

ARRHYTHMIA ANALYSIS AND CLASSIFICATION USING LONG SHORT TERM MEMORY NETWORK

Submitted in partial fulfilment of the requirements for the award of Bachelor of
Engineering degree in Electronics and Communication Engineering

By

V.HEMALATHA (37130160)

S. KANTHARI (37130184)



**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
SCHOOL OF ELECTRICAL AND ELECTRONICS**

**SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade “A” by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119**

APRIL – 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **V. HEMALATHA (37130160)** and **S.KANTHARI(37130184)** who have done the Project work as a team who carried out the project entitled "**ARRHYTHMIA ANALYSIS AND CLASSIFICATION USING LONG SHORT-TERM MEMORY NETWORK**" under my supervision from December 2020 to April 2021.

Internal Guide

Dr. M. SUBRAMONIAM, M.E., Ph.D.

Head of the Department

Dr. T. RAVI M.E, Ph.D.

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

We **V. HEMALATHA (37130160)**, **S. KANTHARI (37130184)** hereby declare that the Project Report entitled **“ARRHYTHMIA ANALYSIS AND CLASSIFICATION USING LONG SHORT-TERM MEMORY NETWORK”** done by us under the guidance **Dr. M. SUBRAMONIAM** at Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Electronics and Communication Engineering.

DATE:

PLACE:

1.

V. Hemalatha

2.

S. Kanthari

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

We are pleased to acknowledge our sincere thanks to Board of Management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. We are grateful to them.

We convey our thanks to **Dr. N.M. Nandhitha,M.E,Ph.D**, Dean, School of Electrical and Electronics and **Dr. T Ravi,M.E.Ph.D**, Head of the Department, Dept. of Electronics and Communication Engineering for providing us necessary support and details at the right time during the progressive reviews.

We would like to express our sincere and deep sense of gratitude to our Project Guide**Dr.M. SUBRAMONIAM, M.E, Ph.D.**,for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of our project work.

We wish to express our thanks to all Teaching and Non-teaching staff members of the Department of **Electronics and Communication Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

An arrhythmia is a problem with the rate or rhythm of the heartbeat. The heart can beat too fast, too slowly, or with an irregular rhythm. Arrhythmia is one of the types of cardiovascular disease which is detected by the Electrocardiogram. Cardiovascular disease is the number one deadly disease in the world. The detection will reduce the number of dead happening due to these cardiovascular diseases. If arrhythmia is left undiagnosed, the heart will fail to pump enough blood to other organs. This lead to fail other organs of our body like brain and kidney. We proposed the method to classify the Electrocardiogram signals to normal and arrhythmia signals by using Long-short term memory network of five layers which is the part of deep networks.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	(v)
	LIST OF FIGURES	(viii)
	LIST OF ABBREVIATION	(ix)
1	INTRODUCTION	1
2	LITERATURE REVIEW	4
3	AIM OF THE PROJECT	10
4	MATERIALS AND METHODOLOGY	11
4.1	INTRODUCTION	11
4.2	HARDWARE REQUIREMENTS	11
4.3	SOFTWARE REQUIREMENTS	11
4.4	TECHNOLOGY USED	11
4.4.1	MATLAB programming	11
4.4.2	Features of MATLAB	12
4.4.3	Characteristics of MATLAB	13
4.4.4	Uses of MATLAB	13
4.5	APPLICATIONS OF MATLAB	14

CHAPTER NO.	TITLE	PAGE NO.
	4.5.1 <i>Signal Processing</i>	14
	4.5.2 <i>Image processing</i>	14
	4.5.3 Mapping	15
	4.5.4 Control system	15
4.6	DEEP LEARNING IN MATLAB	15
4.7	LSTM ARCHITECTUTRE	16
	4.7.1 Block diagram	17
4.8	IMPLEMENTATION	18
	4.8.1 Module Description	18
	4.8.2 Collecting Dataset	18
	4.8.3 Loading and Examine the data	18
	4.8.4 Prepare the data for training	19
	4.8.5 LSTM Network Architecture	22
	4.8.6 Train and Test the LSTM Network	23
	4.8.7 Feature Extraction	25
5	RESULT AND DISCUSSION	27
6	CONCLUSION AND FUTURE SCOPE	29
	REFERENCES	31

LIST OF FIGURES

FIGURE NO.	FIGURE TITLE	PAGE NO.
1.1	ANALYSIS OF ECG SIGNAL	2
4.1	MATLAB INTRODUCTION	12
4.2	BLOCK DIAGRAM	17
4.3	LOADING DATA	18
4.4	VISUALIZATION OF THE SAMPLE	19
4.5	SPECTRAL ENTROPY OF THE SAMPLE	20
4.6	INITIAL STAGE OF TRAINING	21
4.7	LSTM ARCHITECTURE	22
4.8	TRAINING PROGRESS	23
4.9	CONFUSION MATRIX OF TRAINING SAMPLE	24
4.10	FEATURE EXTRACTION	26
5.1	CONFUSION MATRIX OF TESTING SAMPLE	29

LIST OF ABBREVIATIONS

S.NO	ABBREVIATION	EXPANDED FORM
1	CVD	Cardio Vascular Diseases
2	BPM	Beats Per Minute
3	AF	Atrial Fibrillation
4	ECG	ElectroCardioGram
5	MATLAB	Matrix Laboratory
6	LSTM	Long Short Term Memory

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Cardiovascular diseases (CVDs) are the leading cause of human death, and seventeen million people over the world lose their lives annually due to CVDs. The common other symptoms of cardiovascular diseases are weakness, dizziness, fainting, and severe pain in the chest and left part of the shoulder. These heart diseases are happening due to the imbalance of electrolytes in the blood which is passing in through the heart to whole body. An arrhythmia is the heart disease which is the condition of irregular heartbeat. These irregularities can be caused by genetics or by injury. An arrhythmia is not only mean that the heart is beating too fast or slow, Arrhythmia is known as the heartbeat is following irregular heartbeat or heart rhythm. If that the heart is beating too fast, it is known as tachycardia. The heart beat rate of the tachycardia will be more than 100 beats per minute (bpm). And if heart is beating too slow, then it is known as bradycardia. In bradycardia the heart beat is less than 60 bpm. Some more effects of arrhythmia are skipping a beat, or in extreme cases, cardiac arrest. Atrial fibrillation, atrial flutter, and ventricular fibrillation are some other types of irregular heart beat or arrhythmia. Atrial fibrillation is the most common type of arrhythmia which affects the humans frequently (AF), Atrial fibrillation causes an irregular and fast heartbeat. Atrial fibrillation is an irregular and often rapid heart rate that will leads to heart stroke and severe heart failure. This may lead to death. The result is a fast and irregular heart rhythm.

The heart rate in atrial fibrillation ranges from 100 - 175 beats per minute where the normal heart rate is 60 - 100 per minute. Human Heart is made up of four chambers-two upper chambers and two lower chambers. The upper chambers is called as atrial and the two lower chambers are called as ventricles. In the normal cases the signal travel through the two upper heart chamber and through a connecting path between the upper and lower chambers. This connection node is called as atrioventricular node. But in case of Atrial fibrillation the signals in upper chambers are chaotic. The Electrocardiogram is the common method of diagnose to detect the arrhythmia. Electrocardiogram which is known as ECG is used to record the rhythm of heart beat and heart rate. This electrocardiogram will record the electrical signal from the heart over the period of the time. This signal will be captured in the form of sinusoidal waveform. This waves also called as 'PQRST' wave.

The main function of the heart is to collect the impure blood from other parts to the body and purify the blood and distribute among the other organs of the human body like brain, kidney, lungs etc., The electrical impulse will be generated due to this purification process. When this process is affected, then

the rhythm of heart is affected. The electrodes which are presented in the ECG playing the major role to detect the cardiovascular diseases.

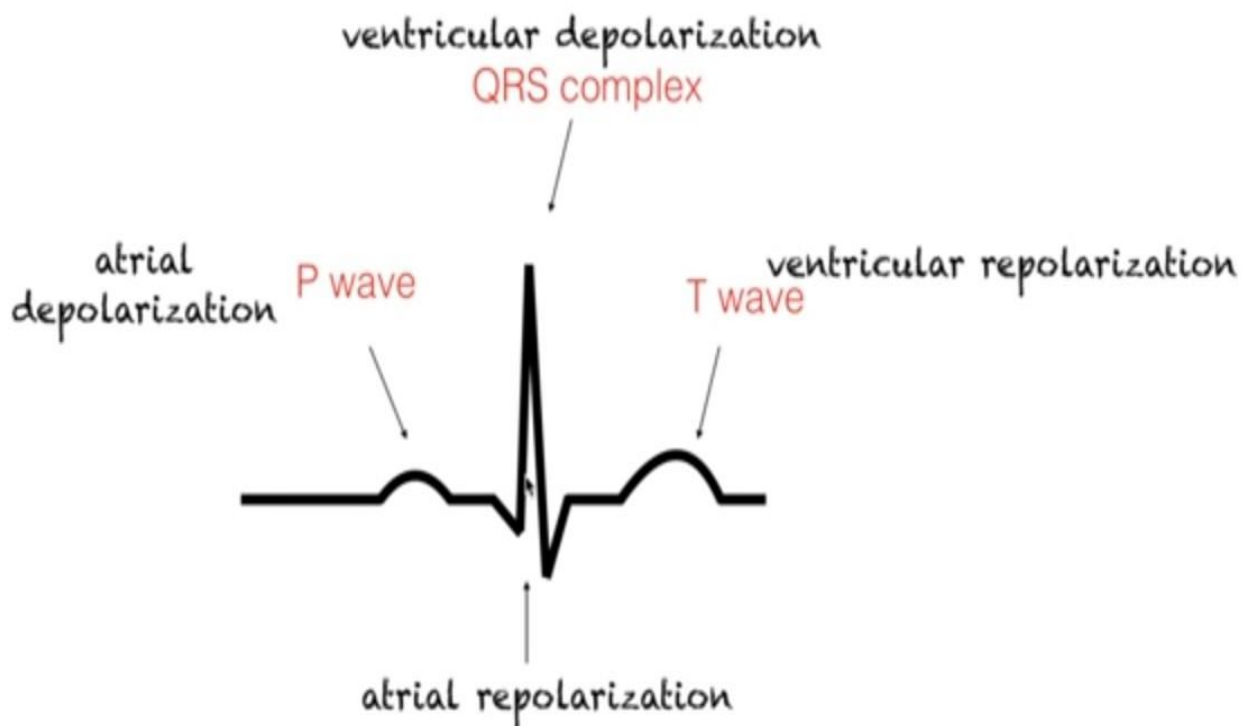


Fig 1.1 ANALYSIS OF ECG SIGNAL

P wave is the beginning of the ECG signal and it is used to record the electrical activity of the upper heart chambers. This is also known as atrial depolarization. For normal P waves always be in the upright direction. QRS complex consists of Q wave, R wave and S wave. This QRS wave is the largest one and this complex wave is used to record the electrical activity of the ventricles of the heart. The Q wave lies on the negative and that the R wave is the first positive and only peak of the QRS complex which lies on positive axis of the graph and again S wave lies on negative axis.

Followed by the QRS complex of ECG signal T waves are used to indicate ventricular repolarization. T waves are always positive that is the wave will be upright always in the normal heart beat. And the T wave is created when during the ventricular repolarization of heart.

A new neural network architecture was proposed for arrhythmia detection in ECG time signals. This deep neural network will work with high percentage of accuracy. In this report we proposed the model to detect and classify the Arrhythmia using Long-short term memory network.

CHAPTER 2

LITERATURE SURVEY

Real time machine learning for early detection of heart diseases using big data approach:

Authors: Abderrahmane ED-Dauody: Khali MAalmi(2019), Embedded and intelligent systems over that las few decades.

Coronary illness is one ordinary reason for the worldwide death.so the location of hear infection ahead of time will diminish the demise rate. The development of innovation improves the personal satisfaction, and furthermore innovation has assumed a critical part in clinical area. The association of AI and large information examination built up the medical care field to the following level. This innovation can be more powerful and terrible. To beat this savage coronary illness this paper has proposed a constant coronary illness forecast framework dependent on apache Sparkle which remain as a solid enormous scope conveyed processing stage that can be 14 utilized effectively for streaming information occasion against AI through in-memory calculations. This framework comprises of three principle areas like

information stockpiling, streaming interaction and representation. The principal utilizes Sparkle MLlib with Flash streaming and applies arrangement model on information occasions to visualize coronary illness. For the optional part they have used Aapache Cassandra to store the gigantic volume of datasets.

Authors: H. Mai ; T. T. Pham ; D. N. Nguyen ; E. Dutkiewicz, (2018).

To develop a heart recognition model assumes a crucial part in forestalling life of a few people, particularly for the individual , who are experiencing hurl coronary illness .This technique permits the clinical expert to perceive the heart

infection in a superior manner. In this task we primarily engaged to build up a non-hazard strategy for the robotized coronary illness location that includes age, sex, rest pulse, greatest pulse, and rest electrocardiography. We review four public datasets from 1071 people who were expressed for an exceptional X-beam of the heart's courses.This work has gone through actual assessment and furthermore includes three obtrusive states. To identify the coronary illness status, we utilized an instigated straight model with regularization ways by means of facilitate drop. we saw that the forecast precision as high as 72%, contrasted and 76% of other exhaustive models. This examination set forward that couple of non-obtrusive variables utilize new advances in information investigation can return the momentum practices of coronary illness hazard appraisal.

Authors: K. M. Zubair Hasan ; Shourob Datta ; Md Zahid Hasan ; Nusrat Zahan,2019 .

Linear Discriminant Analysis (LDA) is a uncomplicated and well organised method for pattern classification, while it is also widely make use for the initial discovery of diseases using Electronic Health Records (EHR) data. However, the performance of LDA for EHR data classification is recurrently impacted by two major factors like poor evaluation of LDA parameters (e.g., covariance matrix), and “linear inseparability” of the EHR data for classification. In this work,we suggest a novel classifier SDA -Sparse Discriminant Analysis method for heart disease detection. The time complexity will be decreased in this algorithm creates an ideal in scoring analysis of LDA and will be comprehensive to accomplish sparse discrimination through the combination of Gaussians if limits between classes are nonlinear or if subgroups are accessible inside every class. In most cases, compared to preceding techniques, our proposed technique is more applicable for the diagnosis of heart disease patients with higher rate.

Authors: Aditi Gavhane ; GouthamiKokkula ; Isha Pandya ; Prof. Kailas Devadkar,(2018)

Due to the uncontrolled expansion in the heart stroke rates at adolescent ages, we need to set up a framework to have the option to identify the indications of a heart stroke at a development stage and hence forestall it. It is improper for an everyday person to routinely go through expensive tests like the ECG and along these lines there should be a framework set up which is convenient and at a similar immovable, in anticipating the odds of a coronary illness. Subsequently we acquainted a technique with build up an application which can foresee the interruption of a coronary illness given essential manifestations like age, sex, beat rate and so forth The AI calculation neural organizations has developed to be the most proper and all aroundgrounded calculation.

Authors; G. Shanmugasundaram; V. Malar Selvam; R. Saravanan; S. Balaji, (2018)

Coronary illness addresses the sort of condition which has been the primary driver to heart breakdown. Not all patients with coronary conduit illness have chest torment as the sign. There are some different reasons, for example, R-Pulse, S-Cholesterol, F-Glucose, R-ECG, Ex-Ang, Number of significant Vessels impeded, Thallium Sweep which additionally bring off on coronary illness. At the beginning stage the location of coronary illness can get human lives. Information mining procedures, for example, Innocent Bayes, Choice Tree, K-closest neighbor, and so on, are utilized in the Coronary illness expectation relies upon the few boundaries. The intention of this exploration is to investigate about the few elements and its significance in perceiving the coronary illness. Moreover it likewise investigates about the strategies and models worn for identification. This work close with the key difficulties which are not yet investigated.

Authors: Radhanath Patra; BonomaliKhuntia, (2019)

Coronary illness is the significant justification expanding passing rate in both genders in around the world. Coronary illness is expanding radically at a fast rate in grown-up and youngsters, individuals of the two guys and females now days. Consequently this leads for the quick therapy with right methodologies and furthermore to expand advantages of the clinical arrangements to defeat the pandemic of the cardiovascular sickness.

Along these lines, distinguishing the illness ahead of time, and treatment with the assistance of both norm and current method should be picked. In this work, we have use the UCI AI vault Cleveland coronary illness data set which has 303 example and 76 credits. For this proposed work we utilize the

Data acquire idea for determination of best property and developing of the chose highlights utilizing python. This work shows the hole of exploration on recognition of coronary illness dependent on python Boa constrictor guide. The different philosophy measures which has been used for the model of heart datasets. This outcome shows that choice tree classifier is incredible and exact for the expectation of UCI archiveCleveland heart dataset.

Author: ZHANQUAN SUN; CHAOLI WANG (Member, IEEE); YANGYANG ZHAO;

ECG signal automatic classification is a very famous research topic and also some of the effective research work has been done on it. Most of the recent work mainly focused on single ECG label classification, i.e, one ECG signal record correlates to one particular label. In this work, one ECG signal usually accept the several cardiovascular diseases atthe same time. It is more complicated to study multi-label ECG signal classification. Several researchworks have been done on this topic. To determine the multi-label ECG signal classificationproblems, we propose a novel ensemble multi-label classification model in this work. The model joins several multi-label classification methods to produce a high-rateperformer classifier. Mutual information can be usedto measure the weight of each of the classifier. At last, the ensemble multi-label classification model is utilizing a clinic ECG signal dataset. The final report shows that the overall classification performanceis has been expanded and It provides a reliable analysis method for multi-label ECG signal automatic classification.

Authors: MonikaRani;Ekta;Reeta Devi(2017).

support vector machine (SVM)classifier is mainly proposed for the categories of two typesof arrhythmias likepremature ventricular contraction(PVC) and atrial premature contraction (APC).The ECG signal trapped from the MIT-BIHdatabase is manipulated by different types of noises like baseline drift, power line interference and musclinecontraction etc.Low pass and high pass filters are utilised for the decomposition of the signal.Discrete wavelet transform (DWT) is mainly used for the feature extraction of the ECG signal.The DWT utilises features that are derived using DWT and these features. This SVM shows high accuracy rate for the PVC and APCclassification.

Authors: Naveen Kumar Dewangan, S; P. Shukla, (2016).

artificial neural network (ANN) based classifier is evolved, where discrete wavelet transform (DWT) is utilised for preprocessing and feature extraction purposes and neural network designed is worn to classify five categories of arrhythmias namely Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), Paced Beat (PB), Atrial Premature Beat (APB) and First-degree AV Block (AVB) beats apart from normal (N) beats. In ECG analysis, this work aim is to establish them degree of accuracy and induced more number of heart diseases that can be categorised. There are several Numbers of algorithms have been developed to classify ECG heartbeat based on the number of features that are extracted from the ECG signals.

Author: Atrija Singh; DebanshuBhowmick, (2017)

The conventional methodology incorporates wavelet decay as the transcendent component extraction strategy yields characterization exactness of 81.8%. This work proposed the investigation utilizing entropy and Hjorth descriptor that gives higher characterization rate at 82.9%. In this work, we use a few number of highlights that had been utilized independently previously.

CHAPTER 3

AIM OF THE PROJECT

This project aims to provide an algorithm for the detection of Cardiovascular disease that operates directly on ECG data. Nowadays most of the people are affected by the cardiovascular diseases in world wide. The heart stroke and heart attack are some of the examples of the cardiovascular diseases. There are many deaths are happening due to this CVD, So the detection and prediction of this types of diseases will save many lives. The damage of the heart muscle causes the cardiac arrhythmia. When there is a block in heart veins, then the flow of oxygen to the heart muscle also gets blocked, which leads to damage of the heart muscle. The damage is irreversible, so it may lead to death. Cardiovascular diseases (CVD) are increased drastically among all other Non-Communicable diseases, which are peaked over the past few years. In 2018, around 17.9 million that is equal to 31% of population worldwide of the people have died worldwide due to CVDs.

The manual detection and classification these cardiovascular diseases will take such a long period of time. Having a programmed framework that analyse different kinds of heart arrhythmias would help cardiologists to start suitable preventive measures and to improve the investigation of cardiovascular sickness. The electrocardiogram signals are utilized as the primary sources to identifies the arrhythmia. In this paper acquaints another methodology with recognize and group naturally cardiovascular arrhythmias.

CHAPTER 4

MATERIALS AND METHODOLOGY

4.1 INTRODUCTION

This chapter mainly deals with hardware of the project. For this project we used the MATLAB version R2020b software as our software tool. This MATLAB software explained in the following topics.

4.2 HARDWARE REQUIREMENTS

The most well-known arrangement of prerequisites characterized by any working framework or programming application is the actual PC assets, otherwise called equipment. The minimal hardware requirements are as follows,

Operating system: 64 bits

Processor : AMD pro

Main memory : 4GB RAM

Key board : 104 keys

4.3 SOFTWARE REQUIREMENTS

Software requirements deals with defining resource requirements and prerequisites that needs to be installed on a computer to provide functioning of an application. The minimal software requirements are as follows,

1. Front end: MATLAB

2. Dataset: csv

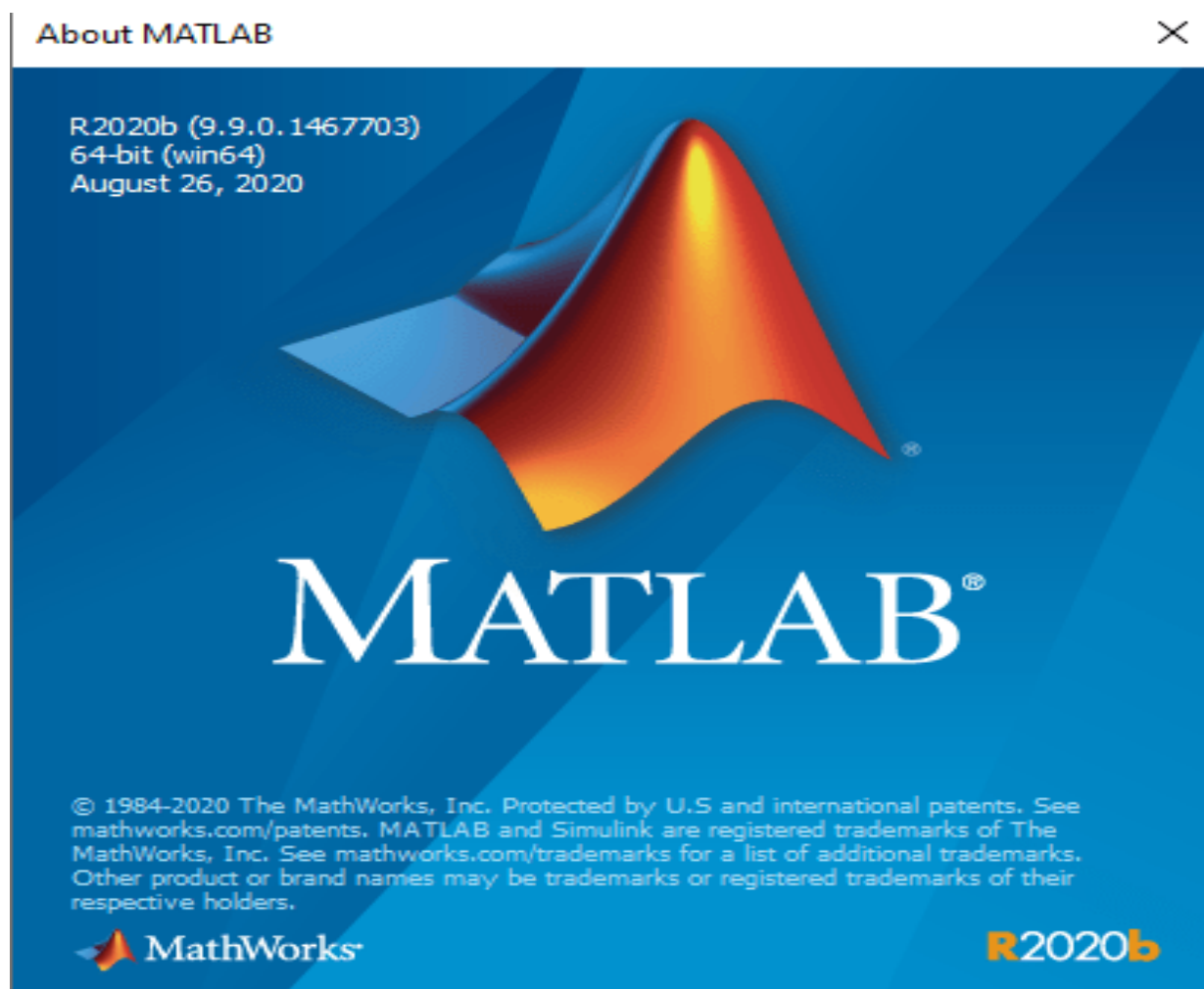
3. Operating System: Windows 10

4.4 TECHNOLOGY USED:

4.4.1 MATLAB Programming

MATLAB is abbreviated as Matrix Laboratory. MATLAB was invented by mathematician and computer programmer Cleve Moler. MATLAB programming language is a building block of MATLAB software. "Command Window" in MATLAB application is the usage that involves an interactive mathematical shell or executing text files containing MATLAB code. MATLAB application can support object-oriented programming that includes classes, inheritance, virtual dispatch, packages, pass-by-value semantics, and pass-by-reference semantics. The structure datatype also available in

MATLAB application. For this project we used the latest version of the MATLAB software which is MATLAB R2020b.



MATLAB INTROUDCTION

Fig 4.1

4.2.2 Features of MATLAB

The MATLAB application have more features and some of the useful features are given the following list.

Some of the features of MATLAB are,

- Mathematics calculation
- Computation
- Algorithm development
- Modelling,

- Simulation
- Prototyping
- Data analysis,
- Exploration data
- Visualization of inputs
- Scientific calculations
- Engineering graphics
- Application development including Graphical User Interface building

4.2.3 Characteristics of MATLAB

- MATLAB is an interactive system and array is basic data element of MATLAB. The MATLAB is used to solve many technical computing problems. Especially used to solve the problems with matrix and vector formulations. Within a fraction of the time it would take to write a program in a scalar non-interactive language such as C. The result will be given in the short span of time.
- Originally the MATLAB was written to provide easy access to matrix software developed by the LINPACK and EISPACK projects, which together represent the state-of-the-art in software for matrix computation in studies.
- The evolution of MATLAB has developed over a period of years with input from many users. MATLAB is the standard instructional tool in mathematics, engineering, and science in university environments.
- MATLAB tool is the great choice in industries like research, development and analysis for high productivity.
- The most important feature of MATLAB is that have the family of toolboxes. Those toolboxes allow us to learn and apply specialized technology in our fields.
- Toolboxes are also known as collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems.
- Signal processing, control systems, neural networks, fuzzy logic, wavelets, simulation and mathematics are the some of the field that are available as toolbox.

4.4.4 Uses of MATLAB

MATLAB is widely used as a computational tool in science and engineering encompassing the fields of physics, chemistry, math and all engineering streams. It is used in a range of applications including –

- Signal Processing and Communications
- Image and Video Processing
- Control Systems
- Test and Measurement
- Computational Finance
- Computational Biology

4.5 Applications of MATLAB

MATLAB is a very useful tool for Artificial Intelligence, Data analytics, Robotics, Image processing, Wireless communication, satellite communication, deep learning and Machine learning. It is used in wide range of uses. Even though it is used by circuit branches and mechanical in the engineering domain to solve problems by comparing to other departments but the application of this MATLAB is very vast. MATLAB as the name suggests is the Matrix or an array matrix laboratory and array is the basic element of this application.

4.5.1 Signal Processing

DSP toolbox is used to provides algorithms for the design and simulation of signal processing systems and the tools which are presents in these toolboxes are used in this designing process. DSP toolbox will provide MATLAB functions, MATLAB Simulink functions and System objects. The signal processing toolbox includes design methods for specialized FIR and IIR filters, FFTs and DFT, DSP techniques for processing and creating real-time prototypes. This tool has a multiple predefined multiple built-in function that can reduces the time and work more efficiently.

4.5.2 Image processing

The very commonly used application of MATLAB is image processing. We use MATLAB in code scanners, editing the background of old pictures, recreating blurred back ground, face detection, image

restoration, remake the colours of ancient drawings and image enhancement. The digital images processing (DSP) is playing a vital role in satellite communication also. In satellite communication the image processing is used to decode the images transmitted by the satellites.

4.5.3 Mapping

Mapping has multiple applications in various domains. For example, in Big data, the MapReduce tool is quite important which has multiple applications in the real world. Financial fraud detection, regression models, predicting techniques in social media, data monitoring, etc., these are the examples where the mapping toolbox of MATLAB is used.

4.5.4 Control Systems

The Systems nature can be obtained. Factors such as closed-loop, open-loop, its controllability and observability, bode plot, Nyquist plot, etc., can be obtained. Control systems toolbox is used to visualize the Various controlling techniques such as PD, PI and PID. The frequent domain and time domain analysis can be done by this control System toolbox and we can visualize the results as analysis graph.

4.6 DEEP LEARNING IN MATLAB

MATLAB makes deep learning easy. This toolbox is used to work with algorithms which have large amount of data that is large datasets. MATLAB has capable to handle and manage large datasets. MATLAB also provides specialized toolboxes for working with machine learning, deep neural networks, computer vision, computer networking and automated driving with less number of lines of code. MATLAB is to learn and gain expertise in the area of deep learning.

4.7 LSTM ARCHITECTURE

We propose an ECG arrhythmia classification method using Long-short term memory network. The classifier achieved the highest accuracy and the lowest loss. The main advantage of LSTM compared to its predecessors is that it automatically detects the important features without any human supervision. For example, given many pictures of ECG it learns distinctive features for each class by itself. LSTM is also computationally efficient.

4.7.1 Block Diagram of working

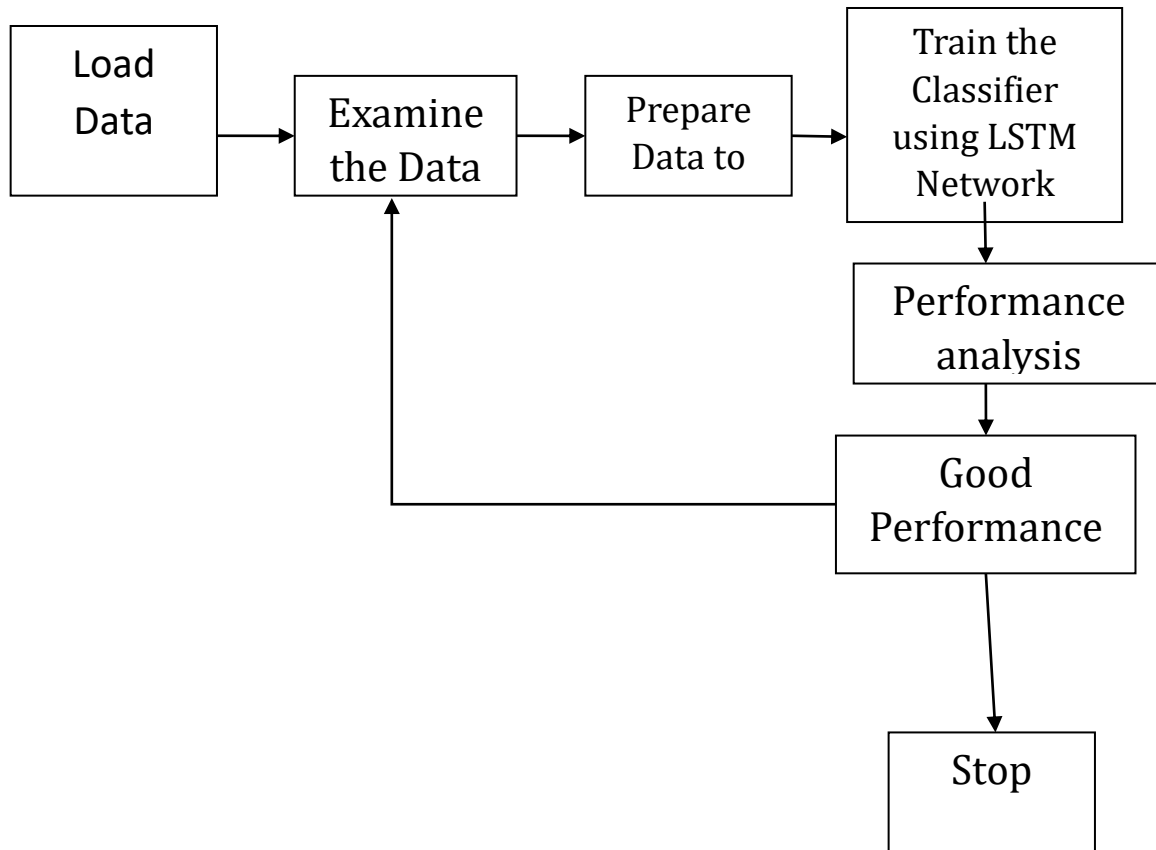


Fig 4.2

BLOCK DIAGRAM OF THE PROPOSED WORK

Block diagrams are fundamental building blocks for any process. Block diagram giving the flow of the process. The first and foremost step is to load the data to system. Initially the raw dataset is used to train the network. Once the data is loaded to the system the training process is about to begin. The datasets are splited for training and testing. Then design the classifier network and with the input datasets. If we attain the good accuracy as we expect than start test the datasets which splited of test. If the accuracy of the network is not good then do feature extraction using time domain spectrum.

4.8 IMPLEMENTATION

4.8.1 Module Description

The proposed work is to detect and to classify the Cardiac Arrhythmia using a Long short term memory neural network.

4.8.2 Collecting Dataset

The data for the proposed method is taken from the [physionet.com/challenge2017/ECG](https://physionet.org/challenge2017/) signals. The ECG signals from the dataset is sampled at 300 Hz. The challenge2017 dataset was already classified by the experts into four categories. There are normal signal(N), Atrial filtration signal (AFib), noise signal and others. By removing the noise and other signal we can make our database that consist of only normal and atrial fibrillation signal. The signals in the datasets are in the form of .cvs(.mat) files are taken and loaded to the MATLAB software. The data were already labeled as whether the ECG signal is normal or affected by arrhythmia. So, there is no need to classify the data manually. Now the signals from the dataset are noiseless.

4.8.3 Loading and Examine the data

Load the data from the PhysioNet.com website that contains the ECG signals in the appropriate format to load to MATLAB. After loading the signal that creates two work space in MATLAB window. They are Signals and Labels. Signals is a cell array that is used to store the ECG signal values. The size of this cell array is 1*9000 and some of the signals cell array has the size of 1*18000. The label work space has the label of the category. For example, the below diagram for the normal signal and Atrial fibrillation signal to visualize the difference. The signal will help us to plot the corresponding signal plot. This label will help us to our signal to name whether it come under normal signal or AFib signal some respective features.

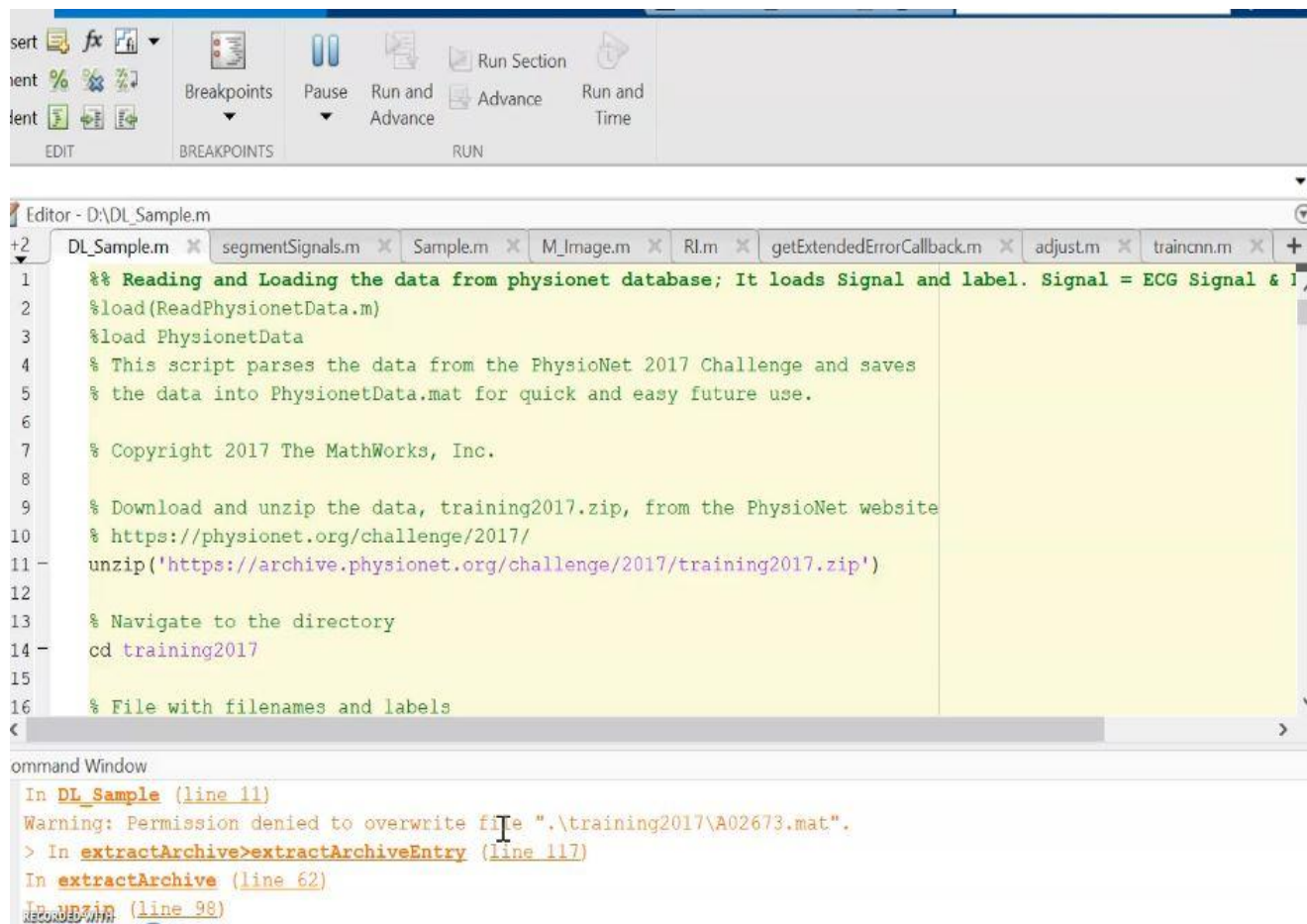


FIG 4.3 LOADING DATA

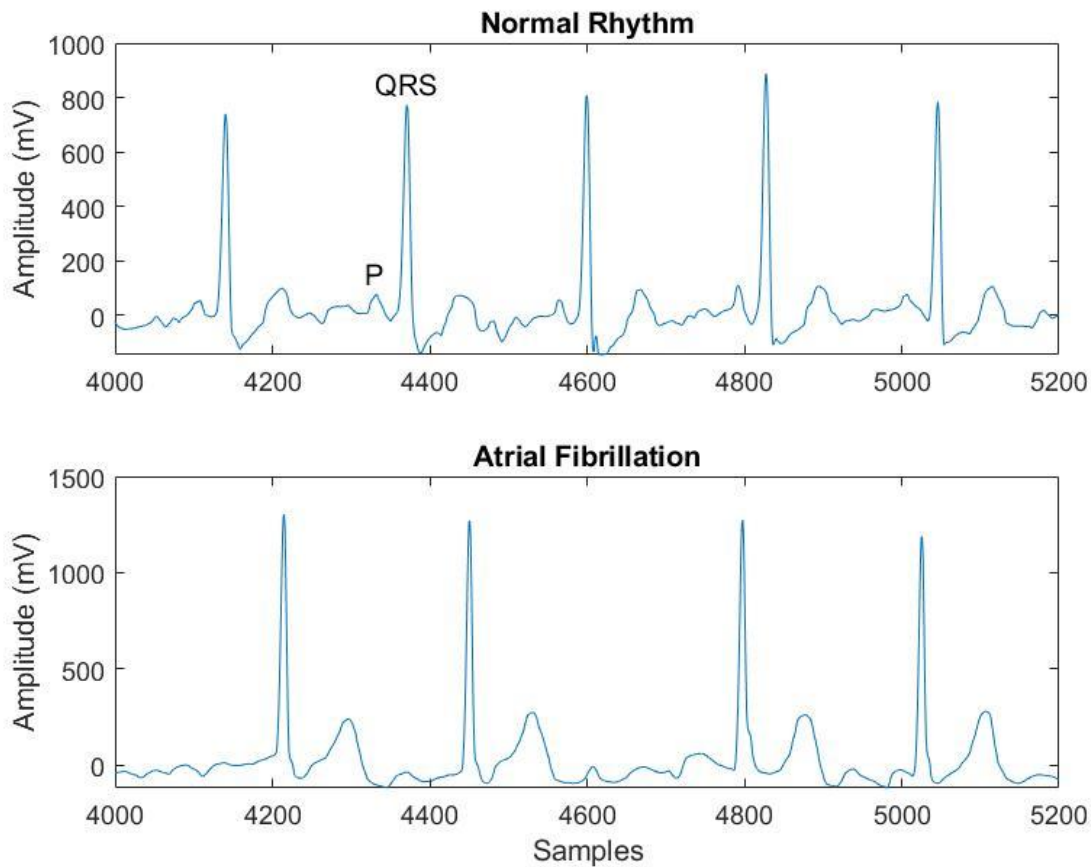


Fig 4.4

VISUALIZATION OF THE SAMPLE

4.8.4 Prepare the data for training

The data is split into many mini batches for train the network. When we print the number of normal signals and AFib affected signal from the input, the count of AFib signal is very less, it would affect the overall accuracy of the LSTM-network. Apply the `segment Signals()` function to the input ECG signals to avoid excessive padding or truncating. The raw input signals are used to design a classifier network. Split the signals into a training set and a testing set to train the classifier and test the accuracy of the classifier on new data respectively. Once the signals are segmented using `segment signal ()` function, again the normal and AFib signal are plotted to visualize the signals after truncation.

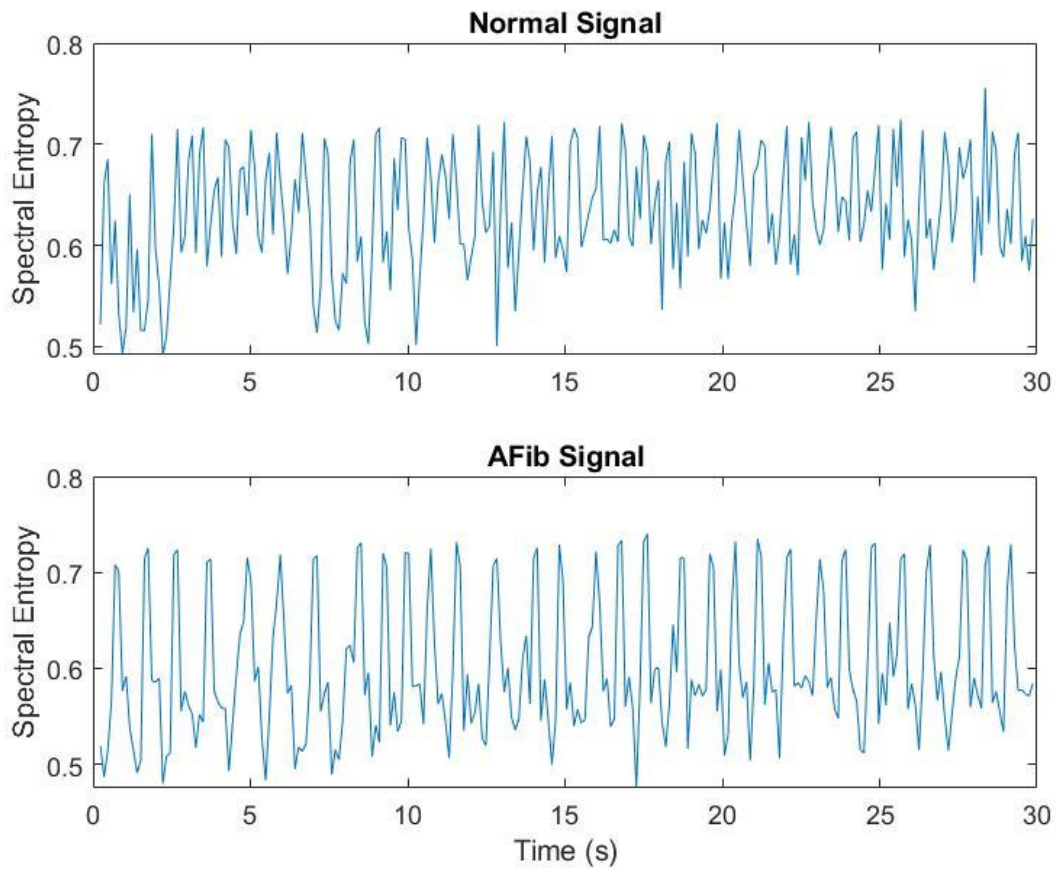


Fig 4.5
SPECTRAL ENTROPY OF THE SAMPLE

The training consists of ten iterations. Initially the accuracy of the network is very low and loss is very high. Once the number of iterations is increased then the accuracy is getting high. Training may take approximately two and half hours for the first attempt. The learning rate of the network is 0.01. If we increase the learning rate or changing the learning rate, the accuracy of network will also change.



FIG 4.6
INITIAL STAGE OF TRAINING

4.8.5 LSTM Network Architecture

LSTM is abbreviated as Long-short term memory neural network. Here in this proposal, we used the bidirectional LSTM layer (bi-lstm

Layer), as it looks at the sequence in both forward and backward directions.

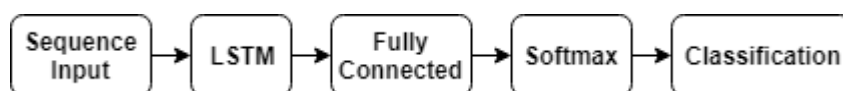


Fig 4.7
LSTM ARCHITECTURE

This LSTM consist of five layers. They are listed below

1. Sequence Input layer
2. BiLSTM layer
3. Fully Connected layer
4. SoftMax layer

5. Classification Output layer

The sequence layer is act as the input layer to take the input for training and testing. The BiLSTM is the bidirectional layer which is used to move in forward and back direction. The fully connected layer is used to define the number of classes available in the output. The classification layer is also called as output layer. The dataset which splitted of minibatches are used to train the network. If we take minibatch less than 150 then the accuracy of the system will also reduce. An 'InitialLearnRate' of 0.01 helps speed up the training process. We can train the network with different values of learning rate also. This will give us the different accuracy percentage.

4.8.6 Train and Test the LSTM Network

The LSTM network training is the prolonged process. since there is a large dataset available . Early the training process took more than two hours to complete.

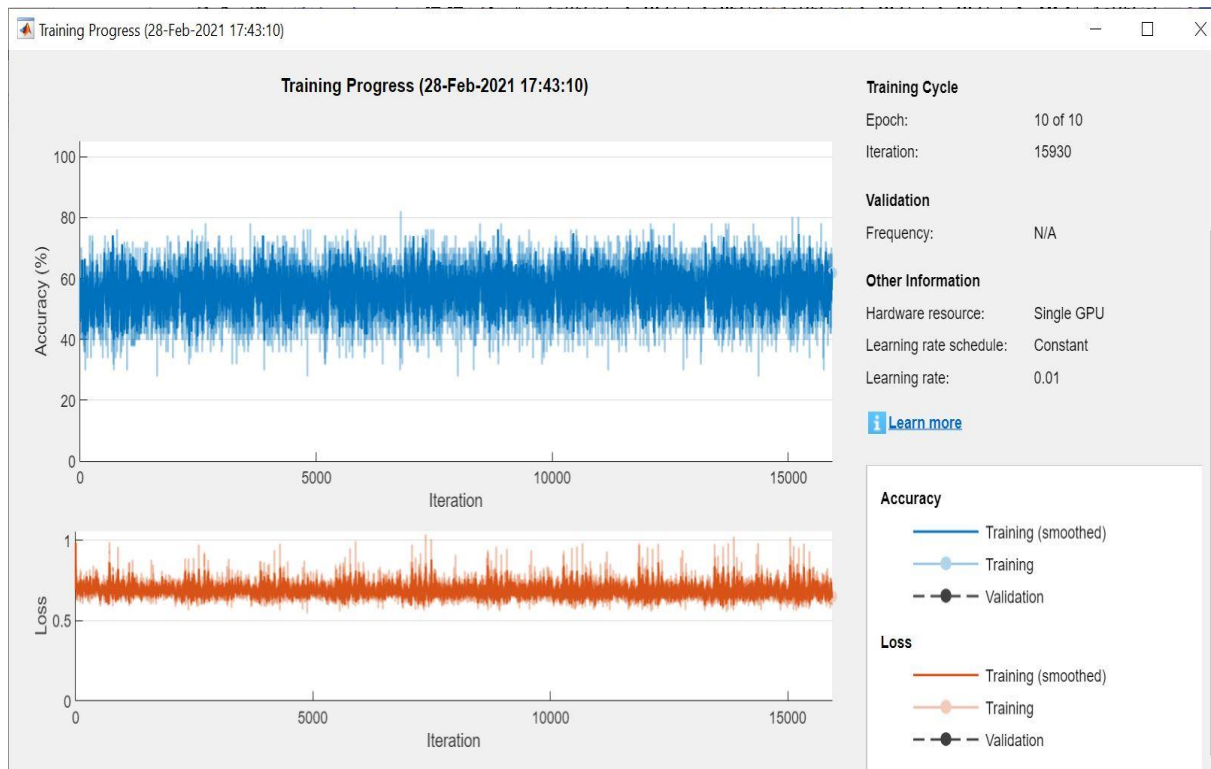


Fig 4.8

TRAINING PROGRESS

The loss rate is huge and the accuracy of the testing network is minimum at the initial stage of training. Calculate the training accuracy, which denotes the accuracy of the classifier on the signals on which it is been trained. In classification problems, confusion matrices are used to envisage the performance of a classifier on the dataset for which the original values are known. The Target Class

is the truth label of the signal, and the Output Class is the label is given to the signal by the network. The axes labels denotes the class labels, AFib (A) and Normal (N). The training accuracy of the LSTM network is 61%. When the network completes the training process it will starts with the testing process.



Fig 4.9
CONFUSION MATRIX OF TRAINING SAMPLE

Confusion matrix is a matrix that can be used for predicting the performance of a classification network, where N is the number of target classes. This matrix matches the actual target values with those detected by the machine learning model. This gives us a comprehensive view of how suitably our classification model is performing and what types of errors it is showing

Coming to the binary classification problem, we would have a 2 x 2 matrix with four sets of values

The row variable has two values: Positive or Negative
The columns denotes the actual values of the target variable
The rows denotes the predicted values of the target variable

True Positive (TP)

The predicted value matches the actual value
The original value is positive and also the model classified as a positive value
True Negative (TN)

The predicted value matches the actual value
The original value is negative and the model also detected as a negative value
False Positive (FP) – Type 1 error

The predicted value was wrongly predicted
The real value was negative but the model classified as a positive value
Also known as the Type 1 error
False Negative (FN) – Type 2 error

The predicted value was wrongly predicted
The true value was positive but the model also detected as a negative value
Also known as the Type 2 error

Training network:

True Positive (TP) = 4277; meaning 4277 positive class data points were correctly classified by the model

True Negative (TN) = 293; meaning 293 negative class data points were correctly classified by the model

False Positive (FP) = 161; meaning 161 negative class data points were incorrectly classified as belonging to the positive class by the model

False Negative (FN) = 4145; meaning 4145 positive class data points were incorrectly predicted as belonging to the negative class by the model

4.8.7 Feature Extraction

In order to increase the training and testing accuracies of a network we improved our method by incorporating feature extraction techniques.. To resolve that, which features to extract, this example modifies an approach that estimates time-frequency images, such as spectrograms, and uses them guiding the LSTM network. To envisage the spectrogram of individual type of signal.

```
fs = 300;  
  
figure  
subplot(2,1,1);  
pspectrum(normal,fs,'spectrogram','TimeResolution',0.5)  
title('Normal Signal')  
  
subplot(2,1,2);  
pspectrum(aFib,fs,'spectrogram','TimeResolution',0.5)  
title('AFib Signal')
```

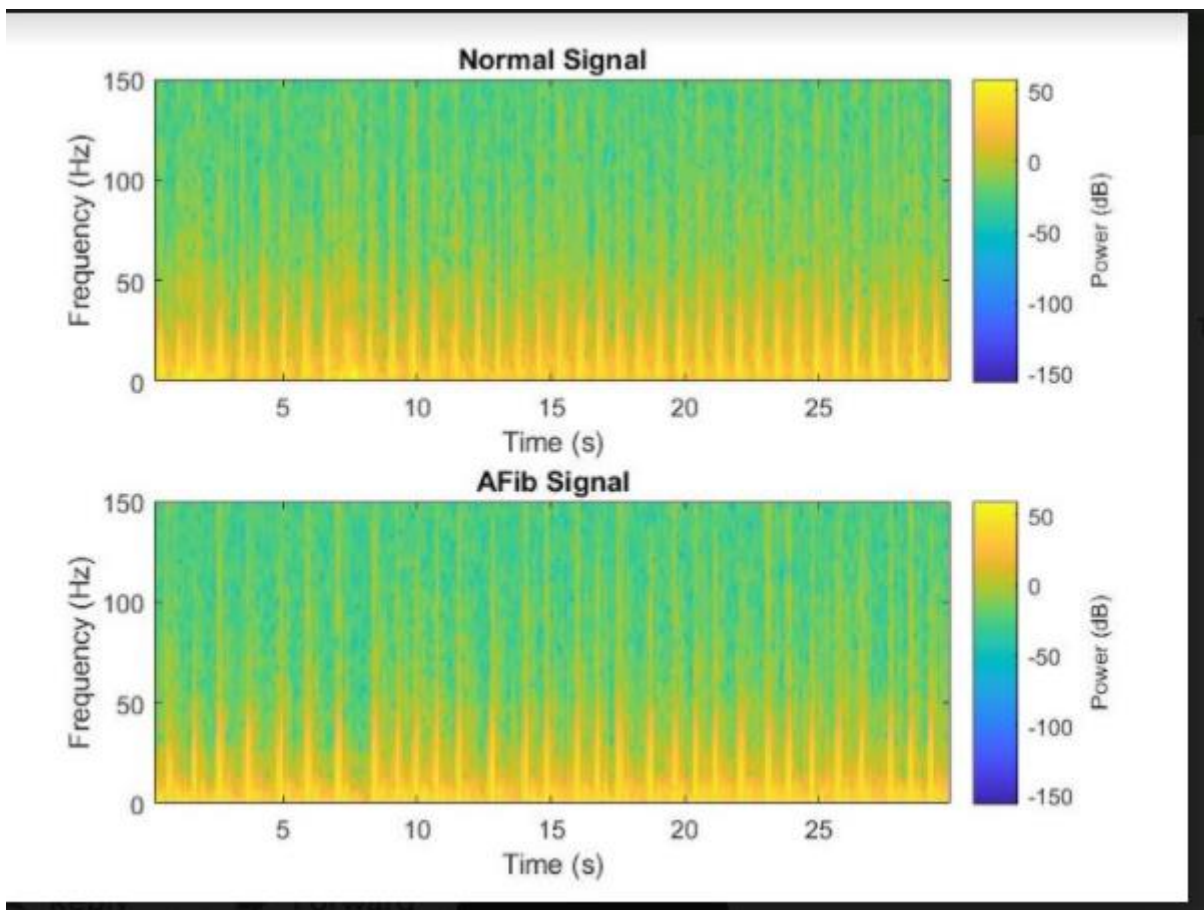


FIG 4.10
FEATURE EXTRACTION

CHAPTER – 5

RESULTS AND DISUCSSION

Confusion matrix is a matrix that can be used for predicting the performance of a classification network, where N is the number of target classes. This matrix matches the actual target values with those detected by the machine learning model. This gives us a comprehensive view of how suitably our classification model is performing and what types of errors it is showing

Coming to the binary classification problem, we would have a 2×2 matrix with four sets of values

The row variable has two values: Positive or Negative

The columns denotes the actual values of the target variable

The rows denotes the predicted values of the target variable

True Positive (TP)

The predicted value matches the actual value

The original value is positive and also the model classified as a positive value

True Negative (TN)

The predicted value matches the actual value

The original value is negative and the model also detected as a negative value

False Positive (FP) – Type 1 error

The predicted value was wrongly predicted

The real value was negative but the model classified as a positive value

Also known as the Type 1 error

False Negative (FN) – Type 2 error

The predicted value was wrongly predicted

The true value was positive but the model alos detected as a negative value

Also known as the Type 2 error

Training network:

True Positive (TP) = 483; meaning 483 positive class data points were correctly classified by the model
True Negative (TN) = 36; meaning 36 negative class data points were correctly classified by the model
False Positive (FP) = 7 ; meaning 7 negative class data points were incorrectly classified as belonging to the positive class by the model

False Negative (FN) = 454; meaning 454; positive class data points were incorrectly predicted as belonging to the negative class by the model

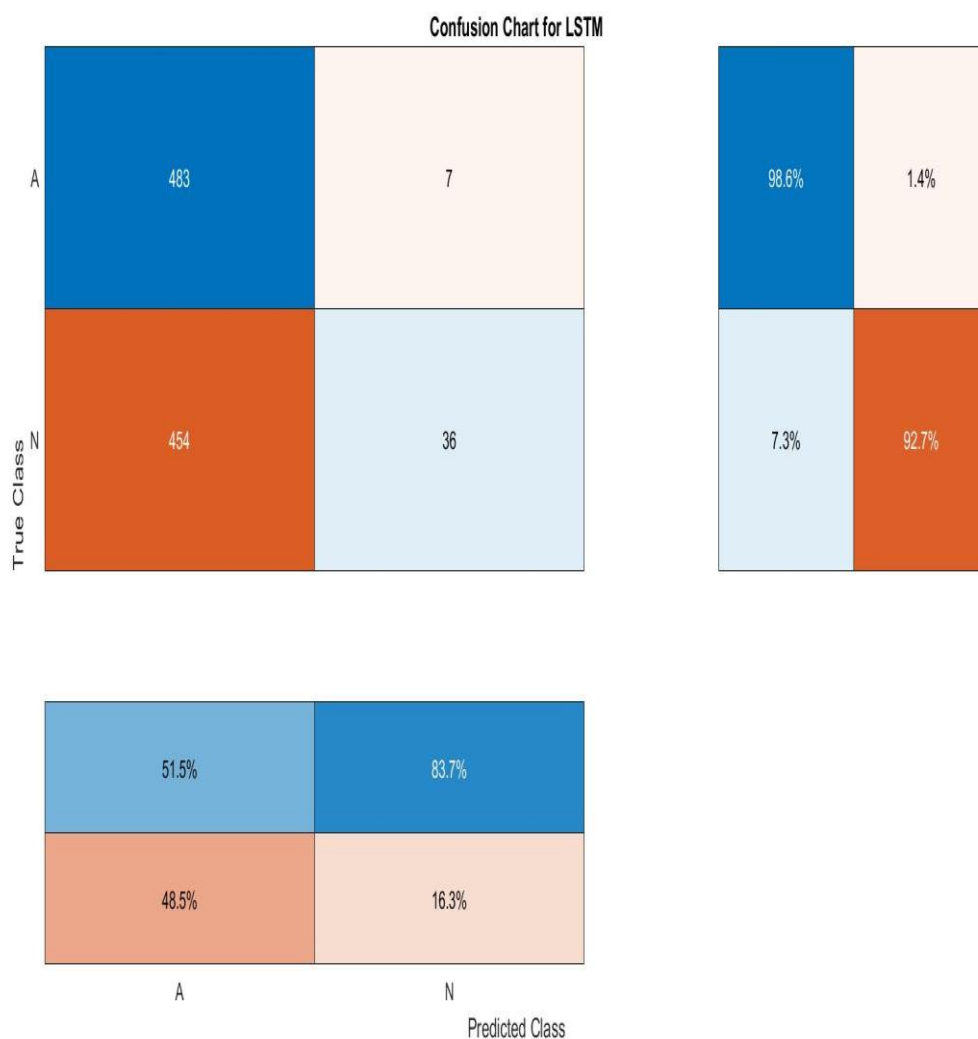


FIG 5.1

CONFUSION MATRIX OF THE TESTING SAMPLE

CHAPTER – 6

CONCLUSION

The results obtained portrays that the proposed system produces precise output and is user-friendly. LSTMs have additional retentive power over other recurrent architectures because of their ability to overcome the vanishing gradient problem. Stacking recurrent hidden layers allows for the processing of the data at different time scales and generation of a richer set of temporal features. In this work we have investigated the applicability of Deep LSTM networks for detecting cardiac arrhythmias in ECG signals. The technique offers several advantages in that these networks are quite fast once they are trained, and do not require knowledge of the abnormalities, hand crafted features or pre-processing of the data. The proposed LSTM network achieved 96% accuracy and 99.6% specificity. Additionally, if this method is to be used by a medical practitioner, recall may be given higher priority, although the number of false positives tend to increase substantially. Overall, it appears that stacked LSTMs may be a viable candidate for anomaly detection in ECG signals.

REFERENCES

1. Addison P, Watson J, Clegg G, Holzer M, Sterz F, Robertson C. Evaluating arrhythmias in ECG signals using wavelet transforms. *IEEE Eng Med Biol Mag.* 2000;19(5):104-109. <https://doi.org/10.1109/51.870237> 16.
2. Jung Y, Tompkins W. Detecting and classifying life-threatening ECG ventricular arrhythmias using wavelet decomposition. Paper presented at: 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2003; 3; Cancun, Mexico. 17.
3. Luz E, Schwartz W, Cámara-Chávez G, Menotti D. ECG-based heartbeat classification for arrhythmia detection: a survey. *Comput Methods Programs Biomed.* 2016;127:144-164. <https://doi.org/10.1016/j.cmpb.2015.12.008> 18. Ye C, Kumar B, Coimbra M. Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Trans Biomed Eng.* 2012;59(10):2930-2941. <https://doi.org/10.1109/tbme.2012.2213253> 19.
4. Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng in Med Bio.* 2001;20(3):45-50. 20. Mallat S. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Mach Intell.* 1989;11(7):674-693. <https://doi.org/10.1109/34.192463> 21. Josko A. Discrete wavelet transforms in automatic ECG signal analysis. Paper presented at: IEEE Instrumentation and Measurement Technology Conference; 2007; Warsaw, Poland. 22. Stantic D.
5. Detecting abnormal ECG signals utilising wavelet transform and standard deviation. *World Acad Sci Eng Technol.* 2012;71:208-214. 23. Lu Z, Kim DY, Pearlman W. Wavelet compression of ECG signals by the set partitioning in hierarchical trees algorithm. *IEEE Trans Biomed Eng.* 2000;47(7):849-856. <https://doi.org/10.1109/10.846678> 24. Martis R, Acharya U, Min L. ECG beat classification using PCA, LDA, ICA and discrete wavelet transform.
6. *Biomed Signal Process Control.* 2013;8(5):437-448. <https://doi.org/10.1016/j.bspc.2013.01.005> 25. Jatmiko W, Nulad W, Elly M, Setiawan I, Mursanto P. Heart beat classification using wavelet feature based on neural network. *WSEAS Trans Syst.* 2011;10(1):17-26.

Document Information

Analyzed document 37130160,37130184_project_report_tocheck_plg.docx
(D101746262)

Submitted 4/15/2021 6:53:00 PM

Submitted by S.Poornapushpakala

Submitter email poornapushpakala.enc@sathyabama.ac.in

Similarity 9%

Analysis address poornapushpakala.enc.sathya@analysis.urkund.com

Sources included in the report

	71060521006-TS.pdf		3
SA	Document 71060521006-TS.pdf (D36117888)		
W	URL: https://www.medigoo.com/articles/tachycardia/shri New - 16.docx Fetched: 4/15/2021 6:57:00 PM		1
SA	Document shri New - 16.docx (D90452028) URL: https://www.researchgate.net/publication/324988448 Cardiac Arrhythmia Classificati ...		1
W	Fetched: 9/25/2019 10:37:32 AM URL: https://www.researchgate.net/publication/312326362 ECG arrhythmia classification u ...		2
W	Fetched: 12/9/2019 6:33:10 AM URL: https://www.researchgate.net/publication/273031959 Classification Of Ecg Arrhythmi ...		1
W	Fetched: 12/26/2020 5:47:00 AM URL: https://www.researchgate.net/publication/324056729 A novel wavelet sequence based u ...		2
W	Fetched: 11/26/2019 7:03:40 AM updated file.docx		3
SA	Document updated file.docx (D21488114)		
SA	king elango Thesis(07.01.2018).docx Document king elango Thesis(07.01.2018).docx (D34450196)		1
SA	mtech file.docx Document mtech file.docx (D40930967) URL: https://www.researchgate.net/publication/291385336 Arrhythmia Recognition and Clas ...		1
W	Fetched: 11/6/2019 7:48:23 AM king elango Thesis(19.01.2018).docx		1
SA	Document king elango Thesis(19.01.2018).docx (D34917203)		
	URL: https://www.mdpi.com/1424-8220/20/5/1461/pdf		