Department of Computer Sciences and Software Engineering

Army Public College of Management & Sciences (APCOMS), Rawalpindi

Affiliated with

University of Engineering & Technology Taxila

(Session 2016 – 2020)

Thesis Report

Detection of Cardio vascular Diseases Using Electrocardiogram

By:

ARSLAN SHAFIQUE	(UET-16S-BSCS-19)
CH FAWAD JAVED	(UET-16S-BSCS-17)
WAHAB HASEEB BHATTI	(UET-16S-BSCS-25)

Supervised by

____

ENGR. SADAF KHAN GONDAL

ACKNOWLEDGEMENT

It is not possible for a person to achieve the ultimate goal without the assistance and help of others. Self-efforts also matter a lot, thanks to our ALLAH ALMIGHTY and his great blessings upon us, we are very thankful to our parents and friends and their experience, useful advice and cooperation has helped us a lot to move forward in right path for the completing of your final year project.

We owe our thankfulness and bow our self to THE ALLAH ALMIGHTY for guiding us till the completing of our fascinating and complex final year project. We extend our profound acknowledgement and appreciation to Engr. Sadaf Khan Gondal, who not just give us this chance to do this task yet additionally gave a reasonable consideration to our issues and she demonstrated an unmistakable fascination for our exercises. Great educators are in every case hard to discover however we have been hugely honored by the all-powerful in such manner.

Submissive thanks to Engr. Sadaf Khan Gondal, she demonstrated individual consideration regarding our venture and helped us getting all the stuff. Without her participation, we probably won't have had the option to complete it in time yet she was so enlightening and humble that she never looked exhausted of our occasionally wayward and unessential inquiries.

Our thanks also go out to all our friends, family and other people for morally supporting us during the project. Our appreciation also goes to the people who have consistently and willingly helped out with their skills and abilities.

UNDERTAKING

This is to declare that the project entitled "**DETECTION OF CARDIO VASCULAR DISEASES USING ELECTROCARDIOGRAM"** is an organic work done by signatories, in partial fulfillment of the necessities for the degree "Bachelor of Science in Computer Science engineering" at Computer Science Department, Army Public College of Management & Sciences, affiliated with UET Taxila, Pakistan. All the analysis, design and system development have been accomplished by the signatories. Moreover, this project has not been succumbed to any other university or college.

Date: - 1 - 2020

Arslan Shafique
Ch Fawad Javed
Wahab Haseeb Bhatti

BSCS-25

CERTIFICATE

This is to certify that the project titled "**DETECTION OF CARDIO VASCULAR DISEASES USING ELECTROCARDIOGRAM"** is the Bonafede work carried out by **Arslan Shafique**, **CH Fawad Javed**, **Wahab Haseeb Bhatti** who are the students of Bachelor of Science in Computer science(BSCS), from Army Public College of Management & Sciences, which is affiliated with university of Engineering and Technology, Taxila .Throughout the academic year 2016-2020, in partial fulfillment of the condition for the award of the degree of Bachelor of Science in Computer Science (BSCS), and that the project has not formed the foundation for the award before of any other degree, diploma, companionship or any other alike title .

Signature of the Supervisor

____

Date: - 1 – 2020

ABSTRACT

Classification of electrocardiogram (ECG) signals plays an important role in diagnoses of heart diseases. In this project of thesis, we offered a computerized Artificial Neural Network (ANN) based Classification of Atrial Fibrillation, Ventricular Fibrillation and Normal. The proposed model was trained and tested on MIT-BIH arrhythmia database. ECG recording of 26 datasets of patients with paroxysmal or sustained atrial and ventricular fibrillation. For the preprocessing the signal had to go through multiple noise removal techniques like moving average filter, simple low pass FIR filter, Butterworth high pass filter, low pass Butterworth filter. The ANN model is fed with the 30,450 training examples and validated across 6,525 examples and then tested with 6,525 examples with the accuracy of 0.99%.

Table of Contents

List of Tables

List of Figures

Chapter 1

1. Introduction

1.1. Background Information

In the medical field of cardiology, the cardiologist study human's electrical activity of heart to distinguish cardio diseases from patient's ECG. Myocardial infarction is cardio Problem that occurs because of the obstruction in the trail of coronary blood vessels which deliver the blood to the cardio muscles. Irregularities in cardio muscles can be recognized via difference in patient's ECG.

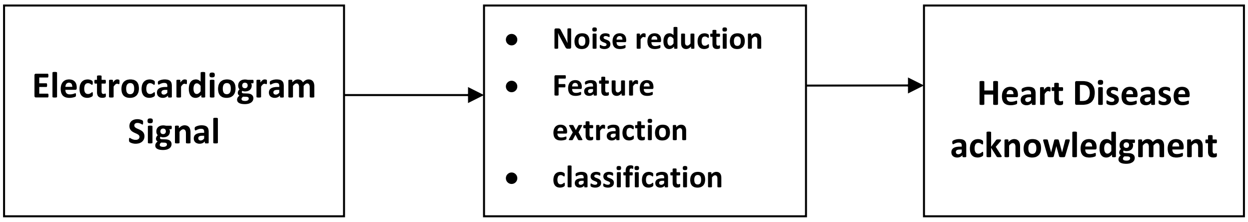

Figure 1. Myocardial infarction detection process.

1.2. Problem Statement

To propose a system that will detect cardiovascular diseases (Atrial fibrillation, Ventricular fibrillation).

1.3. Motivation

Life is vital for everyone, a little carelessness can be a cause of life loss. Doctors are busy in big cities or in emergency and may intentionally neglect the reads from ECG. It will assist cardiologist to detect early or already having cardio diseases. It will be easy to use. In future it can provide assist on multiple cardio vascular diseases.

1.4. **Aims and Objectives**

The main objective:

- To propose a system that will help cardiologists and emergency staff to detect myocardial infarctions (Atrial fibrillation, Ventricular fibrillation).

1.5. **Scope of Project**

- System will accept only digital signal
- System will only cover some of the MI (myocardial infarctions) including (atrial fibrillation, ventricular fibrillation)

1.6. **Breakdown of Project**

Table 1: Breakdown of Project

Task Name	Start Date	End Date	Duration
Problem Statement Analysis	05/03/19	15/03/19	10d
Dataset Gathering	16/03/19	06/04/19	15d
Existing systems Analysis	07/04/19	07/05/19	30d
Literature Review	08/05/19	08/06/19	30d
Model Design & Analysis	09/07/19	09/08/19	30d
Proposed Model Implementation	10/08/19	10/10/19	120d
Model Training	11/10/19	11/11/19	30d
Model Testing	12/11/19	12/12/19	30d

Error Evaluation	13/12/19	12/01/20	30d
GUI development	13/01/20	23/01/20	10d
Result Analysis	23/01/20	07/02/20	15d
Documentation	08/02/20	05/03/20	27d

1.7. **Summary:**

Detection of myocardial infarctions via an ECG includes three steps, that include doing the preprocessing on signal then doing feature extraction and lastly classification. The initial phase in the preprocessing which include removing of noise from the given ECG digital signal and make it less noisy so next step become easy. Next stage is training and testing of digital signal.

Chapter 2

2. Literature Review

2.1. Problem domain

The doctors use the electrical activity that are generated by the human heart to detect cardio diseases from the electrocardiogram. A myocardial infarction (MI) is a cardio disfunction that happens when there is an obstruction in the pathway of one or more coronary blood vessels (veins/arteries) that deliver the blood to the heart. The irregularities in the Cardio muscle can be acknowledged by the difference in the ECG signal.

Life is vital for everyone; a little carelessness can be a cause of life loss. Doctors are busy in big cities or in emergency and may intentionally neglect the reads from EKG. It will assist cardiologist to detect early or already having cardio diseases. It will be easy to use. In future it can provide assist on multiple cardio vascular diseases.

Hence, we want to propose a system that will detect cardiovascular diseases (Atrial fibrillation, Ventricular fibrillation). There is a massive research carrying out through decades in the field of Medical.

2.2. Existing work

The human ECG is the complex and versatile of all signals. Ventricular fibrillation (VF) and are dangerous and risky arrhythmic diseases leading to unavoidable demise if no defibrillation shock is applied to the subject within a few minutes. [1]

In [1] author have worked by using single bit classifier with adaptive technique, they have detected ventricular Arrhythmia and atrial fibrillation. First using prepressing and then extracting feature from the ECG. Firstly, in preprocessing they have used bandpass filter then second task is the QRS detection using PAT algorithm. The third step of preprocessing phase is T & P wave delimitation. While designing a detection system for the feature extraction two main factors were considered, the difficulty and correctness of the feature extraction stage to provide the best result A set of six attribute were selected, which showed great importance in detection of arrhythmia. The features are P, RR, T, TP, PR width and QRS width. A single-bit classifier is considered to spot arrhythmia.

In [2] author investigated VT via heart beat annotation using Pan Tompkins algorithm. They used five different hypothesis that help them in getting the right results. The first was measuring the RR intervals in the measured ECG. Second hypothesis was that peak in heart rate appear at same time as of RR peaks. The next theory tells regarding the form of the QRS waveform. The QRS wave in the ECG signal turn out to be broader. the 4th theory, states if the signal power spectrum standard deviation more than the threshold set manually. The 5th theory which states that the low level of coherence among the ECG and V waveform signals may show that the ECG waveform is too disturbed and should be not used for alarm incident inspection the false negative, true positive, false positive and true negative rates are Used to measure the output result.

]. In [3] the author has used three step processing GUI, starting with taking the ECG and applying Preprocessing with Bandpass filtration. Then PSD analysis is gained followed by feature extraction and classification using threshold values. The Bandpass filter allows the single to be passed only if it lies between under certain regularity. Then they use a power spectral density which is used for dissimilar topics to sense the alteration.

Then on the bases of PSD the feature is extracted and threshold classification is applied to acquire the result. The QRS segment complexes are then detected. Different features are extracted such as the form factor, heart rate, kurtosis and ratio of LF and HF is used to distinguish between non-infected and infected.

M. Botsivaly et al [4] proposed a method according to which every digital signal was taken from the database and then the digital signal was converted into an image that was divided in different segments and further the segment was divided into four areas. These areas were defined by the Vmax limit, VZ max, -1/2 and 0.the regular sinus these ROI paper bigger than normal pixel density about the isoelectric band that corresponds to the Vmax.O, Vmax regions.

The images have nearly persistent pixel circulation across all the ROI. Conservative Detection Methods include Threshold Crossing Intervals and Spectral Analysis in which the Hz of VF is calculated which is between 4 to 7 Hz.

In [5] the author has used machine learning, which is used on electro cardiology. For including data compression Wavelet transform is used, examination of ventricular late possibilities, and the recognition of ECG signal distinctive attributes. The wavelet transform can separate high frequency component a low frequency component in time domain. In addition, wavelet transform is able to decompose signals at various resolutions. It allows accurate feature extraction from non-stationary signals like ECG.

For these two values were entered, initial value signify number of example and other value is size of the window. when wavelet transform is applied the ECG signal, the next step is the R peaks detection. The position of the other peaks is found with position of R peaks. calculated RR and PR intervals. These intervals are very important for detecting diseases. These intermissions are significant for defining these kinds of disease.

Heart wellbeing is obligatory for presence as it is the most basic organ that fills in as the middle for blood flow. Atrial Fibrillation (AF) is a kind of arrhythmias with hindered atrial exercises bringing about harms to atrial mechanical capacities Basic Features through which AF is identified on electrocardiogram are unpredictable RR interim examples and absence of clear, detectable P waves.

According to **[6]** The electrocardiogram basic shape consists of three waves: P, QRS, and T Wave. A chamber experiencing AF would experience issues in contracting in this manner delivering irregular P-wave. In QRS wave is an aftereffect of ventricle constriction that pumps blood into the remainder of the human body. QRS complex has the higher profusion in the event of AF.

In the interim, T wave is an aftereffect of ventricular unwinding process when the compression closures, and blood starts to siphon from the chamber into the ventricle. In [6], the creator has proposed KNN (k closest neighbors) to group kinds of AF maladies. KNN is an order strategy which utilize past information of articles like it to characterize. KNN is a kind of directed understanding calculation, KNN expects to arrange new articles dependent on the characteristics and tests of preparing information.

The principle undertaking of arrangement performed by K-NN is deciding attribute K choices. The finest K esteems for the KNN calculation change contingent upon information. By and large, higher K esteems will diminish impacts on grouping, yet they additionally obscure the limits between every arrangement. Consequently, parameter advancement required in deciding the ideal K esteem.

The element extraction is done in following advances:

a) R peak signal data input and finding R peak XY coordinates.
b) Finding RR interval distance difference.

Extracting the most dominant number of RR distances.

K-Nearest Neighbor

1. Preparing the data input and Testing it.

2. Deciding k parameter.

3. Estimating testing information separation against preparing information dependent on every datum measurement with Euclidean Distance technique.

4. Arranging separation estimations in climbing request.

5. Deciding closest neighbors dependent on the quantity of k determined. On the off chance that the aftereffect of climbing request does exclude the closest neighbors, at that point the information can't be sorted.

6. Deciding classifications from closest neighbors.

7. At long last is utilizing greater part classifications from the closest neighbors as expectation esteems for new information.

Utilizations parameter with Euclidean separation strategy for separation estimation with 54 information. An AF is distinguished utilizing RR interim with five heartbeats ECG sections and differentiator limit at one hundred heartbeats is developed to identify tops and lessen inadmissible R top identification. The framework structure is accomplished utilizing K-NN characterization techniques. This examination utilizes parameter with train/test gap to acquire the most extreme exactness. K-NN calculation execution is estimated utilizing disarray network. In light of the exactness of the general plan, the best outcome is k = 1 with a normal precision at 91.75% and the most noteworthy precision, affectability, explicitness level at

95.45%, 91.67%, and 100% with train/test split information at 60:40 percent. The outcome shows that heart AF discovery with K-NN strategy is best for restorative assessment. Thus, it is required to decrease rate number populace of AF.

Author **[7]** introduced Discrete wavelet transform constructed component withdrawal; the tops of R are predicted to decide the HRV signal highpoints. Beats classification is achieved to identify irregularities in ECG signal using Support vector machine classifier. R-top recognition system is predominantly founded on heart rate which is usually used to determine the RR provisional. Dividing one moment by the rapid pulse gives the RR provisional of the given signal. Consecutive RR breaks of the signal are determined from start interim of the beat work. Among the Meyer, turn around.

Coiflet wavelets and Biorthogonal, Coiflet is chosen to eliminate R-tops. In handy RR interim approximation framework, joining approach with timing goals ± 1 ms is used. The accurate RR interim approximation can be developed by best progressive sign processor or different processor. Substantial measure of design in signal is disqualified using the multipurpose filtering based preprocessing.

frequency domain and Time realm structures can be consequent from the mined HRV attribute. .It portrays the mechanisms involved in the arrhythmic beat classification based on preprocessing and HRV attribute withdrawal. The evaluation consists of

i) Assortment of raw electrocardiograms signal from MITBIH.

ii) Signal Preprocessing utilizing filters.

iii) Detection of the R-peak.

iv) Frequency area attribute withdrawal from HRV signal

v) Support Vector Machine classification.

The acquired result of the preprocessing, highlight removal and characterization are utilized to close the planned system. In signal preprocessing, DENLMS calculation constructed on versatile filter is used to get better-quality filtering implementation with little computational trouble. In R-top acknowledgment, Coiflet wavelet is used to mine everything the conceivable R-pinnacles and offers increasingly exact heartbeat frequency. The picked-up heartbeat rate and HRV Frequency space highlights are applied to SVM classifier for arrhythmic beat classification which is less difficult than other AI draws near. Different classification strategies dependent on PCA, ANN, information-based framework, KNN and SVM are utilized utilizing parameters, for example, ECG and HRV. The most extreme classification precision of 94.2% is obtain. Be that as it may, the test consequence of SVM based classifier gives a greatest exactness of 96 % on grouping ordinary and arrhythmic hazard unusual subjects. (C. VENKATESAN, 2018)

In [8] creators stress that spikes and dunks in the line tracings are called arrangement of waves. These arrangement of waves comprise of six distinct waveforms, are recognizable, and are separated as P,Q,R,S,T and U. The vast majority of previous strategies for ECG signal investigation for distinguishing PQRST were grounded on DSP (digital signal processing) strategy for instance (Wavelet Transform and Fast Fourier Transform) and Artificial Neural networks. While in this paper it proposes a basic strategy to recognize the P, Q, R, S and T estimations of an electrocardiogram (ECG) signal. This strategy depends on finding a scientific connection between the most noteworthy qualities (pinnacles and valleys) of the ECG waveform and time. In this proposed technique is exhibited by planning a graphical UI (GUI) by utilizing MATLAB for recognizing PQRST by utilizing straightforward scientific calculation to get PQRST qualities and draw these qualities on ECG wave simultaneously. This paper is increasingly disposed to the reasons for logical research rather than clinical finding.

In this work lead determination alternative is given to contribution of information alongside document type choice. After that low recurrence parts are evacuated and afterward a windowing channel and afterward threshold and afterward again modifying channel (window separating) and distinguish R-tops to identify pulse and afterward applying basic scientific computations by utilizing MATLAB conditions to ascertain P, Q, R, S and T.In **[9]** authors emphasize on detection of RR intervals considering them as most important for detection of arrhythmia. Hence correct identification of this interval is very essential for making accurate decision. Normal sinus rhythm was detected in four subjects, Atrial Fibrillation (AF), and arrhythmia are taken into consideration by examining the distribution of both statistical features and power features that includes LF/HF ratio, Heart Rate (HR), Kurtosis, Form Factor (FF), and Skewness. One of the major features is the affected ECG signal has RR interval not equal to that of normal heart beat. Additional feature which was seen was that skewness, heart rate and kurtosis are higher but LF/HF ratio is lower for affected ECG signals. The resulting output were analyzed and compared with real reports of patients that were obtained from medical. The outcome offers major benefit, and matches the doctor's judgement.

In this paper **[9]** a prediction technique is proposed which was divided into three modules such as

1. Data acquisition & Data filtration.
2. Power Spectral Density analysis.
3. Feature Extraction and Classification.

Band-pass filter was used with frequency range of 2-250Hz for noise reduction, this resulted into the identification of QRS complexes of ECG signals.

The platform manages to stores the data and provides assists to affected patients to observe their own condition of heart. In addition, the platform provided visual representation to observe the infected and non-infected parts of ECG signals and analyze results. To observe the variation in heart rate variability for large number of patient's, more research work is required to increase accuracy. Further research work is required to acquire acceptable results by the platform for clinical implementation.

Table 2: statistical features for different subjects

Subjects	Heart Rate (bpm)	FF	Kurtosis	Skewness	LF/HF ratio	Status
S1	70.8051	3.7898	1.3749	0.0864	0.9210	Normal
S2	88.9621	5.3844	1.8045	0.2347	0.8140	Infected
S3	110.2857	7.3344	1.9310	0.3955	0.7398	Infected
S4	98.3163	7.0107	4.7611	1.6646	0.5991	Infected

In **[10]** research paper aims to develop an automated model which outclass the cardiologist performance in diagnosing a wide range of heart diseases from ECG records of single-lead. Researcher used a large descriptive dataset and a Deep-CNN which was able to generalize sequences of ECG samples and acquired cardiologist level of accuracy for cardiac diseases detection with CNN. Researcher clearly stated that the future work could be done by training the model on different classes of arrhythmias or any other type of heart disease, by acquiring

ECG records from single or multiple lead a high accuracy can be achieved. For example, system was not trained Ventricular Flutter or Ventricular Fibrillation so it cannot detect those diseases also system didn't detect Left or Right Ventricular Hypertrophy, Myocardial Infarction or a many of other heart diseases which didn't fall under the category of arrhythmia. There are some diseases which are difficult to detect on single lead ECG signal but can be detected through multiple-lead ECG signals. As it is known that more than Three hundred million ECG signals are acquired annually, highly accurate diagnosis from ECG can be achieved easily.

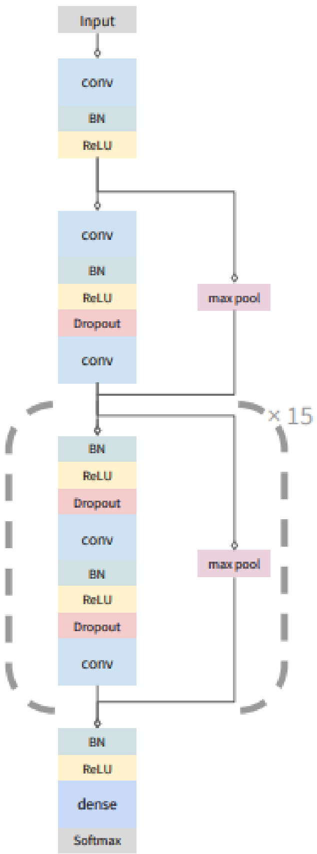

Figure 2: CNN block diagram

In **[11]** the author has used a deep Convolutional Neural Network which has been previously trained on millions of images. The deep CNN is used for feature extraction from ECG signals. On classifying a small dataset of 7000 samples of ECG signals containing four classes of rhythms, 97.23% accuracy was achieved, Model summary was a pre-trained 161 layered Dense Network was used for feature extraction and Support Vector Machine was used for classification. The results show that using deep CNN based spectrograms to transform ECG signals into the images preserved their fine-grained details. Secondly features learnt from such a big amount of generic data in a deep neural network can act very well to represent spectrograms of ECG signals.

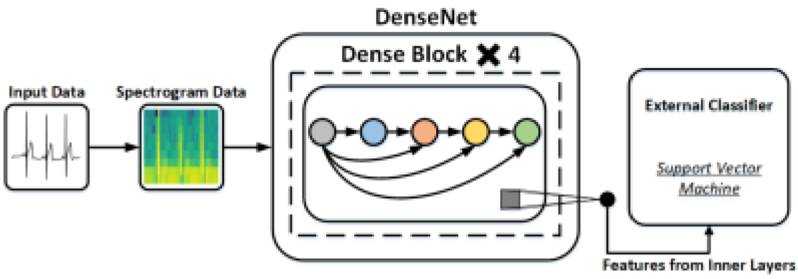

Figure 3: Dense Net block diagram

In **[12]** Ventricular Fibrillation or VF is a life-threatening arrhythmia which could result into cardiac arrest which may cause death of the patient. Doctors cannot continuously monitor each patient which motivated the author to develop an automated system to perform the task.

It was mandatory that the system should have detected patients not affected by Ventricular Fibrillation i.e. system should have higher specificity. This defines the system to prioritize on detecting 'Not VF' class correctly but the accuracy of 'VF' class detection was low. Such a system cannot be considered rational if it fails to detect VF accurately as the objective is to detect VF, this may result into the death of the patient. The goal of this research is building an assisting tool for the doctors. Researcher combined the algorithms of both signal processing and machine learning i.e. they have used Support Vector Machine and Random Forest Algorithm for classification. This algorithm successfully resulted 99.99% sensitivity, 98.40% specificity and 99.20% accuracy. System also resulted 99.19% G-Mean Accuracy, on a 5 seconds long time interval. The increase in performance was seen when the window length was increased.

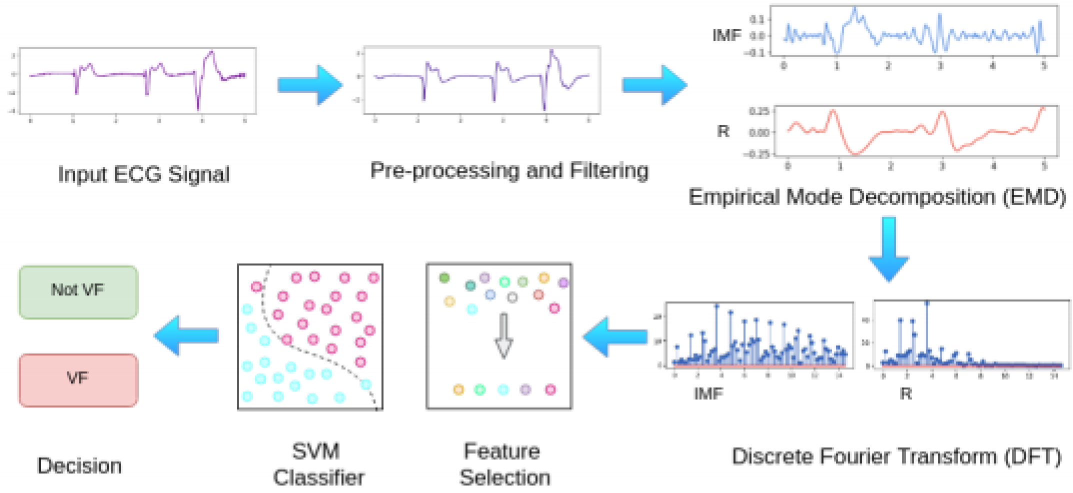

Figure 4: output system diagram

[13] Research paper is about the preparation of a Clinical Decision Support System a.k.a. CDSS is performing multi-class classification using ECG signals to categorize certain cardiac diseases. The approach being used is machine learning algorithm i.e. Artificial Neural Network and also Discrete Wavelet Transformer (DWT) for features extraction. This CDSS categories ECG signals into five classes. For ECG signals MIT-BIH Arrhythmia Database has been used. The study of the proposed work revealed that it has an accuracy of 93.8%. **(Hela Lassoued, 2018)**

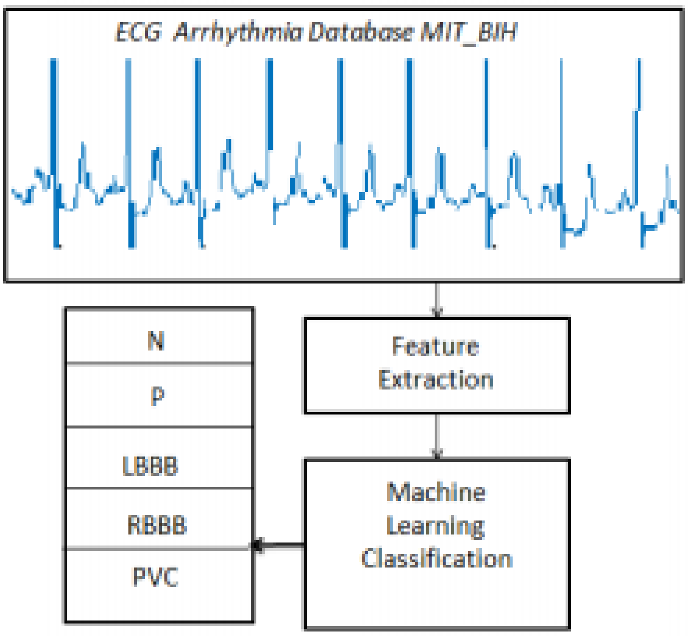

Figure 5: Discrete Wavelet Transformer

2.3. State of Art methods

Atrial Fibrillation is shown by sporadic beats in the Cardio electrical framework from the chamber into the ventricle. On this examination, an AF location framework dependent on RR interim with five heartbeats ECG sections and differentiator edge at one hundred heartbeats is developed to recognize tops and limit unseemly R top recognition. The technique is expected to recognize the distinction among AF and ordinary heart flag additionally can decrease the quantity of information measurements. There are three phases in this examination; they are pre-handling as a procedure of unfirming information measurement, include extraction, and K-NN order. Highlight removal applied by looking at the R to R interim of Atrial Fibrillation sign and the ordinary one.

Framework structure is utilizing K-NN grouping techniques. This examination utilizes parameter with Euclidean separation technique for separation estimation with fifty-four information. In preparing and testing information division, train/test split is done creating an information. The test and preparing information utilized are irregular. Exactness can be looked at dependent on said train/test split-NN calculation execution is estimated utilizing disarray lattice. In light of the exactness of the general plan, the best outcome is k = 1 with a normal precision at 91.75% and the most elevated precision, affectability, particularity level at 91.67%,. It infers that heart Atrial Fibrillation finding with K-NN technique is sufficient for therapeutic assessment. In this way, it is relied upon to decrease rate number populace of Atrial Fibrillation.

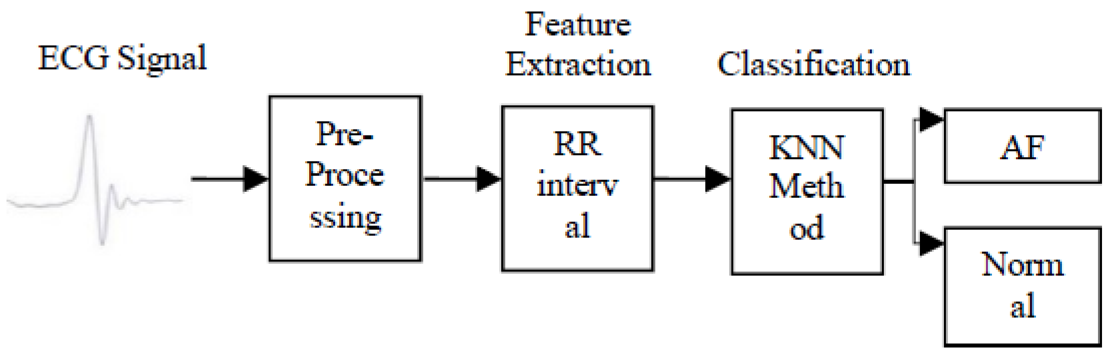

Figure 6: KNN Block diagram

Table 3: result of different K values on different splits

K	60:40	50:50	80:20	Average
1	95.45%	88.89%	90.90%	91.75%
3	95.45%	85.18%	90.90%	90.51%
5	90.90%	85.18%	90.90%	88.99%
7	86.36%	85.18%	81.81%	84.45%
9	81.81%	74.07%	81.81%	79.23%
11	81.81%	70.37%	81.81%	78.00%

2.4. Comparison and Conclusion

Development of an automated system that accomplishes recognition of Cardio vascular Diseases Using Electrocardiogram is difficult. Various approaches have been made toward robust of Cardio vascular Diseases detection, by different signal trait detection, and feature extraction, analysis and classification methods. We have overviewed the different methods of detection of cardio vascular disease. Feature extraction is important stage for expression recognition system because extracted feature is used for classification stage.

Table 4: Comparison of Different Extraction Segmentation Techniques

Feature Extraction Method	**Description**	**Advantages**	**Disadvantages**
Using single bit classifier with adaptive technique S	Work by dividing the elements into set of two group on the basis of a classification rules as defined	Super simple as compared to another classifier. System can be proficient on less training data.	If prediction values are very near the boundaries then error metric may be very sensitive to false positive or false negative
Pan Tompkins algorithm	Derivative function is used in pan and Tompkins algorithm which highlights the slopes of R-wave. additional improve the high regularity features of the QRS complex.	Greatly influence the QRS detection as compared to other.	Used only for QRS detection.

K-NEAREST NEIHBOR (KNN)	The feature extraction is carried out in following steps: a. R peak signal data input and finding R peak XY coordinates. b. Finding RR interval distance difference c. Extracting the most dominant number of RR distances.	Using different values of k with Euclidean distance method Used to detect RR interval with five pulses ECG segments and differentiator threshold at one hundred pulses is constructed to detect peaks and reduce unsuitable R peak detection	The accurate RR interval measurement cannot obtain without high performance digital signal processor or customized processor Baseline trend is still present
Discreet Wavelet Transformation based attribute withdrawal	The R-peaks are spotted to check the HRV signal attribute. Arrhythmic beat classification is achieved to perceive irregularities in signal using Support Vector Machine classifier	1.DENLMS algorithm created adaptive filter is used to get better-quality filtering performance 2.little computational struggle , Coifed wavelet is used to mine all the likely R-peaks and offers more exact beat rate.	Too much computation required

Convolutional Neural Network (CNN)	Used for the sequence-to-sequence learning tasks.	Does not requires manual feature extraction,	Requires a lot of computational power. Complex structure.
Random Forest	It is an ensemble learning algorithm that employs a collection of decision tree classifiers.	It can also be used to determine the importance of the features and thereby rank them accordingly.	Large number of trees can make the algorithm to slow and ineffective for real-time predictions.
Discrete Wavelength Transformation (DWT)	It decomposes into a set of wavelets (functions) that are orthogonal to its translations and scaling	DWT uses Discrete values for the scale and translation factor	Computationally intensive Less efficient

2.5. Problem Solution

There is a great need to detect the different variation in cardio vascular disease and a system that accurately identify these changes. Latest work is on machine learning. Various techniques have been used for detection of these characteristics. The feature abstraction technique plays a very important role in the precision of recognition so choosing them is our first priority to improve the accuracy of our system as much as possible. So, we focus on performance of feature extraction techniques and analyze its accuracies and efficiencies in recognition of detection of various characteristics of cardio vascular disease. System will accept only digital signal, System will only cover some of the myocardial infarctions including atrial fibrillation, ventricular fibrillation. The initial phase in the preprocessing which include removing of noise from the given ECG digital signal and make it less noisy so next step become easy. Next stage is segmentation of digital signal into more simple samples and then segments of R peak are obtained.

Chapter 3

3. Methodology

3.1. Introduction

Today Computer and technology is playing a vital role in better development of biomedical field. Technology is providing fast and reliable way to detect and find cures for various diseases.

Any object is detected by its attribute or features. A better recognition depends upon the quality of feature selected to determine the entity. Cardiology is the field of medical which relates to study of cardiac system, it includes the study of the electrical movement of the human cardio muscle in direction to detect various heart conditions. Doctor uses the ECG signal which is an inspection which records the electrical activity of heart to show whether or not it is working normally. An ECG records the heart's rhythm and activity of the heart.

A myocardial infarction is heart problem that happens when there is any obstruction between path of more or one coronary blood vessels that provide the blood to the heart. The irregularities in the cardio muscle can be predictable by the difference in signal.

We are working with the ECG. The ECG disease is detected using an Artificial Neural Network. ANNs is deep learning computing systems which were created to model the working of human brains. Such systems can gradually learn to perform tasks by getting training from training samples and process complex inputs and make decision by themselves. ANN itself isn’t an algorithm but rather it provides framework for many algorithms. The algorithm we are using is called Backpropagation. The Backpropagation algorithm points for the least value of the error function in weight space using a method called the gradient descent. The weights that minimize the error function are then considered to be answer to problem.

Our Project aims to distinguish two major cardiovascular diseases Atrial fibrillation and Ventricular fibrillation. The system will take digital signal as input from the user and by using various features vector.

3.2. System Requirements

System requirements are the configuration that a system must have in order for hardware of software application to run smoothly and efficiently.

Following are the System software and hardware requirements.

Hardware Requirements:

- Core i3 or Higher Processor with 4GB Ram (DD3) or higher

Software requirements:

- Windows OS (window 7 or higher)
- Python
- Digital Signal

User Interface Requirements

The user interface will provide a button for “Select ECG signal”, clicking the button will open a new window which will allow user to upload ECG Digital Signal.

The user interface will provide an image view to show the Signal as input.

The user interface will provide a Button for “Detect”, clicking it will run the algorithm on the ECG and will tell which disease is detected.

The user interface will provide an image view that will also segment the portion of the ECG wave where it will detect the variation.

3.3. **System Overview**

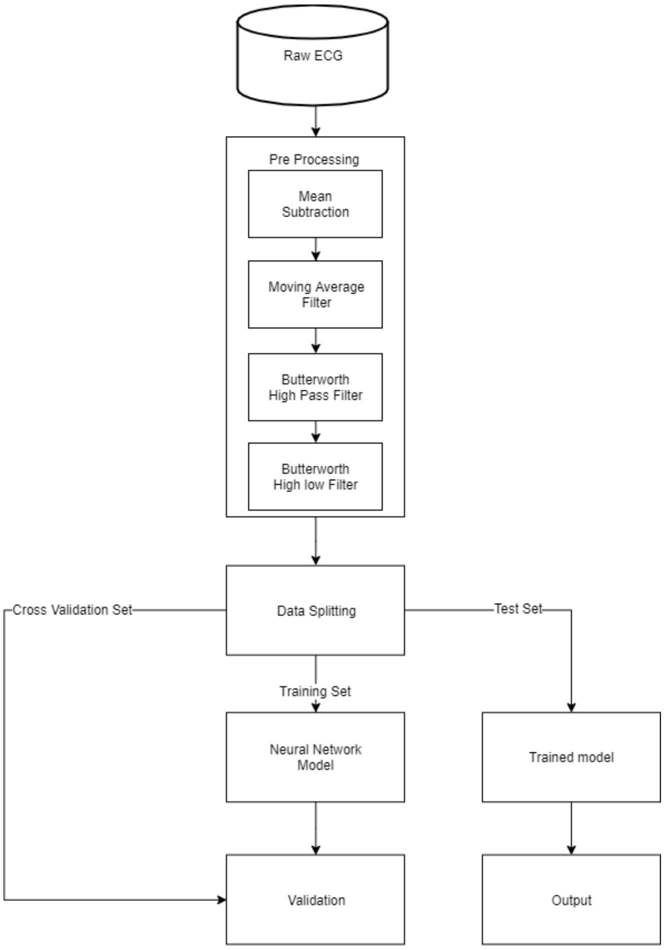

Figure 7: System Overview Diagram

3.4. **Inputs**

Input is a digital signal of ECG will be taken from the database of physionet. Databases of the ECG signal recordings from the MIT physionet-ATM, the signal is sampled at 250 Hz. Then system will fetch the digital signal. The input maybe in raw form we need to convert the raw unit to mV physical unit by subtracting the giving base and dividing it by gain.

Then we will use the preprocessing technique to remove the noise of the digital signal we can use bandpass filtering as the first phase of the preprocessing stage. Initially a low pass filter is used to reduce the high frequency noise from the ECG signal. The noise can be caused by EMG noise or power line interface. Then a high pass filter can be used to remove low frequency from the given signal.

3.5. Method/Algorithm/Approach/Technique

Method

- The ECG signal is acquired as an input to the system.
- Applying the preprocessing techniques on an ECG signal: first is to apply noise reduction and then splitting the data for training and testing. Splitting of data will be totally random.
- By extracting the features from an ECG signal, we will be creating a feature vector. Feature that will be highlighted are:

  - Distance of R-R interval

  - Presence of a P wave

  - Distance P-R time interval
- Artificial Neural Network will be trained on the training data. For the training of a Neural Network back propagation algorithm will be used.
- Feature vector will be provided to the Artificial Neural Network.
- Neural Network will classify the disease on the basis of the feature vector.

Algorithm

Artificial Neural Network is a Machine Learning algorithm that is inspired by a human brain neural system. Artificial neurons or nodes are interconnected with each other.

A Neural Network is divided into three parts: Input layer, hidden layers and an output layer.

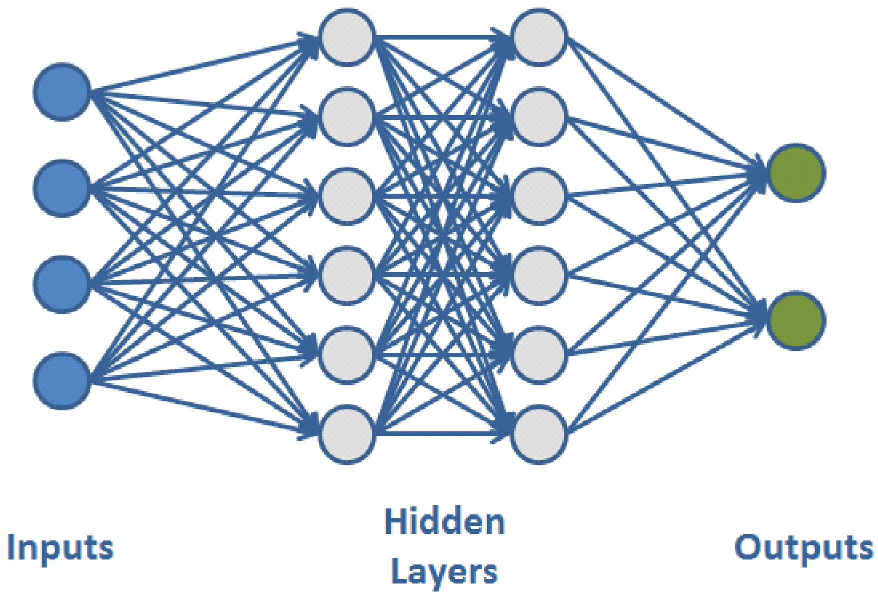

Figure 8: A Neural Network

Each connection has a specific weight which has a random value which is adjusted during the time of training.

Each node has two functions to perform:

1- To calculate the weighted sum of inputs and weights connected to it.

2- To apply an Activation Function on the output of a weighted sum. Activation function can be step, sigmoid, ReLu, Softmax etc.

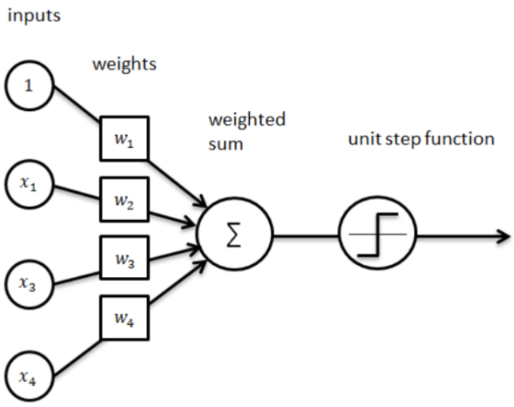

Figure 9: Single Neuron

Training of a Neural Network will undergo through Back propagation algorithm.

How does Back propagation work:

1- Calculate the error: To check the difference between actual output and the predicted output.

2- Minimum error: Using the gradient descent technique we will check whether the error is minimum or not.

3- Update the parameter: If the error is not minimum back propagate into the network, check the weights contribution in the output and update the weights accordingly.

4- Model is ready to make prediction: As if the error is minimum the model is ready to be given an input and detect a disease.

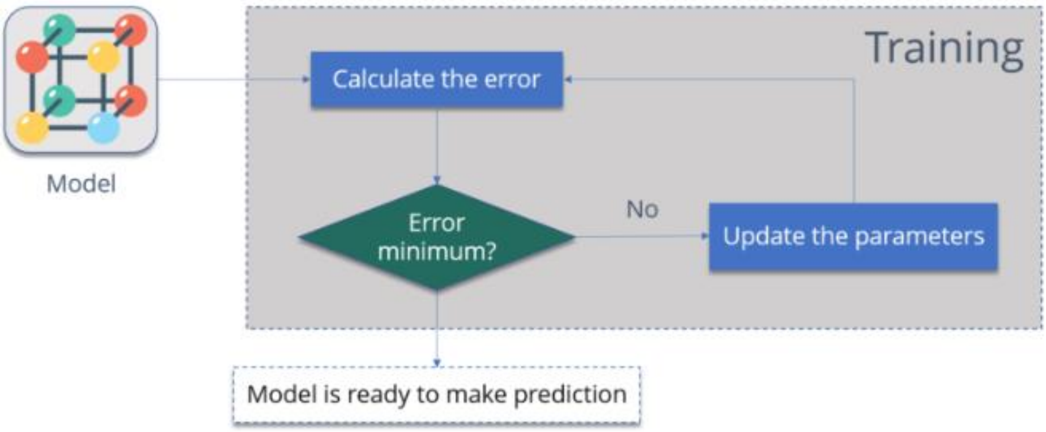

Figure 10: Cycle of Back propagation

How does a complete cycle work.

1. **Forward Propagation**: Getting the input taking a weighted sum and applying activation function on each layer and getting and output.

2. **Error Estimation**: Calculate the error and apply Gradient Descent to check if the error is minimum.

3. **Back Propagation**: If the error is not minimum then back propagate and change the weights values accordingly.

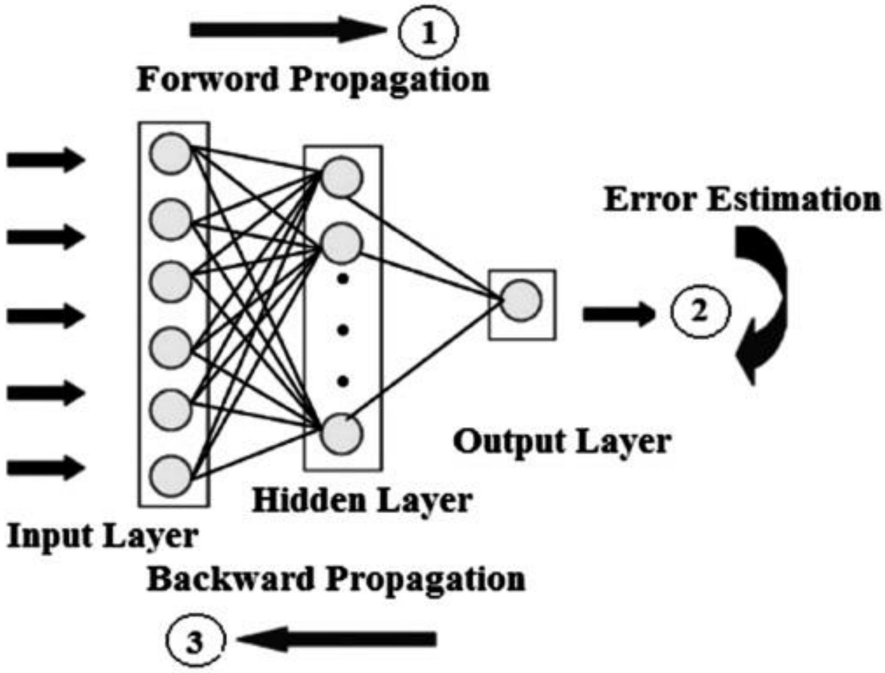

Figure 11: Back Propagation

3.6. **Pseudo Code**

- Detection of myocardial in infection using ECG.
- Fetch data from computer
- Input the file and apply preprocessing
- Apply noise reduction.
- Feature extraction.
- Evaluation of feature using ANN.
- Classification and tanning of feature vector using ANN.
- Displaying result along with disease symptoms.

3.7. **Expected output**

Since our whole system is a two steps/stages process, first being the "noise reduction" stage and the second one being the "disease detection" stage. Noise detection is very important stage because it is very difficult to extract features from raw ECG that look like below figure.

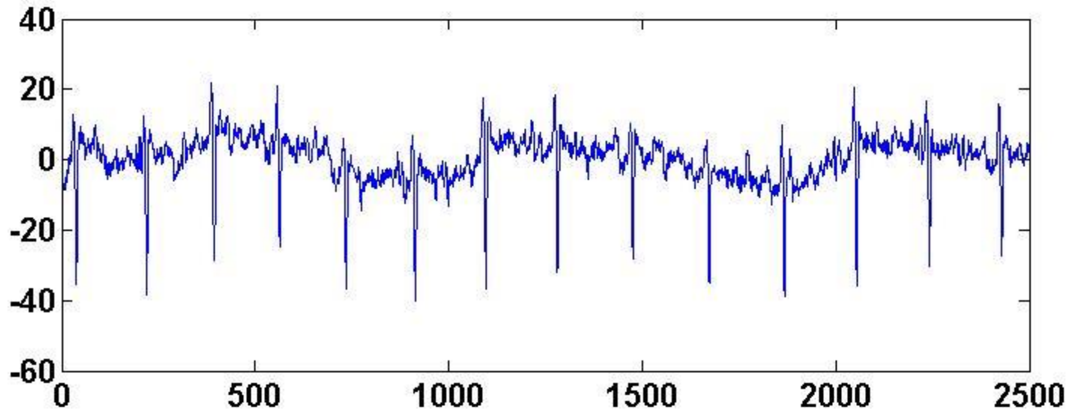

Figure 12: ECG strip

So, in first step of noise detection several irregularities from the ECG signal are removed. This is considered as first output which will act as input to ANN system which will detect two out of several myocardial infarctions.

The output will comprise of predicted disease (atrial fibrillation or ventricular fibrillation) along with symptoms of that disease or normal. In case of if no disease is found and ECG is still abnormal then system will also inform about that "specified myocardial infarction not found". It will also locate the affected area in signal using reshape method and will help doctors to effectively diagnose disease and locate effected area.

3.8. Summary

Atrial fibrillation and ventricular fibrillation are among those arrhythmias which if not detected early may lead to heart attack or stroke. Doctor while analyzing ECG manually may miss some of the information. So, our system will assist doctor in this case so that no details are missed. The system covers the following steps. Initially raw ECG signal is taken from physionet database. After this noise reduction techniques are applied. Then after this, features will be extracted which will create feature vector which will be then used for training and testing of the neural network. After the training, system will able to classify myocardial infarction as affected or not affected on the basis of training. If the myocardial infarction is found affected then it will classify it as atrial fibrillation or ventricular fibrillation, the system will recognize these heart abnormalities.

Chapter 4

4. Implementation (Block Diagram)

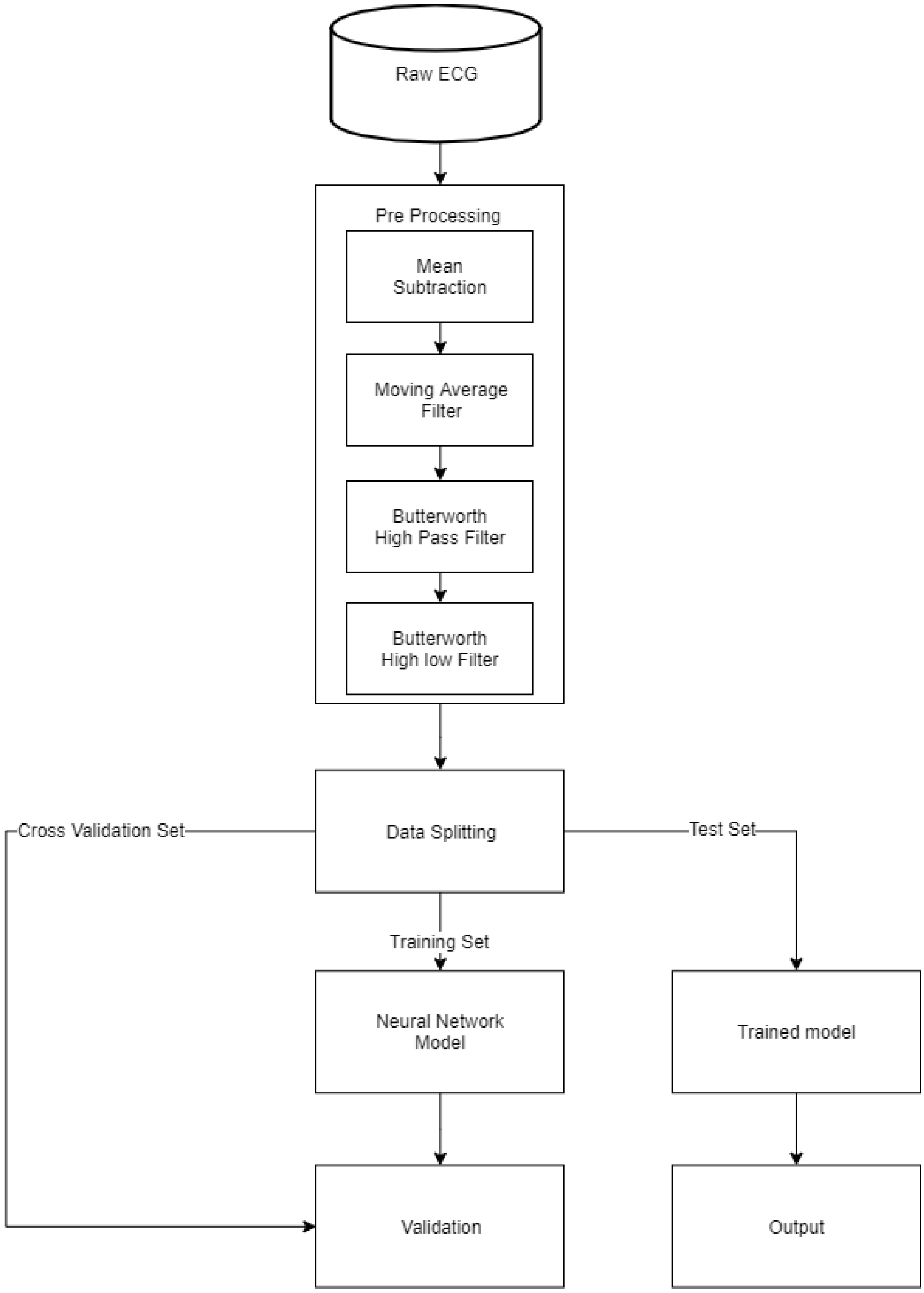

Figure 13: Block Diagram of project

4.1. Development Tool:

For training our Neural Network we choose Google's online research platform named "Google colab”. This is because in Google colab we get access to the GPU for free, compared to our systems this online platform helps us to train our model faster. We used Keras library with Tensor flow framework on the backend to build and train our model, other than that we used libraries like sci-kit learn, numpy and pandas for multiple operations in the process.

4.2. Implementation Issues:

While implementing our modules we face number of issues. These issues are stated below:

- The first issue was to select python library that will do our proposed task efficiently and accurately.
- Then the second task was to find reliable database for patients. Which we found on physio net MIT BIH database. In MIT BIH arrhythmia database, there was real-time datasets of 100 patients.
- The MIT BIH dataset was in .dat format. We extracted the useful information such as name, signal, time period, signal base, signal gain.
- Initially the signal was raw and had to be converted into mV. For this the signal has to be subtracted from base and divided by gain.
- The obtained signal was noisy hence different noise removal techniques were applied to stabilize the signal while preserving important features
- Also, there is baseline wandering in the obtained ECG Signal which may lead to wrong and inaccurate results.
- Another issue was that the ratio proportion of datasets of three categories was highly varying.
- The dataset of all categories must have equal ratio. But we don’t have issue. One way to deal with this problem was to drop the data from A.F. and V.F. down to the Normal ECG's number of tuples. But this would have cost us with overall decrease in the dataset. We tackled this issue by duplicating the Normal ECG dataset three times

4.3. **Framework Section for Development:**

4.3.1. Signal Acquisition:

We will acquire ECG signals directly from Dataset stored in our PC downloaded from MIT BIH arrhythmia database.

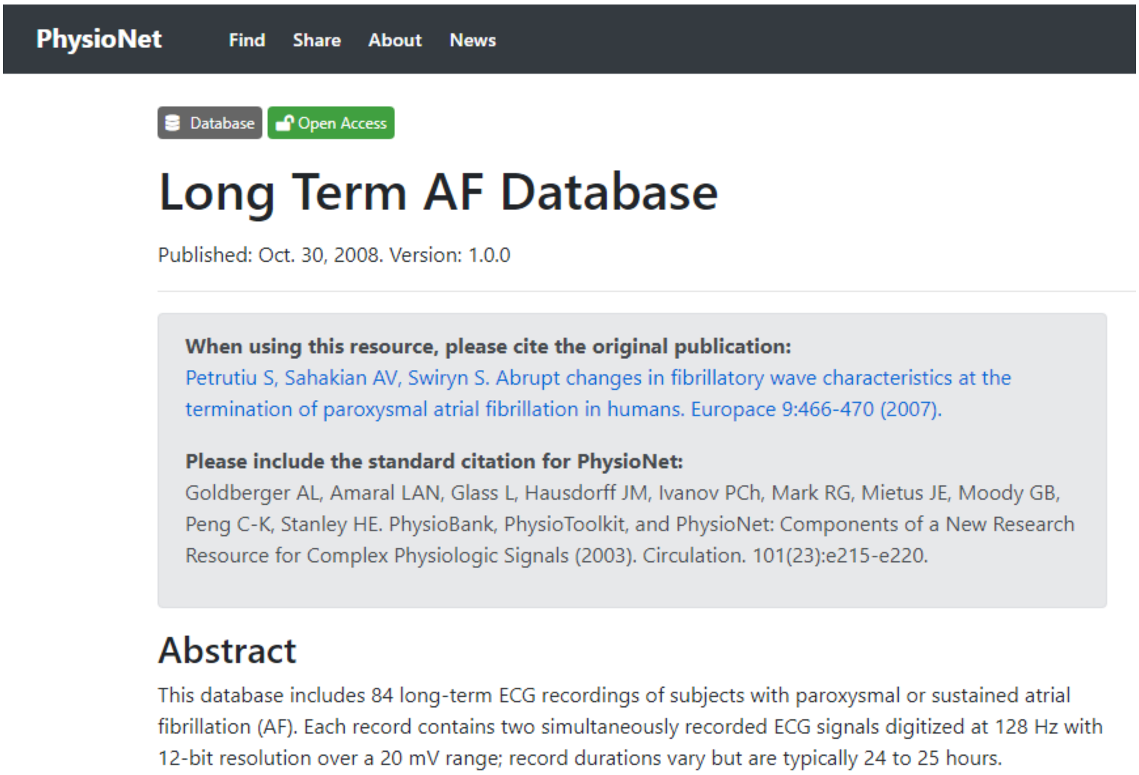

Figure 14: Benchmark Website for ECG signal extraction

We will be using ECG signals as our input Signal. This dataset has real-time ECG recordings of patients. It has separate database for both diseases (Atrial and Ventricular Fibrillation). We will use data of patients of each disease including normal.

The file obtained from the database was of .dat format and when imported to python that was found to be a dictionary structure. The following picture show the structure of the file obtained from the Physio net website

```
{'record_name': '103',
 'n_sig': 2,
 'fs': 128,
 'counter_freq': None,
 'base_counter': None,
 'sig_len': 11059200,
 'base_time': datetime.time(13, 24, 7),
 'base_date': datetime.date(2005, 5, 2),
 'comments': [],
 'sig_name': ['ECG', 'ECG'],
 'p_signal': array([[ 0.06422005,  0.14820011],
        [ 0.06422005,  0.14326011],
        [ 0.06916005,  0.14820011],
        ...,
        [-0.07410005,  0.00988001],
        [-0.05928004,  0.03458003],
        [-0.03458003, -0.00988001]]),
 'd_signal': None,
 'e_p_signal': None,
 'e_d_signal': None,
 'file_name': ['103.dat', '103.dat'],
 'fmt': ['16', '16'],
 'samps_per_frame': [1, 1],
 'skew': [None, None],
 'byte_offset': [None, None],
 'adc_gain': [202.429, 202.429],
 'baseline': [0, 0],
 'units': ['mV', 'mV'],
 'adc_res': [0, 0],
 'adc_zero': [0, 0],
 'init_value': [13, 30],
 'checksum': [-26557, -18915],
 'block_size': [0, 0],
 'base_datetime': datetime.datetime(2005, 5, 2, 13, 24, 7)}
```


Figure 15: Information obtained from signal file

All the important information related to the signal was provided in that .dat file which was used to further subtract the import information needed for our project such as the file name, signal, time, gain, baseline, signal length, sampling frequency. After loading signal, the preprocessing phase will come.

4.3.2. Preprocessing:

For the preprocessing the signal had to go through multiple noise removel methods in order to achieve a noise free ECG signal. For this purpose we applied four methods on the ECG signal. First here is the raw signal obtained form MIT BH database coverted into mV.its is just a 2 second sample of ECG for demo purpose.

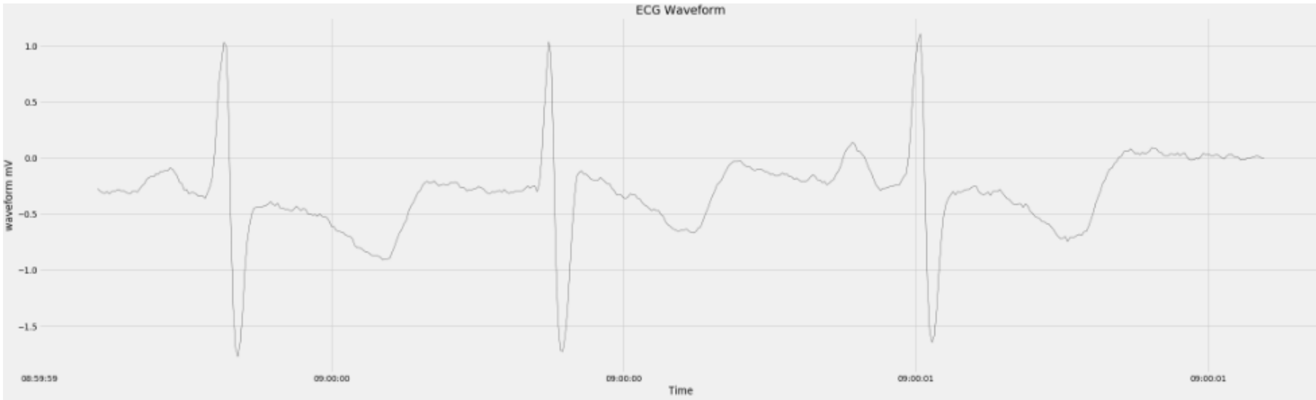

Figure 16: 3 cycles of an orignal ECG strip

The upper signal is noisy and if we look closly on the ECG strip we can observe the different noise. Mainly the noise in the ECG are caused by the

1. **Electromagnetic intervention** by power line, Electromagnetic field.
2. **The Electromyogram (EMG) noise** is generated from electrical activity of the muscle.
3. **Baseline wander** is a low-frequency noise component which is mainly due to respiration, and body movement.
4. **Motion artifacts** are transient base line changes caused by changes in the electrode-skin impedance with electrode motion.

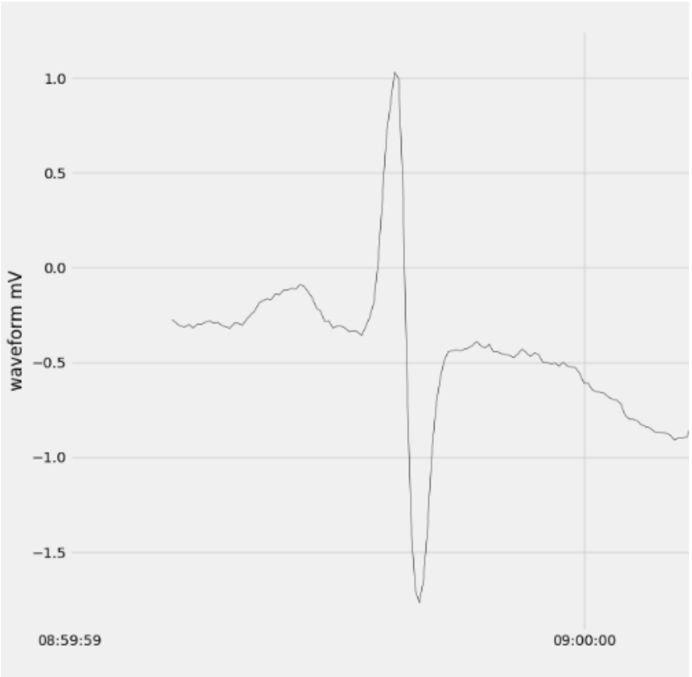

Figure 17: Noisy Signal Zoomed to one Cycle

First of all the mean value is calculated from the noisy signal and then the obtained mean of the signal is subtracted from it hence making the mean value of ECG signal to zero.

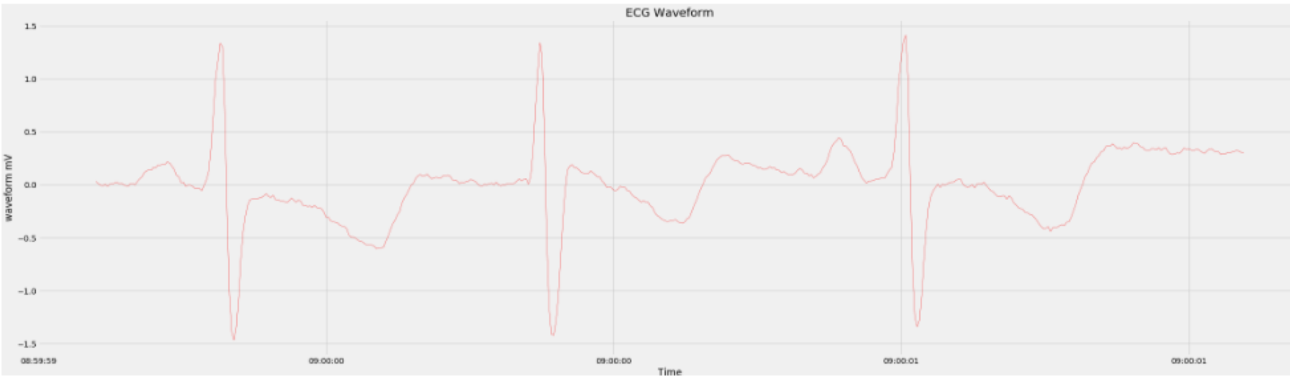

Figure 18: After subtracting mean from signal

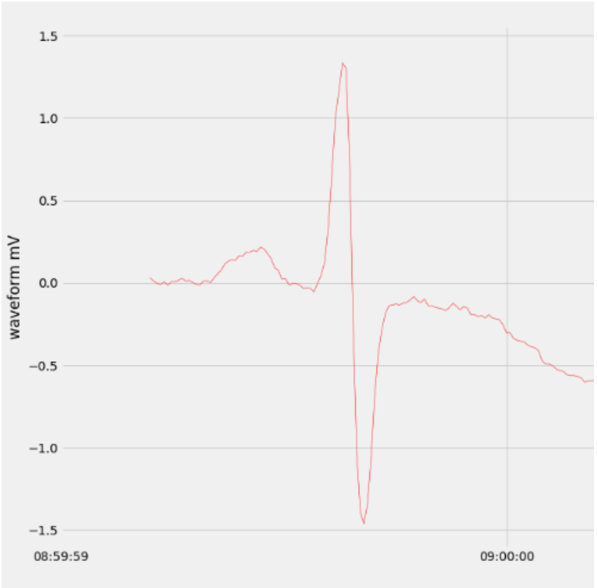

Figure 19: One cycle after mean substraction

a MV filter(moving average) is applied to the ECG signal which is a very simple low pass FIR filter which is used for smoothing the signal. As the filter size increases the output smoothness also increases.

Figure 20: ECG Strip after appling MA filter

Signal After appling Moving Average filter to remove MA filter is applied of order 5. This should remove most interception and muscles noise from the ECG signal.

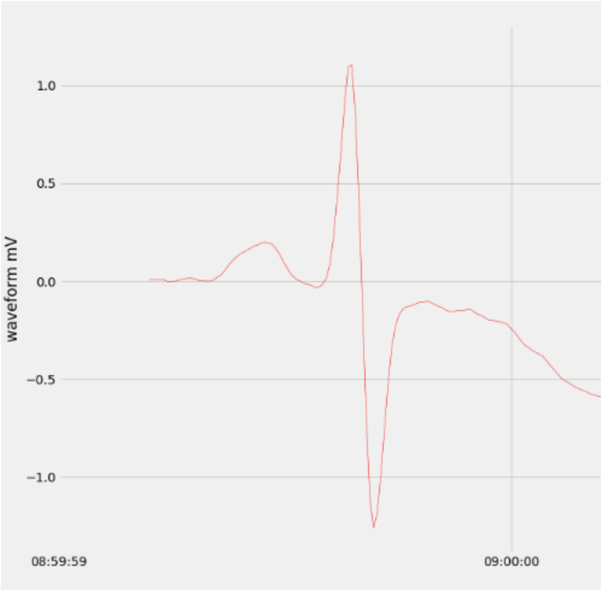

Figure 21: Moving Average Filter of One cycle

In the third step we have applied a butterworth high pass filter with the cutoff frequency of 1 Hertz the signal. This signal will impose drift suspension on the signal. a high pass filter is a circuit that diminishes all the signals below a specified cut off frequency denoted.

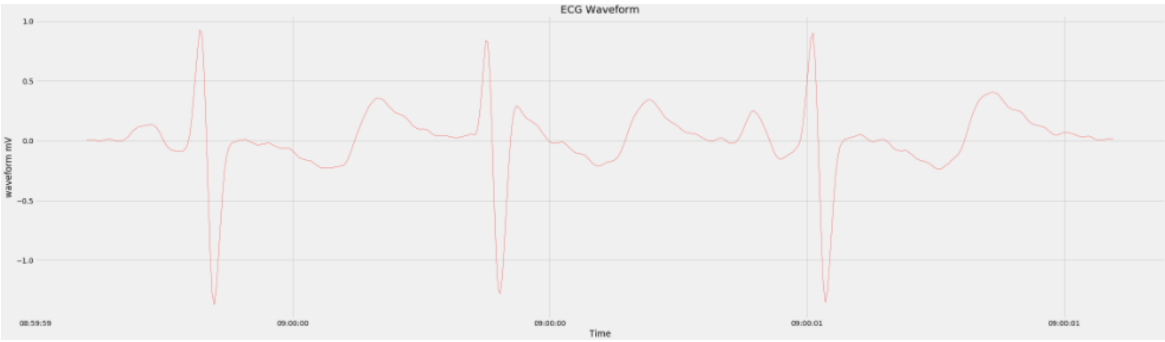

Figure 22: Strip after appling butterworth high pass filter

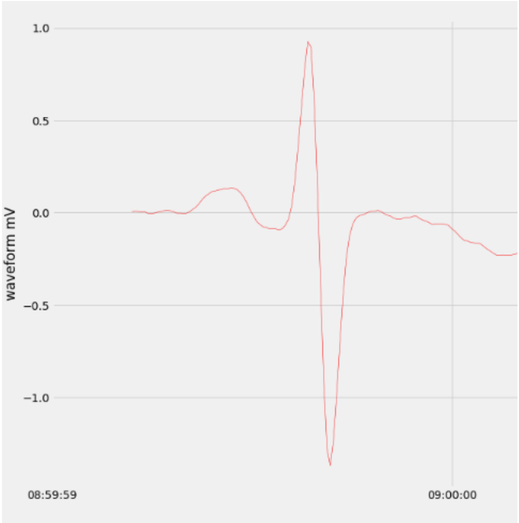

Figure 23: One cycle after High pass filter

At last we have applied a a low pass butterworth filter of order 12 with a cutoff frequency of 20 Hertz a low pass filter is applied to remove unnecessary high frequency information.

Figure 24: ECG strip after appling Low Pass Butterworth Filter

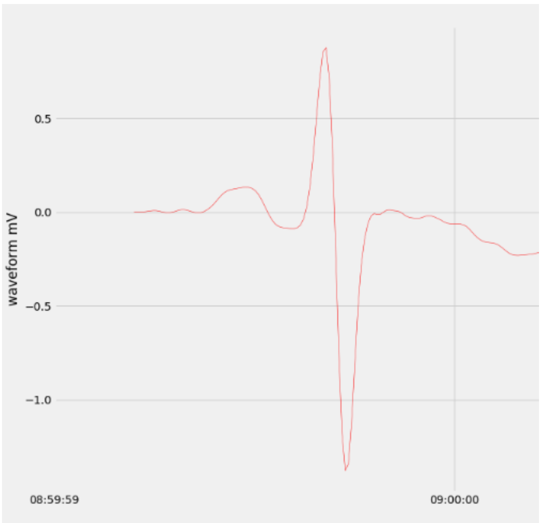

Figure 25: Efffect of Low pass filter (one cycle)

The purpose of applying the butterworth filter was not only to completely reject the unwanted frequencies but should have the uniform sensitive from the Wanted frequencies. Butterworth filters are one of the most commonly used digital filters used in motion analysis and in audio circuits.

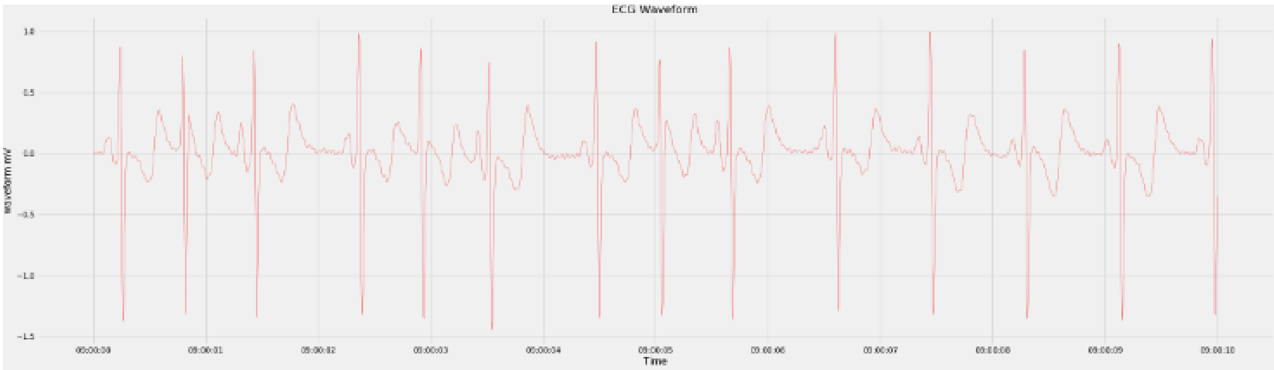

Figure 26: After applying filters

After applying the pre processing on obtained file of Normal ,atrial fibilation and venticualr fibilation the the dataset were formed in an oranzied way which is also a part of data cleaning.A matrix was formed per row 200 sample of signal were stored colum wise and in the end a lable was given to indentife each signal if it is of normal , AF or VF.the obatain matrix was.for example for atrial fibilation we obtained datasets in which 14972 dataset were extracted from signal, orgained rows wise and each row conataing 200 samples of signl and 1 colume to identifty the ECG type.

	0	1	2	3	4	5	6	7	8	9	...	191	192	193	
0	4.360680e-08	6.410188e-07	4.422439e-06	0.000019	0.000057	0.000124	0.000207	0.000269	0.000281	0.000252	...	-7.949701e-04	-0.000924	-0.000948	-
1	-9.809577e-08	-1.426375e-06	-9.707027e-06	-0.000041	-0.000119	-0.000253	-0.000398	-0.000467	-0.000398	-0.000233	...	-2.641408e-04	-0.000319	-0.000379	-
2	4.653737e-08	6.958202e-07	4.888055e-06	0.000021	0.000065	0.000149	0.000260	0.000366	0.000438	0.000479	...	-2.834607e-04	-0.000310	-0.000331	-
3	-8.002395e-08	-1.164628e-06	-7.935598e-06	-0.000033	-0.000096	-0.000196	-0.000267	-0.000168	0.000248	0.000976	...	-3.521987e-04	-0.000331	-0.000334	-
4	-8.000442e-08	-1.183878e-06	-8.228626e-06	-0.000036	-0.000107	-0.000239	-0.000406	-0.000536	-0.000564	-0.000494	...	-2.250120e-05	0.000001	0.000048	
5	-4.610755e-08	-6.973168e-07	-4.963250e-06	-0.000022	-0.000069	-0.000157	-0.000276	-0.000379	-0.000415	-0.000378	...	-5.215552e-04	-0.000701	-0.000721	-
6	-1.088216e-08	-1.521527e-07	-9.613984e-07	-0.000004	-0.000008	-0.000008	0.000011	0.000065	0.000152	0.000236	...	-1.281615e-04	-0.000067	-0.000014	
7	-7.730830e-08	-1.159874e-06	-8.173229e-06	-0.000036	-0.000110	-0.000247	-0.000423	-0.000564	-0.000603	-0.000545	...	1.960172e-03	0.000663	-0.000505	-
8	-8.512314e-08	-1.239585e-06	-8.448830e-06	-0.000036	-0.000104	-0.000223	-0.000356	-0.000434	-0.000407	-0.000314	...	2.167566e-04	0.000330	0.000594	

Figure 27: Dataframe of dataset

After that the datasets of signal were organized and stored in an ordered manner to be feed to our ANN algorithem.the files was stored as a CSV (Comma sperated vlaues) for further working.

4.3.3. Data Splitting:

After the pre-processing of the datasets next process is to split the dataset. It should be noted that the previously acquired dataset was categorized into its specific class in a separate CSV files (i.e. Signals of Arterial Fibrillation, Ventricular Fibrillation and Normal were in separate files). It was mandatory to integrate the whole dataset into a single CSV file in order to perform data splitting.

Another issue was that the ratio proportion of datasets of three categories was highly varying. Ventricular Fibrillation dataset had 57750 tuples, Arterial Fibrillation dataset contained 14972 tuples whereas Normal ECG dataset had only 3649 tuples. The dataset of all categories must have equal ratio. One way to deal with this problem was to drop the data from A.F. and V.F. down to the Normal ECG's number of tuples. But this would have cost us with overall decrease in the dataset.

We tackled this issue by duplicating the Normal ECG dataset three times, which resulted to Normal ECG dataset up to 10947.

```
[ ] normal_extd=pd.concat([normal]*3, ignore_index=True)

[ ] normal_extd.shape

(10947, 202)
```

Figure 28: Normal ECG dataset up to 10947

We fix a round figure for all three classes i.e. 10000 tuples for all three classes. Next thing was to append all three datasets which resulted into 30000 tuples dataset of all three classes.

```
[ ] dataset=normal_final
    dataset=dataset.append(AF_final, sort=False)
    dataset=dataset.append(VF_final, sort=False)

[ ] dataset.shape

(30000, 203)
```

Figure 29: 30000 tuples dataset of all three classes

Next thing was to shuffle the dataset. We shuffled the dataset three times, just to make sure that all three classes that we appended previously should be placed completely random.

Now our dataset is ready for splitting. We will be splitting our dataset into three parts of 70-15-15. 70% for training, 15% for cross validation and 15 % for test. Why do we need to divide our dataset into three parts? Training set is for training our model. Cross-Validation set or Validation set is to monitor the performance of our model. Note: The ideal condition for validation of our model is when there is no bias and variance in our model.

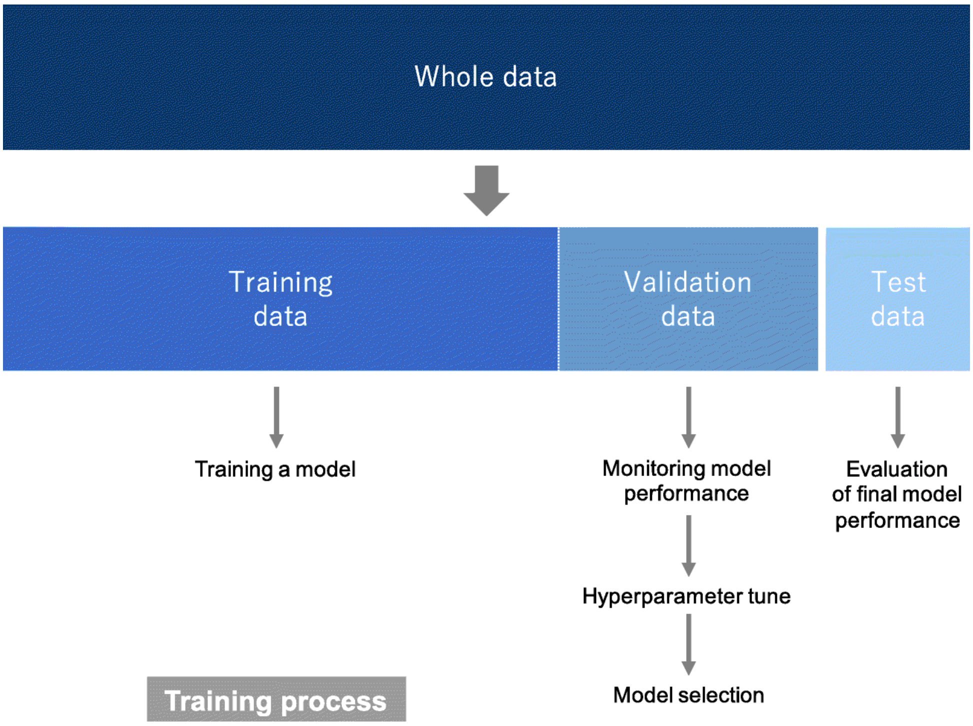

Figure 30: difference between train, validation and test set

4.3.4. Feature Engineering:

Feature Engineering or Features Computation is supposed to be the major part of every Machine Learning project. But in case of Neural Network we can rely on our model to handle it for us.

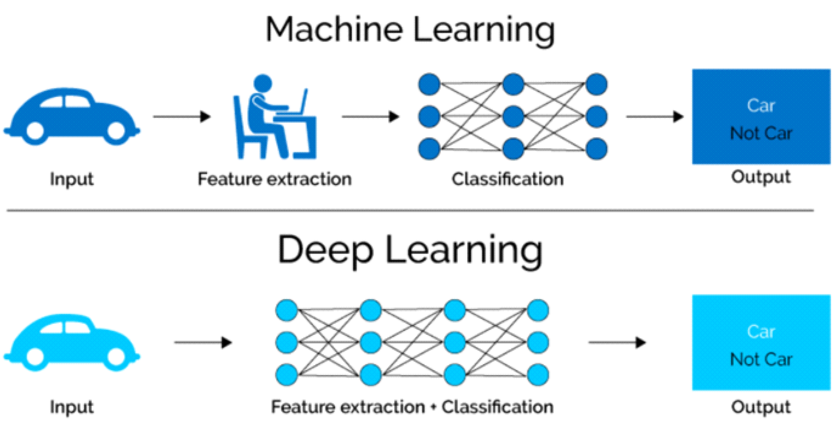

Figure 31: difference between Machine Learning and Deep Learning Feature Extraction

In case of NN, the no of input neurons defines the number of features. In our case of ECG signal, defining features manually will not be an efficient way because each part of ECG signal contains some information which might be an important feature.

4.3.5. Model:

The model we trained is an Artificial Neural Network using Back Propagation technique. As we have already split our dataset into train, validation and test. We first imported train and validation dataset and after applying several operations we get a suitable dimension that could be fed into our Neural Network for training. Our input layer consists 200 Neurons and output layer contains 3 neurons defining our three classes. The output layer gives us the probability of each of the class, greatest probability is assigned as 1 and other two are classified as 0.

4.4. Configuration Management

We have used the external physio net MIT BIH Dataset for input to the system.

4.5. Deployment factor:

There are several factors that can affect our system when deployed. It is not able to detect myocardial infarctions other than Atrial and Ventricular Fibrillation. This system may give inaccurate results if there is systematic or human error while taking ECG. This system is still dependent on Doctor.

4.6. Summary

In this phase, implementation of our project is under process. We are using Python 3.7. There are three phases of our project one is preprocessing, other is training and the third is testing. During the implementation phase, we are encountering issues like finding reliable database, signal was raw which has to be converted into mV and signal was also not very smooth. So, we will be using moving average filter to the signal which is a low pass FIR filter which is used for the smoothing the signal. Another issue was that the ratio proportion of datasets of three categories was highly varying. We tackled this issue by duplicating the Normal ECG dataset three times. In our case of ECG signal, defining features manually will not be an efficient way because each part of ECG signal contains some information which might be an important feature.

Chapter 5

5. Results and Discussion

5.1. Introduction:

We started with 43,500 total number of examples which was divided into three portions i.e. training set (70%), validation set (15%) and test set (15%). In the beginning, ANN model that is classifying between three different classes i.e. Atrial Fibrillation, Ventricular Fibrillation and Normal is fed with the 30,450 training examples and validated across 6,525 examples. In order to achieve higher accuracy multiple hyper-parameters were tuned during multiple iterations of training and validation phase. Finally, after multiple iterations Neural Network started to generalize the complex patterns of the ECG signals and was generating following accuracy on training and testing set.

5.2. Output:

To completely analyze the finding of the respected project, it is necessary to analyze on modular basis.

```
Accuracy on training set is 0.9999343185550083

Accuracy on validation set is 0.9975478927203065
```

After acquiring highest possible outcome from the respected model, it was necessary to test the model with the test set that is of 6525 training examples. The accuracy score for the test set was:

```
Accuracy Score:  0.9969348659003832
```

5.3. **Model Analysis:**

As it was already defined that a Deep Neural Network is used in this project. The Neural Network used contained four layers including the input layer which consists of 200 input neurons. The first hidden layer consists of 400 neurons and in total 80,400 trainable parameters, Second layer consist of 700 neurons and in total 280,700 trainable parameters. Third or the output layer consists of 3 neuron and in total 2,103 trainable parameters. Total number of trainable parameters in our Neural Network are 363,203.

```
Model: "sequential_5"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_31 (Dense)             (None, 400)               80400
_________________________________________________________________
dense_32 (Dense)             (None, 700)               280700
_________________________________________________________________
dense_33 (Dense)             (None, 3)                 2103
=================================================================
Total params: 363,203
Trainable params: 363,203
Non-trainable params: 0
```

Figure 32: Model Details

"ReLU" activation function was used for the intermediate layers and "Softmax" activation was used for the output layer as we are doing multi-class classification. In order to converge the model at the optimal point we used "Adam" optimizer, Learning Rate was "0.0001".

Table 5: Complete ANN Parameters

Parameters	
Trainable parameters	363,203
Hyper-parameters	
No. of Layers	3
No. of Neurons	400, 700, 3
Activation Function (intermediate layers)	ReLU
Epochs	More than 2000
Activation Function (output layer)	Softmax
Learning Rate	0.0001
Optimizer	Adam
Loss calculation	Categorical Cross-entropy

5.1. Statistical and Graphical Analysis:

After achieving the accuracy of 99% on our training and validation set, it is clear that there is no under-fitting or overfitting. Following figure shows the drop in our loss during the time of training.

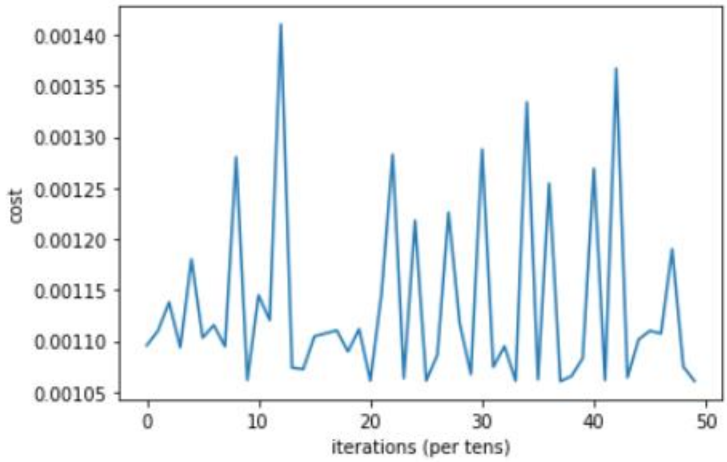

Figure 33: Drop in loss

To analyze or verify the accuracy of our model we performed few evaluations using our Actual Output and Predicted Output. For the similar purpose two of the functions of library "ScikitLearn" were used that were "accuracy_score()" and "classification_report()". Accuracy_score calculated the accuracy of our test data, comparing our Actual output and Predicted output which resulted into following outcomes.

```
Accuracy Score:  0.9969348659003832
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      2167
           1       1.00      1.00      1.00      2142
           2       1.00      1.00      1.00      2216

    accuracy                           1.00      6525
   macro avg       1.00      1.00      1.00      6525
weighted avg       1.00      1.00      1.00      6525
```

Figure 34: Classification Report

We can see that the precision of our model is 100%.

Another evaluation technique that we performed was "Confusion Matrix". Confusion Matrix is one of the best techniques which can be used to check the precision if the model. This technique creates a matrix of N x N dimensions, where the number N is the number of classlabels. The aim of this matrix is to reach to identity matrix. Following figure represents the confusion matrix of our test set.

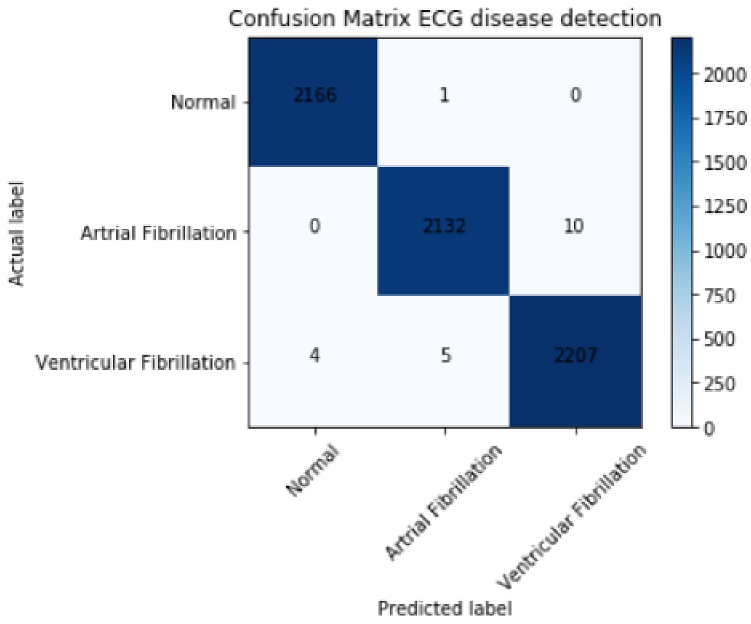

Figure 35: Confusion Matrix ECG Disease Detection

We can see that out of 6,525 examples, 6,505 were predicted accurately and 20 examples were predicted wrong which defines that there is only 0.3% chance of wrong prediction.

5.2. **Summary:**

In this chapter we built and evaluated our network. We developed a Neural Network which consists of four layers, 200 inputs, 400 hidden layer one, 700 hidden layer two and 3 output layer neurons. We used ReLU activation function and Adam optimizer. We were able to acquire the accuracy of 99% through training and validating our ANN model. For testing we used functions accuracy score, classification report and technique confusion matrix to evaluate our model through which acquired 100% precision and out of 6,525 examples 6,505 were predicted accurately which means only 0.3% of error.

Chapter 6

6. Conclusion:

We presented a model that will detect myocardial infarctions through an ECG. Compared to previous work our model works in two steps: pre-processing and classification. Our approach begins by preprocessing which include removing of noise from the given ECG digital signal and make it less noisy so next step become easy. Atrial fibrillation and ventricular fibrillation are among those arrhythmias which if not detected early may lead to heart attack or stroke. For signal acquisition we use the data set from MIT BIH arrhythmia database. The database includes long term ECG recording of subjects with paroxysmal or sustained atrial fibrillation. Recorded ECG signals are digitized at 120 Hertz. We gather around 26 datasets of patients having atrial and ventricular fibrillation. For the preprocessing the signal had to go through multiple noise removal methods. there is various noise in the ECG including Electromagnetic intervention, EMG, Baseline wander and motion artifacts.

First of all, the mean value is calculated from the noisy signal and then the obtained mean of the signal is subtracted from it thus making the mean value zero. Next a moving average filter is applied to the signal which is a very simple low pass FIR filter which is used for smoothing the signal. As the filter length increases the smoothness of the output increases, whereas the sharp transitions in the data are made increasingly blunt. In the third step we have applied a Butterworth high pass filter with the cutoff frequency of 1 Hertz the signal. This signal will impose drift suspension on the signal. A high pass filter is a circuit that diminishes all the signals below a specified cut off frequency denoted. At last we have applied a low pass Butterworth filter to the ECG signal of order 12 with a cutoff frequency of 20 Hertz a low pass filter is applied to remove pointless high frequency material from signal. The main purpose of applying the Butterworth filter was not only to completely reject the unwanted frequencies but should have the uniform sensitive which is little bit of the noise from the Wanted frequencies. Butterworth filters are very famously used digital filters that are massively used in analysis of motion and in audio chip board.

We developed a Neural Network which consists of four layers, 200 inputs, 400 hidden layer one, 700 hidden layer two and 3 output layer neurons. We used ReLU activation function and Adam optimizer. We were able to acquire the accuracy of 99% through training and validating our ANN model. For testing we used functions accuracy score, classification report and technique confusion matrix to evaluate our model through which acquired 100% precision and out of 6,525 examples 6,505 were predicted accurately which means only 0.3% of error.

7. REFERENCES

[1]Ann Brian, S. N. (2017). ECG Based Algorithm For Detecting Ventricular. *International Conference on Intelligent Computing and Control Systems IEEE*, 6.

[2]Arturas Serackis, V. A. (2015). Identification of ECG Signal Pattern Changes. *ISSN 2325-8861*, 4.

[3]Kusum Tara, A. K. (2017). Detection of Cardiac Disorder using MATLAB based. *IEEE Region 10 Humanitarian Technology Conference*, 4.

[4]M. Botsivaly, C. K. (July 23-28,2000). Evaluation of a new technique for the Detection. *EMBS International Conference IEEE*, 4.

[5]Qiao Li, C. R. (JUNE 2014). Ventricular Fibrillation and Tachycardia. *0018-9294 © 2013 IEEE.*, 7.

[6] C. VENKATESAN, P. K. (2018). ECG Signal Preprocessing and SVM Classifier-Based Abnormality Detection in Remote Healthcare Applications. IEEE, 9767-9773.

[7] Kankava, H. A.-Z. (2015). Heart Rate Monitoring and PQRST Detection Based on Graphical User Interface with MATLAB. International Journal of Information and Electronics Engineering, Vol. 5, No. 4, July 2015, 311-316.

[8] Kartika Resaid, A. D. (2018). Detection of Atrial Fibrillation Disease Based on Electrocardiogram Signal Classification Using RR Interval and K-Nearest Neighbor. 6th International Conference on Information and Communication Technology (Ecocity) (pp. 501-506). Bandung: International Conference on Information and Communication Technology (Ecocity.

[9] Kusum Tara, A. K. (2017). Detection of Cardiac Disorder using MATLAB based Graphical User Interface (GUI). 7 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 440-443). Dhaka: IEEE.

[10]Andrew Y. Ng, P. R. (2017). Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. arxiv.org, 9.

[11]Hela Lassoued, R. k. (2018). ECG Multi-Class Classification using Neural Network as Machine Learning Model. 2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET) (p. 6). Hammamet, Tunisia: IEEE.

[12]Milad Salem, S. T.-S. (2018). ECG Arrhythmia Classification Using Transfer Learning from 2-Dimensional Deep CNN Features. arxiv.org, 4.

[13]Nabil Ibtehaz, M. S. (2018). VFPred: A Fusion of Signal Processing and Machine Learning techniques in Detecting Ventricular Fibrillation from ECG Signals. arXiv.org, 24.

[13] Wall Street Journal, "For Men at 40, Risk of Cardiac Death 1 in 8," http://online.wsj.com/article/SB125833149978449651.html

[14] M. Gertsch, The ECG Manual. London: Springer-Verlag, 2009.

[15] M. Khan, Rapid ECG Interpretation, Totowa, NJ: Humana Press Inc., 2008.

[16] Oresko, Joseph John. Portable heart attack warning system by monitoring the ST segment via smartphone electrocardiogram processing. Diss. University of Pittsburgh, 2010

[17] Kora, P. & Kalva, S.R. SpringerPlus (2015) 4: 666. https://doi.org/10.1186/s40064-015-1379-7