simply "remember" all of its training data (possibly transformed into a fast indexing structure such as a Ball Tree or KD Tree.).

Despite its simplicity, nearest neighbors have been successful in a large number of classification and regression problems, including handwritten digits or satellite image scenes. Being a non-parametric method, it is often successful in classification situations where the decision boundary is very irregular.

The KNN classifier is also a **non-parametric** and **instance-based** learning algorithm.

- **Non-parametric** means it makes no explicit assumptions about the functional form of h, avoiding the dangers of mismodeling the underlying distribution of the data. For example, suppose our data is highly non-Gaussian but the learning

model we choose assumes a Gaussian form. In that case, our algorithm would make extremely poor predictions.

- **Instance-based** learning means that our algorithm doesn't explicitly learn a model. Instead, it chooses to memorize the training instances which are subsequently used as "knowledge" for the prediction phase. Concretely, this means that only when a query to our database is made (i.e. when we ask it to predict a label given an input), will the algorithm use the training instances to spit out an answer.

Like most machine learning algorithms, the K in KNN is a hyperparameter that we must pick in order to get the best possible fit for the data set. Intuitively, you can think of K as controlling the shape of the decision boundary we talked about earlier.

When K is small, we are restraining the region of a given prediction and forcing our classifier to be "more blind" to the overall distribution. A small value for K provides the most flexible fit, which will have low bias but high variance. Graphically, our decision boundary will be more jagged.

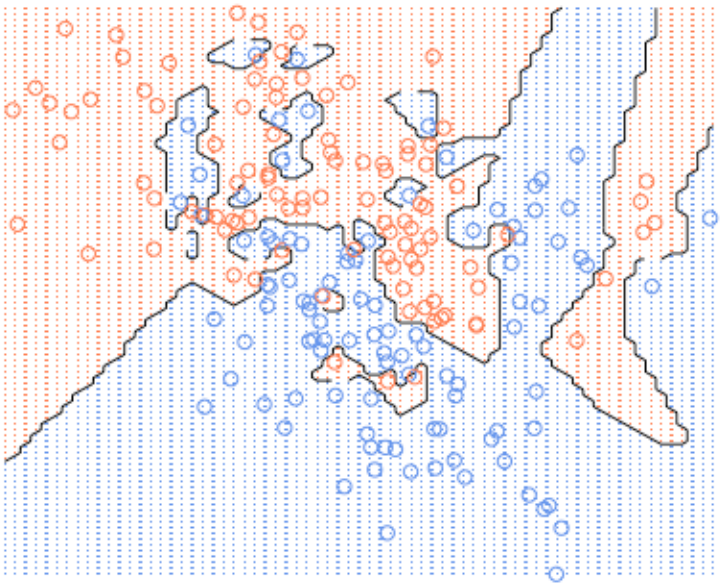

Fig 3.13: k-NN with value of k=1

On the other hand, a higher K averages more voters in each prediction and hence is more resilient to outliers. Larger values of K will have smoother decision boundaries which means lower variance but increased bias.

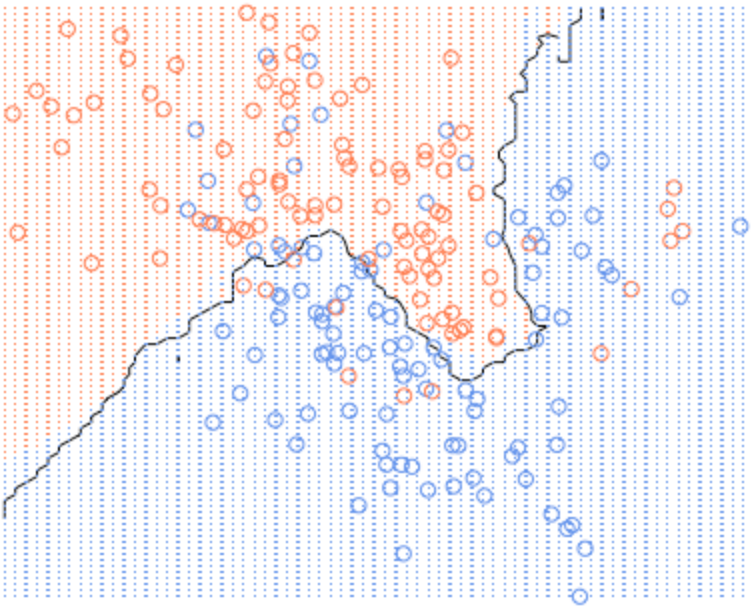

Fig 3.14: k-NN with value of k=20

Chapter 4

Results and Discussions

4.1 Database

In this project database is taken from the MIT-BIH arrhythmia which is accessible from physio net website [8-9]. In the ECG signal recording 2 channels ambulatory contain 48 and half hour signals record which is digitized at 360 samples/second over 11 bits resolution. Each bit is annotated by expert cardiologist which is very important in this signal analysis.

4.2 System Requirements

Beat extraction and feature extraction part of the project was done on MATLAB software. Specifications are:

Operating system: Windows 7 professional (64 bit)

Processor: Intel® Core™ i7-7700 Processor @ 4 x 4.2 GHz

RAM: 32 GB

GPU: NVIDIA GeForce GTX TITAN X

Classification part of the project was done with the following specification:

Operating System: Ubuntu 16.04 LTS (64-Bit)

Processor: Intel(R) Core(TM) i5-2450M @ 4 x 2.50 GHz

RAM: 4 GB

Programming Language: Python 2.7

Machine Learning library: scikit-learn 0.18.1

4.3 Evaluation metrics

For machine learning and deep learning applications, accuracy is not an efficient metric for evaluation of the performance. There are certain other metrics are used which are described below:

4.3.1 Confusion matrix

It is a matrix that describes the performance of any classification system. For confusion matrix C, any element $C_{i,j}$ will represent the number of observations known to be in group i but classified or predicted to be in class j. The confusion matrix is a square matrix that show the count value of the true positive, false positive, true negative and false negative.

Consider the case of simple binary classification where only two classes exist: positive class denoted by P and negative class denoted by N. Confusion matrix for this case can be shown as below:

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Fig:4.1: Confusion matrix

Note:

For multiclass classification problem, as like the case here, the value of TP, TN, FP and FN can be extracted from the confusion matrix as below:

- For any class, total number of examples will be the sum of the corresponding row
  (*i.e.* TP + FN)
- For any class, total number of FN will be the sum of values in the corresponding row
  (excluding TP) while FP will be the sum of values in the corresponding column
  (excluding TP)
- For any class, total number of TN will be the sum of rows and columns (excluding the row and column corresponding to that class)

4.3.2 Precision

It denotes the fraction of prediction which actually have positive class out of the total positive predicted classifications.

$$PRE = \frac{True\ Positives}{Predicted\ Positives} = \frac{TP}{TP+FP} \quad (4.1)$$

4.3.3 Recall

It denotes that of all the samples having positive class, what fraction correctly classified as positive class.

$$REC = \frac{True\ Positives}{Actual\ Positives} = \frac{TP}{TP+FN} \quad (4.2)$$

4.3.4 F1-score

For any classifier, precision and recall should be high. But both the precision and recall can't be high at the same time. Thus, another parameter is used for the analysis called F1-score.

$$F1 = 2\frac{PRE \times REC}{PRE + REC} \tag{4.3}$$

Value of F1-score lies in the range 0 to 1.

4.3.5 Accuracy

It is defined as the ratio of number of correct predictions to the total number of predictions.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.4}$$

4.4 Results

4.4.1 Decision Tree

Confusion matrix

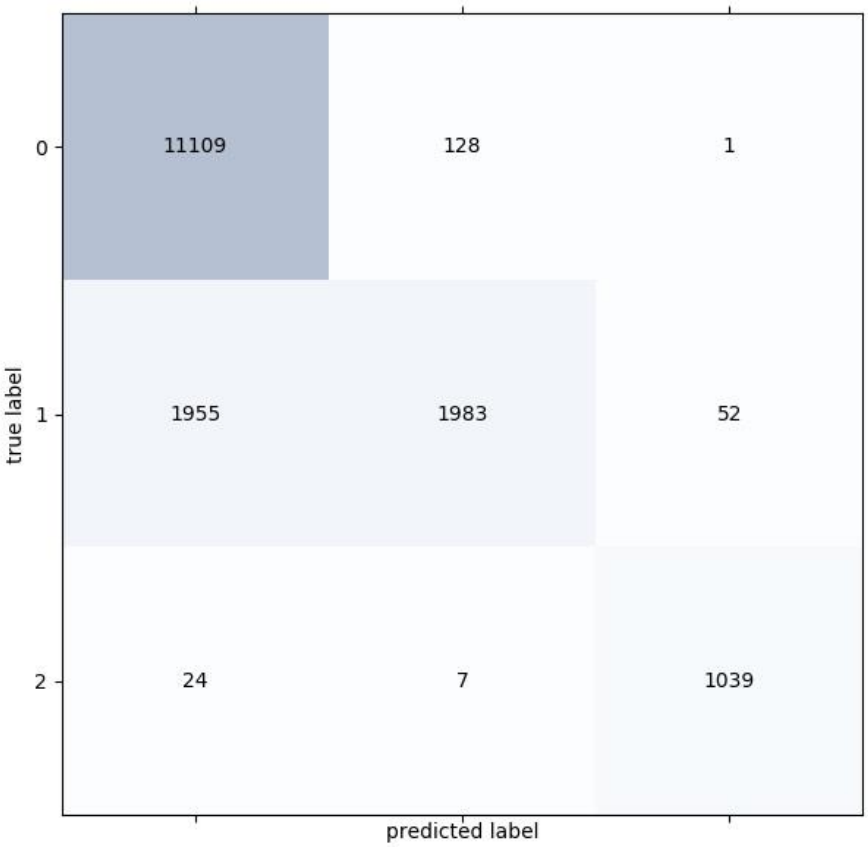

Fig.:4.2: Confusion matrix of decision tree

Precision, Recall and F-1 score

	Precision	**Recall**	**F-1 score**
Normal	0.848	0.989	0.913
Other	0.936	0.500	0.652
SVC	0.951	0.971	0.961

Table: 4.1: Precision, Recall and F1 score of decision tree

Overall accuracy of the decision tree is calculated as 86.70 %.

4.4.2 Multilayer Perceptron

Confusion matrix

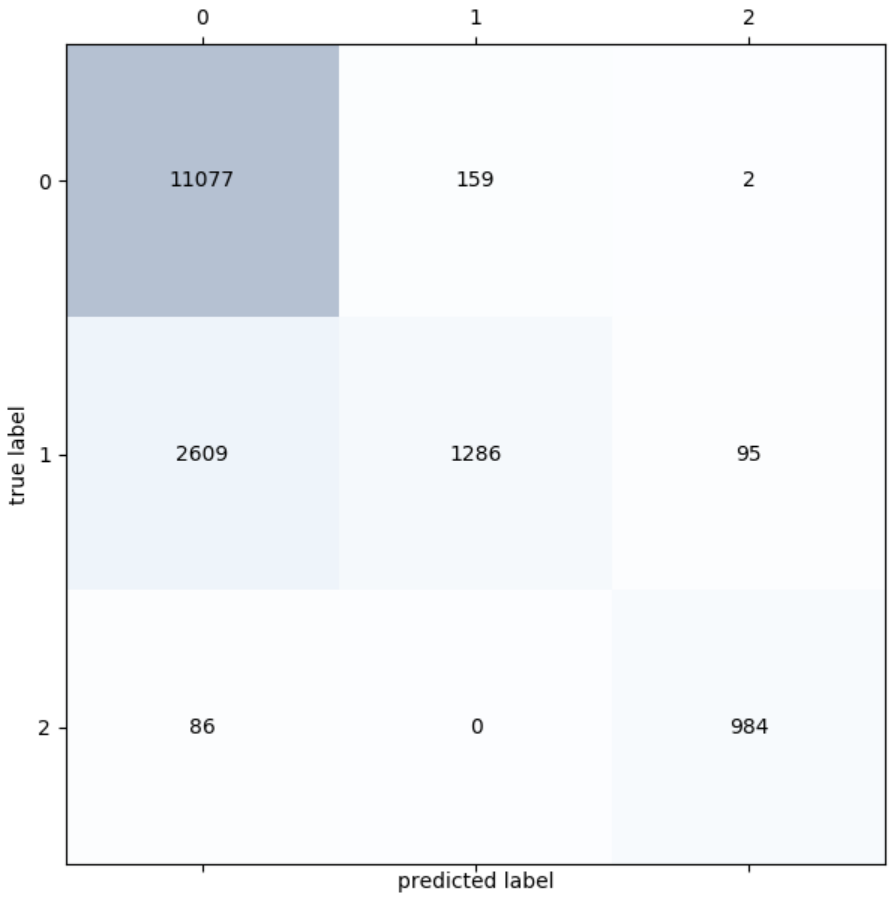

Fig.:4.3: Confusion matrix of Multilayer Perceptron

Precision, Recall and F-1 score

	Precision	**Recall**	**F-1 score**
Normal	0.986	0.986	0.986
Other	0.900	0.322	0.474
SVC	0.910	0.920	0.915

Table: 4.2: Precision, Recall and F1 score of Multilayer Perceptron

Overall accuracy of the multilayer perceptron is calculated as 81.89 %.

4.4.3 Logistic Regression

Confusion Matrix

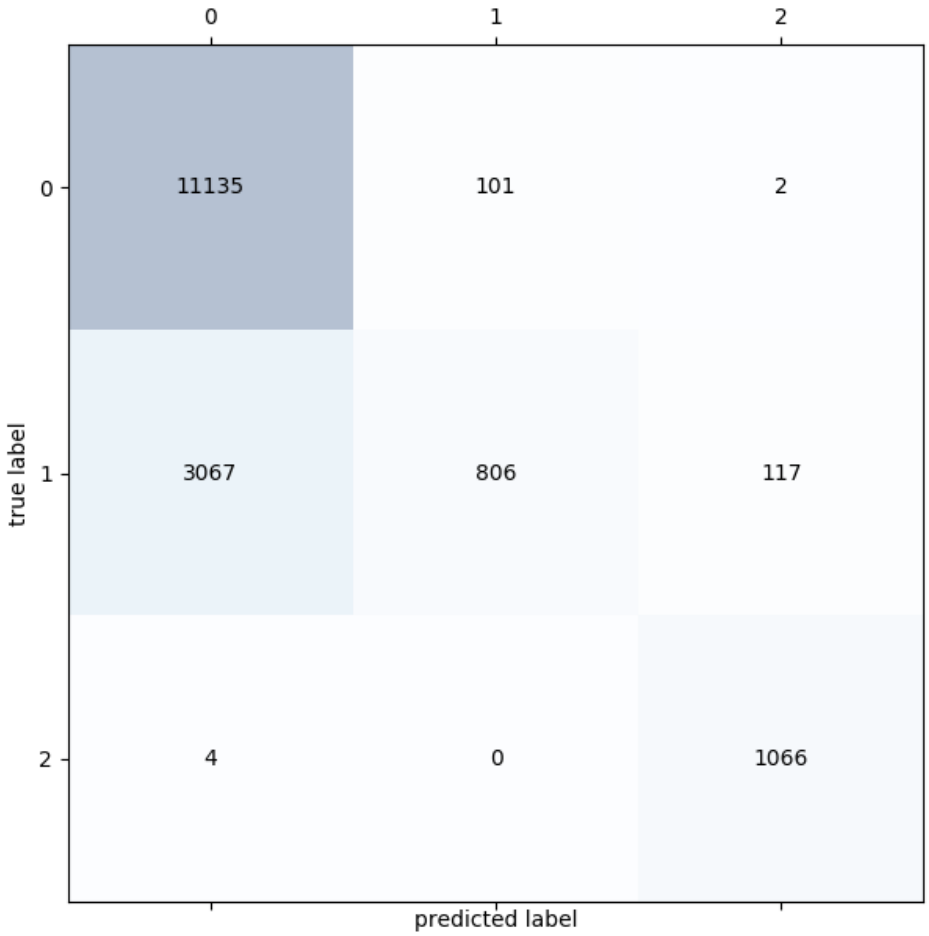

Fig.:4.4: Confusion matrix of Logistic Regression

Precision, Recall and F-1 score

	Precision	**Recall**	**F-1 score**
Normal	0.784	0.991	0.875
Other	0.889	0.202	0.329
SVC	0.900	0.900	0.900

Table: 4.3: Precision, Recall and F1 score of Logistic Regression

Overall accuracy of the logistic regression is calculated as **79.80 %**.

4.4.4 Random Forest

Confusion Matrix

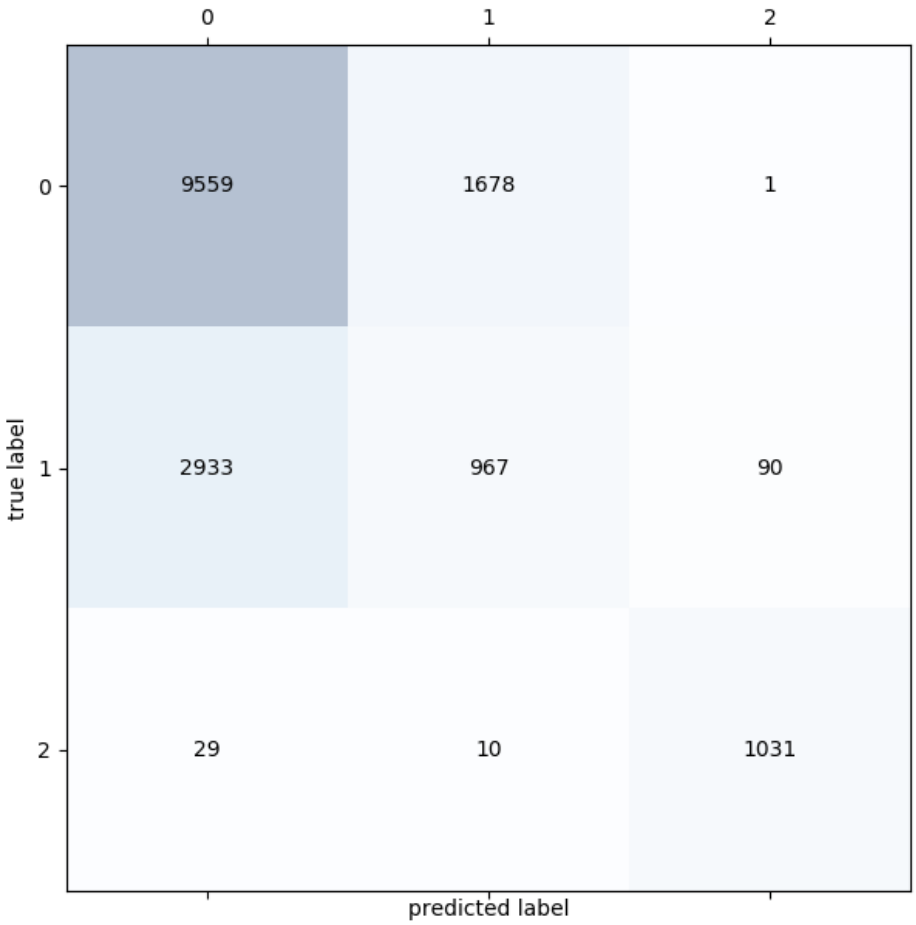

Fig.:4.5: Confusion matrix of Random Forest

Precision, Recall and F-1 score

	Precision	**Recall**	**F-1 score**
Normal	0.763	0.851	0.805
Other	0.364	0.242	0.291
SVC	0.919	0.964	0.941

Table: 4.4: Precision, Recall and F1 score of Random Forest

Overall accuracy of the random forest is calculated as 70.91 %.

4.4.5 Support Vector Machine

Confusion Matrix

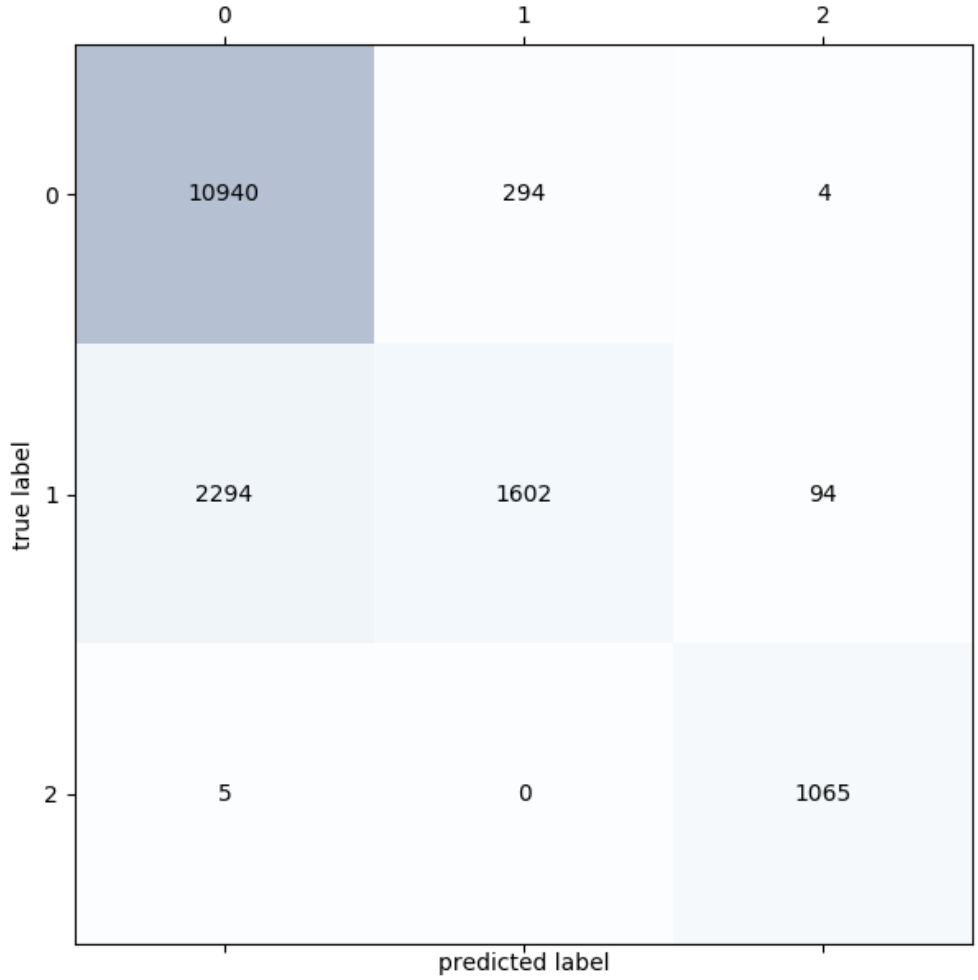

Fig.:4.6: Confusion matrix of SVM

Precision, Recall and F-1 score

	Precision	**Recall**	**F-1 score**
Normal	0.826	0.973	0.892
Other	0.844	0.401	0.543
SVC	0.915	0.995	0.953

Table: 4.5: Precision, Recall and F1 score of SVM

Overall accuracy of the SVM is calculated as 83.48 %.

4.4.6 K-Nearest Neighbors

Confusion matrix

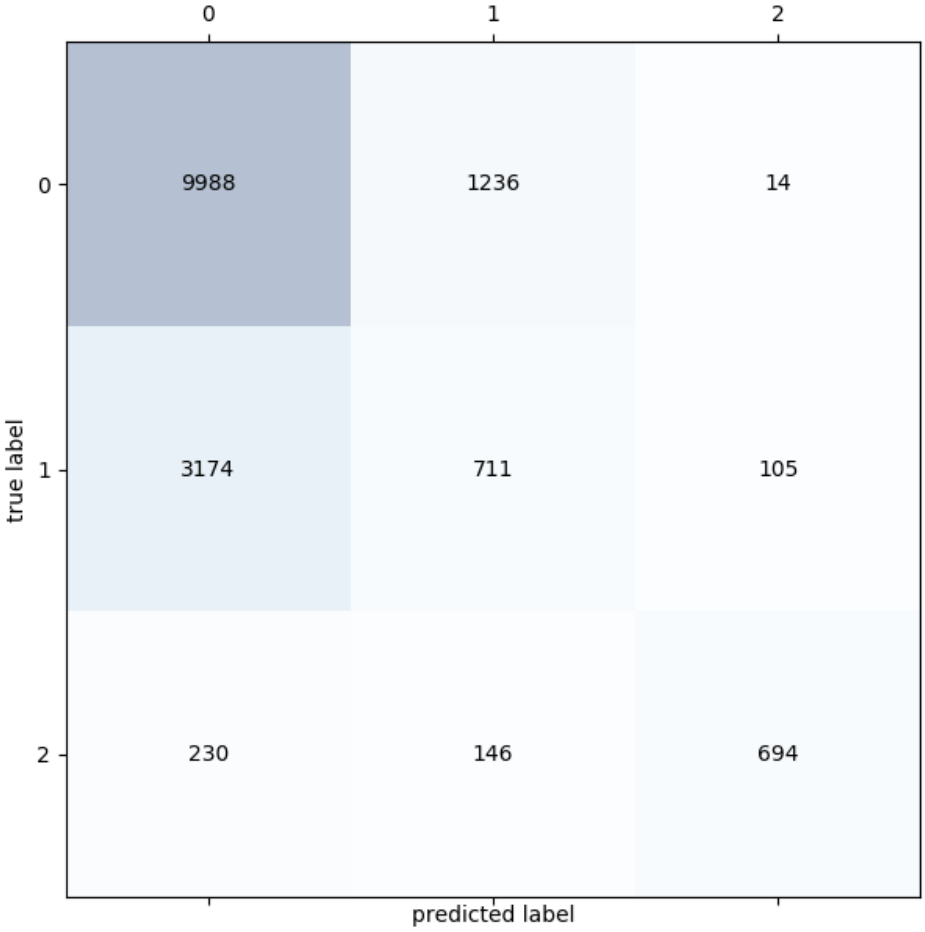

Fig.:4.7: Confusion matrix of KNN

Precision, Recall and F-1 score

	Precision	**Recall**	**F-1 score**
Normal	0.746	0.889	0.811
Other	0.400	0.178	0.246
SVC	0.854	0.649	0.738

Table: 4.6: Precision, Recall and F1 score of KNN

Overall accuracy of the KNN is calculated as 69.90%.

Chapter 5
Conclusion and Future Scope

Conclusion

The main aim of this was to classify the ECG beats as ‘Normal’, ‘SVC’ and ‘other’ with the help of supervised machine learning technique. In particular, I used supervised learning technique like Logistic Regression, Neural Network, KNN, Support vector machine, Random forest, Decision Tree. With the help of proposed method, we easily classify the ECG beat.

From these algorithm Decision tree has the highest accuracy as 86.60%. Even though Logistic regression implemented as a linear classifier still it can classify with a good result for Normal and SVC. From the result, it can be seen that F1 score for Normal and SVC is 0.875 and 0.90 respectively for logistic regression. Because of linear nature of model, it cannot classify correctly labelled as ‘other’. KNN has not good result as a classified as it has low F1 score Normal as 0.811, other as 0.246 and SVC as 0.738.

In case of Decision tree, I can say that the best classifier is this one as seen from the result of precision, recall and F1 score. F1 score in case of decision tree for Normal 0.913, for other 0.652 and for SVC 0.961.

Thus, it can be seen from the confusion matrix that every SVC beat was classified correctly.

Future Scope

An appropriate feature transformation technique will apply to the extracted feature vector in order to improve the performance of classification. Transformed feature vectors will be linearly classified with low dimension to reduce the computation and also it should have decorrelated feature.

Among different patients sometimes the variation of normal and abnormal is considerable and it may lead to misclassification. The best solution for this is to develop an unsupervised learning method to classify ECG beats. An unsupervised learning need not required any predicted knowledge also not required any training, validation and testing of dataset for ECG beat classification.

BIBLIOGRAPHY

[1] Daamouche A., Hamami L., Alajlan N., Melgani F., 2012, A wavelet optimization approach for ECG signal classification, Biomedical Signal Processing and Control, Vol. 7

[2] Nazarahari M., Namin S.G., Davaie Markazi A. H., Anaraki A. K., 2015, A multi-wavelet optimization approach using similarity measures for electrocardiogram signal classification, Biomedical Signal Processing and Control, Vol. 20

[3] Mar T., Zaunseder S.,Martínez J. P. ,Llamedo M., Poll R., 2011, Optimization of ECG Classification by Means of Feature Selection, IEEE Transactions on Biomedical Engineering, Vol.58.

[4] Martis R. J., Rajendra Acharya U., Prasad H., Chua C. K., Lim C. M, Suri J. S., 2013, Application of higher order statistics for atrial arrhythmia classification, Biomedical Signal Processing and Control, Vol. 8

[5] Elhaj F. A., Salim N., Harris A. R., Swee T. T., Ahmed T., 2016, Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals, Computer Methods and Programs in Biomedicine, Vol. 127

[6] Campos Oliveira L. S , Varejao Andreao R., Sarcinelli Filho M., 2016, Bayesian Network with Decision Threshold for Heart Beat Classification, IEEE Latin America Transactions, Vol. 14

[7] Ghorbani Afkhami R., Azarnia G., Ali Tinati M., 2016, Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals, Pattern Recognition Letters, Vol. 70

[8] Martis R. J., Rajendra Acharya U. , Mandana K.M. , Ray A.K., Chakraborty C., 2013, Cardiac decision making using higher order spectra, Biomedical Signal Processing and Control, Vol. 883

[9] Javadi M., Abbaszadeh Arani S. A. A., Sajedin A., Ebrahimpour R., 2013, Classification of ECG arrhythmia by a modular neural network based on Mixture of Experts and Negatively Correlated Learning, Biomedical Signal Processing and Control, Vol. 8

[10] Lamine Talbi M., Ravier P., 2016, Detection of PVC in ECG signals using fractional linear prediction, Biomedical Signal Processing and Control, Vol. 23

[11] C. Kamath, 2011, ECG beat classification using features extracted from teager energy functions in time and frequency domains, IET Signal Processing, Vol. 5

[12] Martis R. J., Rajendra Acharya U., Choo Min L., 2013, ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform, Biomedical Signal Processing and Control, Vol. 8

[13] P. Sharma, K. C. Ray, Efficient methodology for electrocardiogram beat classification, IET Signal Processing, Vol. 10

[14] Maji U., Mitra M., Pal S., 2016, Imposed target based modification of Taguchi method for feature optimisation with application in arrhythmia beat detection, Expert Systems with Applications, Vol. 56

[15] Kim Y. J., Heo J., Suk Park K., Kim S., 2016, Proposition of novel classification approach and features for improved real-time arrhythmia monitoring, Computers in Biology and Medicine, Vol.75

[16] Wiens J., Guttag J. V., 2010, Active learning applied to patient-adaptive heartbeat classification, NIPS'10 Proceedings of the 23rd International Conference on Neural Information Processing Systems

[17] Chazal P., Dwyer M. O., Reilly R. B., 2004, Automatic classification of heartbeats using ECG morphology and heartbeat interval features, IEEE Trans. Biomed. Eng., Vol. 51, no. 7, pp. 1196– 1206, Jul. 2004

[18] Llamedo M., Martinez J. P., 2011, Heartbeat classification using feature selection driven by database generalization criteria, IEEE Trans. Biomed. Eng., vol. 58, no. 3, pp. 616–625, Mar. 2011.

[19] Ye C., Vijaya Kumar B. V. K., Coimbra M. T., 2012, Heartbeat Classification Using Morphological and Dynamic Features of ECG Signals, IEEE Transactions on Biomedical Engineering, vol. 59, no. 10, pp. 2930-2941, Jul. 2012

[20] Mar T., Zaunseder S.Martínez J. P. ,Llamedo M., Poll R., 2011, Optimization of ECG Classification by Means of Feature Selection, IEEE Transactions on Biomedical Engineering, vol. 58, no. 8, pp. 2168–2177, Aug. 2011

[21] Wiens J., Guttag J. V., 2010, Active learning applied to patient-adaptive heartbeat classification, NIPS'10 Proceedings of the 23rd International Conference on Neural Information Processing Systems.

[22] M. Llamedo and J. P. Martínez, “Cross database evaluation of a multilead heartbeat classifier” IEEE Trans. Inf. Technol. Biomed., vol. 16, no. 4, pp. 658-664, Jul. 2007

[23] Llamedo M., Martinez J. P., 2012, An Automatic Patient-Adapted ECG Heartbeat Classifier Allowing Expert Assistance, IEEE Transactions on Biomedical Engineering, vol. 59, no. 8, pp. 2312-2320, Aug. 2012

[24] Al Rahhal M.M., Bazi Y., AlHichri H., Alajlan N., Melgani F., Yager R. R., 2016, “Deep learning approach for active classification of electrocardiogram signals”, Journal Information Sciences, vol. 345, no. C, pp. 340–354, Jun. 2016