TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS NO: 073/MSCS/669

Automated Heart Arrhythmia Classification From Electrocardiographic Data Using Deep Neural Networks

by

SUMITA GHIMIRE

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN COMPUTER SYSTEM AND KNOWLEDGE ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

LALITPUR NEPAL

September 2021

Automated Heart Arrhythmia Classification from Electrocardiographic Data Using Deep Neural Networks

By:
Sumita Ghimire
(073/MSCS/669)

Thesis Supervisor:
Prof. Dr. Shashidhar Ram Joshi

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Computer System and Knowledge Engineering

Department of Electronics and Computer Engineering
Institute of Engineering, Pulchowk Campus
Tribhuvan University
Lalitpur, Nepal

September, 2021

RECOMMENDATION

The undersigned certify that they have read and recommended to the Department of Electronics and Computer Engineering for acceptance, a thesis entitled "**Automated Heart Arrhythmia Classification from Electrocardiographic Data using Deep Neural Networks"** , submitted by **Sumita Ghimire** in partial fulfillment of the requirement for the award of the degree of "**Master of Science in Computer System and Knowledge Engineering**".

…………………………………………………

Supervisor: Prof. Dr. Shashidhar Ram Joshi,

Department of Electronics and Computer Engineering,

Institute of Engineering, Pulchowk Campus

...………………………………………….………..

External Examiner: Dr. Pradip Paudyal,

Deputy Director,

Nepal Telecommunication Authority

...………………………………………….………..

Committee Chairperson: Assoc. Prof. Dr. Nanda Bikram Adhikari,

Department of Electronics and Computer Engineering,

Institute of Engineering, Tribhuvan University

Date: September 2021

DECLARATION

I declare that the work hereby submitted for Masters of Science in Computer Science and Knowledge Engineering (MSCSKE) at IOE, Pulchowk Campus entitled "**Automated Heart Arrhythmia Classification from Electrocardiographic Data Using Deep Neural Networks**" is my own work and has not been previously submitted by me at any university for any academic award. I authorize IOE, Pulchowk Campus to lend this thesis to other institution or individuals for the purpose of scholarly research.

Sumita Ghimire
073/MSCS/669
Date: September, 2021

DEPARTMENTAL ACCEPTANCE

This thesis entitled "**Automated Heart Arrhythmia Classification from Electrocardiographic Data Using Deep Neural Networks**", submitted by **Sumita Ghimire** in partial fulfillment of the requirement for the award of the degree of "**Master of Science in Computer System and Knowledge Engineering**" has been accepted as a bonafide record of work independently carried out by her in the department.

………..…………………………………

Prof. Dr. Ram Krishna Maharjan

Head of the Department

Department of Electronics and Computer Engineering,

Pulchowk Campus,

Institute of Engineering,

Tribhuvan University, Nepal.

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere thanks to my supervisor, **Prof. Dr. Shashidhar Ram Joshi** for his excellent guidance and valuable recommendations on the completion of this thesis work. His thoughtful comments, consistent support and supervision have contributed a lot in this work.

I would like to also express my deep gratitude to Department of Electronics and Computer Engineering for incorporating thesis as a part of our syllabus. I would like to express my gratitude to **Prof. Dr. Subarna Shakya**, **Assoc. Prof. Dr. Diwakar Raj Pant, Assoc. Prof. Dr. Sanjeeb Prasad Panday**, **Asst. Prof. Dr. Aman Shakya**, **Asst. Prof. Dr. Basanta Joshi** and other faculties for their encouragement and precious guidance.

I am thankful to our Head of Department **Prof. Dr. Ram Krishna Maharjan** and program coordinator **Assoc. Prof. Dr. Nanda Bikram Adhikari** for providing suitable platform to carry out this thesis.

Lastly I want to thanks my husband, my family and my friends whose consistent inspirations motivate me for doing this thesis.

ABSTRACT

Arrhythmia is the medical condition where heart beats in an irregular pattern. Arrhythmia is one of the common sources of the Cardio Vascular diseases. To survive from the arrhythmia, the keys are early detection and timely treatment. ECG stands as a diagnostic tool for the detection of the arrhythmia. Human intervention to the ECG is error prone as well as tedious. With the help of the development of the technology, cost effective automated arrhythmia detection framework can be deployed. There are many machine learning as well as deep learning models which can effectively differentiate among various types of heartbeats. Various deep learning models has shown that there is an ease way for predicting arrhythmia which do not require feature engineering and is effective. In order to build automated heartbeat classification model several factors has to be considered which includes data quality, heartbeat segmentation range, data imbalance problem, intra and inter-patients variations and identification of supraventricular ectopic heartbeats from normal heartbeats. This thesis incorporates all of these challenges. In this method, a hybrid method of neural network was deployed. Features were extracted by the two CNNs having two filter sizes. RNN which is BiLSTM was used to classify the ECG signals. Dual channel CNN was used to extract both the temporal as well as frequency patterns. The extracted features were added with the RR information before giving the input to the RNN that are mainly done to classify between the S-type and N-type heartbeats. In particular, BiLSTM learns and extracts hidden temporal dependency between the heartbeats by processing the input RR interval sequence in both the directions. Instead of using raw individual RR-intervals, mutual-connected temporal information provides stronger and more stable support for identifying the S-type heartbeats. The loss used is the focal loss to handle the class imbalance. The results prove that the research of heartbeat classification presented in this thesis brings practical ideas and solutions to the arrhythmia detection. Accuracy of the model presented in this thesis was 93%.

Keywords: *Cardio vascular diseases, arrhythmia, ECG classification, deep learning, dual channel Convolutional Neural Networks, Bidirectional Long Short Term Memory, focal loss*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AAMI	:	Advancement of Medical Instrumentation
AUC	:	Area Under Curve
BiLSTM	:	Bidirectional Long Short Term Memory
CE	:	Cross Entropy
CNN	:	Convolutional Neural Network
CPU	:	Central Processing Unit
CVD	:	Cardio Vascular Disease
CWT	:	Continuous Wavelet Transform
DNN	:	Deep Neural Network
ECG	:	Electrocardiogram
F	:	Fusion
FPR	:	False Positive Rates
GPU		Graphical Processing Unit
LSTM	:	Long Short Term Memory
MIT-BIH	:	Massachusetts Institute of Technology-Beth Israel Hospital
mV	:	Milli Volt
N	:	Normal
Q	:	Unknown
RNN	:	Recurrent Neural Network
ReLU	:	Rectified Linear Unit
ResNet	:	Residual Network
ROC	:	Receiver Operating Curve
S	:	Supraventricular ectopic
V	:	Ventricular ectopic

CHAPTER 1 INTRODUCTION

1.1 Background and Motivation

Cardiovascular disease (CVD) is nowadays a growing disease and has been the cause of many deaths worldwide. About 31% of the global deaths are from the cardiovascular diseases [1]. “As someone gets older, the human cardiovascular system is more likely to be suffered from the CVDs, because the arteries and muscle wall of the left ventricle keeps thickening and shrinking along the age factor. This calls for the less supply of the blood vessels in the arteries”. [2] One of the common sources of the cardiovascular disease is irregular heartbeat pattern. Detection of arrhythmic pattern of the heartbeat can prevent a person to survive from the cardiovascular disease. This call for the need of detecting and interpreting any heart diseases from the ECG signals.

The electrocardiogram (ECG) stands as methodical tool in the field of biomedicine for evaluation of the health of the heart of person for finding the heart related issues and arrhythmia. These conditions of the heart are dangerous for one’s life and can also take the life of the patient. [2] An ECG signal is signal that represents heart rhythms and can be used to detect arrhythmia. Arrhythmia exists where the heart functionality is not constant. This is the situation where the person’s heart rhythms are inconsistent. “Arrhythmia occurs when the heart does not pump blood effectively throughout the body”. [3] Arrhythmia does not necessarily mean that the heartbeats are too fast or too slow. It means the heartbeat is not following a pattern which it is designated to follow.

To survive from arrhythmia the possible remedies may be timely detection and treatment. However cardiologists having many years experience can sometimes mistaken between the normal and abnormal heartbeats as the manual interpretation of electrocardiography signal is simple but it is time consuming process. This process is prone to error as it also contains long-term records. Therefore the exploration and invention in the area of medicine may lead to the easier analysis of electrocardiogram data.

Many technologies have been used for the process of heartbeat classification. Most of them relay on the machine learning algorithm. The features generated are hand crafted and classification is done on the basis of the features extracted. 30 features were extracted for the classification purpose and SVM was used [4]. In [5] RR intervals were also taken in consideration for the purpose of classification. However the machine learning techniques are time consuming as the extraction of the features takes too long and feature extraction process is to be done in a careful manner. The whole process of the classification is totally dependent on the quality of information received in the process of the feature extraction. More than it is tedious, it is error prone.

In the recent years deep neural networks are used in the design of automated heartbeat classification model. Augmentation techniques were used to classify the human heartbeats for the better classification. The use of the Stacked Bidirectional LSTM with two CNNs performed better without requiring feature extraction process. [6] Different CNNs models showed better performance however the data was patient specific [7] [2]. In [8] the model was developed in the inter-patient paradigm. The accuracy was also satisfactory but the problem was with the normal heartbeat. The patients' normal heartbeat accuracy was decreased in comparison with the previous works'. Deep learning technique employed in [9] had better performance. Time-domain features were extracted from raw signals which were merged with the model generated features. In this thesis deep neural network on the MIT-BIH arrhythmia [10] dataset was used to classify among different types of heartbeats.

1.2 Problem statements

There are several challenges that must be addressed when constructing an effective model for the heartbeat classification for arrhythmia detection.

Noise Removal: ECG signal usually come with several background noise and baseline wanders. Quality of ECG recordings is directly related to the heartbeat classification performance.

Intra and inter-patient variations: The heartbeat of the same patient exhibit different pattern at different intervals, and the heartbeat of different patients exhibit different patterns. This causes difficulties in the generation of the automated heartbeat classification.

Class imbalance problem: The arrhythmias are few in number as compared to the normal type heartbeats causing the model to less train in the arrhythmias. The model will not be able to recognize patterns exhibited by arrhythmia making the model to show biasness in heartbeat classification.

Identification of S-type heartbeats: It is less likely to provide accurate identification of the S-type heartbeats from the N-types based on the morphological features.

1.3 Objective of the study

The main focus of this thesis is to incorporate all the challenges explained in the problem statement by developing an automated heartbeat classification model that can effectively differentiate among various heartbeats from the ECG signal. This work is focused on developing of the model from the dual channel Convolutional neural network and Bidirectional Long Short Term memory. The objective can be summarized as:

- To effectively build a classification model that can differentiate among various types of heartbeats.

CHAPTER 2 THEORETICAL BACKGROUND

2.1 The Human Heart

Heart is a muscular organ serving as a pump in the human body for the circulation of blood. Heart beats around hundred thousand times each day. The main function of the heart is to supply the blood around the human body which is equivalent to sixty thousand miles of blood vessels. “The main objectives of the heart are:

- To pump blood around the human body that is collected from the lungs.
- To pump blood back to the lungs from tissues in the body”. [11]

Heart is divided into four chambers, left ventricles, right ventricles, right atria and the left atria. A healthy heartbeat starts with the contraction of the atria. In the mean time, the relaxation of the ventricles takes place. This is followed by the relaxation of the atria and contraction of ventricles.

Heartbeat classification on ECG is an effective way for identifying the irregular patterns. ECG signal can be defined as the pattern of different waves and peaks of heartbeat waveform. There exist waves namely P wave, QRS section and T wave per cardiac section in ECG signal.

Heartbeat starts from the Sinus Node. P-wave is the starting of a cardiac cycle. The P-wave is characterized by a positive and slow wave. The P-wave indicates arterial depolarization. The P-wave is followed by the Q-wave which is the second event of the heartbeat. On the contrary to the P-waves, they are negative and fast waves. The Q-wave represents ventricular septal depolarization. The third one is R-wave. This is powerful, quick and positive wave. The representation of the R-wave in such a way is because of the major depolarization of ventricles. The next is S-waves which are negative and fast wave. The last wave is T-wave. T-wave represents ventricular repolarization.

Figure 2-1 shows a usual cardiac cycle. Isoelectric line is the line connecting two waves. These lines are known as segments. In one cardiac cycle, there exist two segments namely: PR segment and ST segment. These isoelectric lines are the line between two waves exclusive of the waves. There are two intervals in the cardiac cycle which are PR interval and QT interval. Interval can be defined as the time difference between the one or more waves and a segment. The duration for the completion of one cardiac cycle of a normal heart is 600 milliseconds.

Q, R and S wave in combined form is known as QRS complex. This complex is one of the most important deflections that are seen in a ECG signal. This deflection is normally on the centre of the ECG signal. The most prominent figure of an ECG signal can be called as the QRS complex. The QRS signal is the produced when the ventricles relax and the large ventricular muscles contracts. In simply words, the QRS complex corresponds to the duration of ventricular contraction or depolarization.

Along with these morphological information held by the electrical waves of the heartbeat, the information hold by R-R intervals are also crucial while classifying different types of the heartbeats. R-R interval can be defined as the interval of the present R wave and the future R wave or from one peak of the QRS complex to latter peak of the QRS complex. The vital information carrying R-R intervals are the pre, post, average and local RR. For generation of the ECG signals, electrodes are attached to the body surface and electrical activity of the heart is measured. The measured value is the electrocardiogram signal.

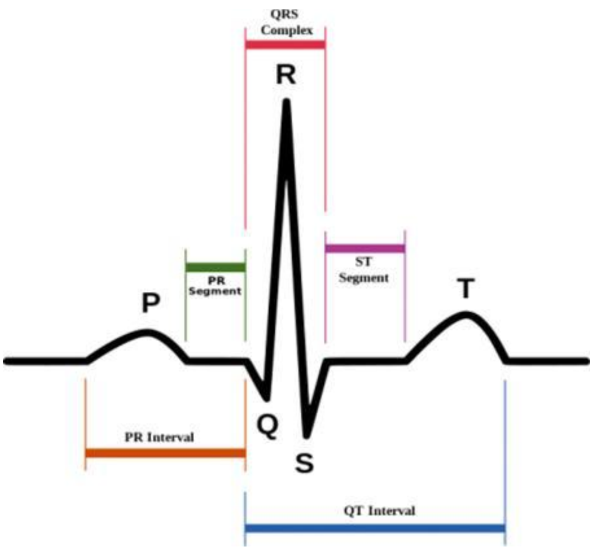


Figure 2-1 ECG Signal Over One Cardiac Cycle [12]

2.2 The 12-lead ECG measurement

For measuring the heartbeat rhythm with a higher accuracy, the standard 12-lead for ECG is used. “These are the ducts of reporting. The leads placed at the different surface have some recording that is the recordings of the electrical voltages produced by the heart. The Limb leads are the leads in which the electrodes are placed in either the arms and/or Legs of a patient. These leads are Lead I, II, III, augmented Voltage Right (aVR), augmented Voltage Left (aVL), and augmented Voltage Foot (aVF). The left six leads are V1, V2, V3, V4, V5, and V6 leads. The recordings from this leads come from the electrodes that are attached on the patient’s chest region. [13] The cardiologists are the specialist who has the idea about the streaming signal. Therefore the doctor must be assigned for the analysis of the ECG signal [14]. It is crucial to investigate the various types of the abnormality present in the ECG signals and figuring the type of the arrhythmia before proceeding for the treatment of the patient’s who is under the examination procedure [2].

The electrodes attaching to the patients skin should have the proper placement otherwise there is the possibility of the inaccuracy in the results after evaluation of the ECG signals. The patient’s health is in the top priority which can be misinterpreted with the false reports. Even a slight deviation in the placement of the electrodes from the proper positioning can cause a large negative impact on the report of the ECG signals which can be the cause of the poor and false analysis of the person’s heartbeat. This situation is not desirable and is very dangerous in the field of medical science.

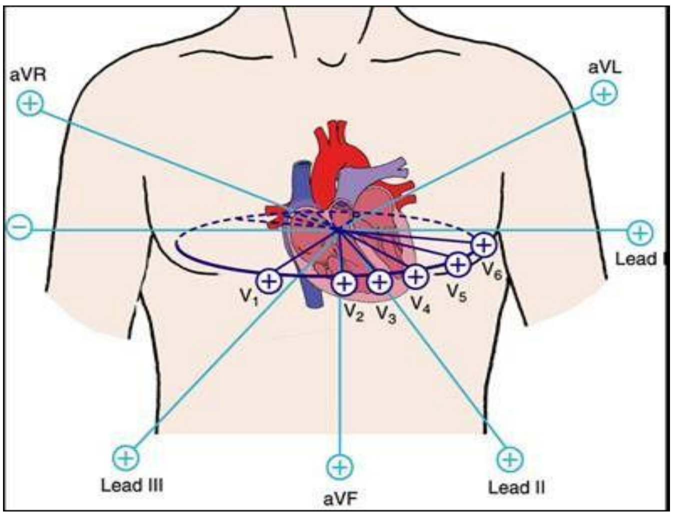


Figure 2-2 Twelve lead ECG Electrode Placement [12]

Studies show that during the analysis of the ECG signals, either consideration or ignorance of the leads V1 to V5 does not impact the results in either ways. The main cause for no significance deviation of the result is because of the V leads not carrying enough information of the electrical activity of the heart. [15]

For a normal heartbeat the amplitudes and durations between the intervals in waves are shown in the following tables.

Table 2-1Amplitude of different waves in ECG signal

P wave	0.25 mV
R wave	1.60 mV
Q wave	0.25 R wave
T wave	0.1 to 0.5 mV

Table 2-2 Duration of intervals in wave and between waves in ECG signal

P-R interval	0.12 to 0.20 sec
Q-T interval	0.35 to 0.44 sec
S-T segment	0.05 to 0.15 sec
P wave interval	0.11 sec
QRS interval	0.09 sec

Analyzing these waveforms one can easily detect arrhythmia. The Advancement of Medical Instrumentation (AAMI) categorizes ECG heartbeats into five super classes namely: Normal (*N*), Supraventricular (*S*) ectopic, Ventricular (*V*) ectopic, Fusion (*F*) and Unknown (*Q*). Figure 2-3 shows the corresponding heartbeat types with the ECG signal. As shown in the figure N-type and S- type heartbeats exhibit similar pattern and the V-type heartbeat exhibits different pattern than other heartbeats. In diagnosis of the S-type of heartbeats RR-interval i.e. the interval from one R peak to another R peak is used to differentiate S-type as *S*-type heartbeats are premature and they have shorter RR-intervals than the N-types.

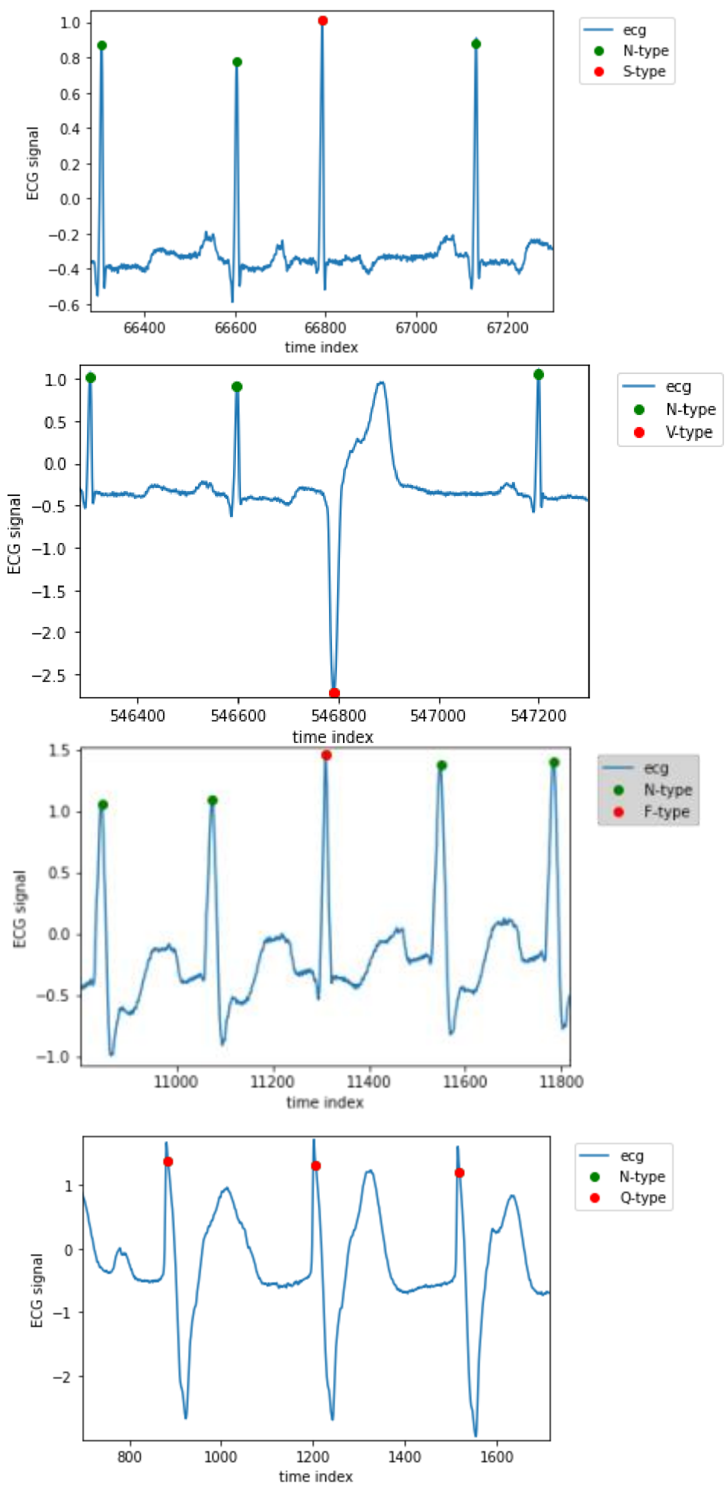


Figure 2-3 Different types of heartbeats

The dots on the graph are pointed at the R wave of the ECG. The green dot on the ECG signal represents the normal heartbeat. The red dot shows some abnormality in the heartbeat. As seen in the Figure 2-3 the V-type heartbeat shows different morphological character than other types of heartbeats and S and N have similar types of the patterns. This call for the problem while trying to differentiate between two of them.

Classification of Heartbeats

ECG data are examined by observing the normal and irregular cardiac activities. The ECG data comprises of different types of noises. These noises only deteriorate the quality of signal and causes obstacles when extracting the information from these signal. After removal of these noises the heartbeats can be classified. The supervised as well as unsupervised learning mechanisms are in use for categorization of the arrhythmia. Machine learning has extensively been used in the arrhythmia detection as the machine learning has a higher desirable performance over the neural network. Machine learning is known to have a greater performance while working with the noisy as well as missing data. Machine learning doesn't require large tests for the model optimization process. It can have a good accuracy on the smaller test. The process of feature engineering makes the machine learning to have a better performance. Features are extracted from the data so that machine can do accurate classification.

Deep learning, a subset of Machine Learning helps machines to learn about the data without the process of feature engineering. It uses examples to classify & detect. The machine can learn the features of the data without human intervention in deep learning at the cost of the training data. The more examples for the training process in deep learning the more accurate result.

2.3 Neural network

A neural network is a subset of machine learning. It is the network that reflects the behavior of the human brain and allows machines to recognize patterns in input data and classify them into their respective output values. Pattern here refers to the underlying relation among the input data. There are three types of layers in a neural network. Input layer is the first layer of the neural network and is responsible for taking the inputs and passing inputs to the second layer, hidden layers. Hidden layers put weights in the input with the help of activation function. The hidden layers processed input to their output values. After that the output layer, the third layer in the neural network outputs the result.

2.3.1 Convolution neural networks

One of the most used neural networks is the convolution neural networks. CNNS are primarily used in image recognition, scene detection, clustering of images based on the learnable weights and biases. CNN consists of neurons that have some weights and biases. Neurons in CNNs receive inputs, take a weighted sum over them and passed these inputs to an activation function, finally producing an output. The main components in CNN are convolution layer, pooling/sub-sampling layers, activation layers and fully connected layer. The advantage with CNN is that it requires less preprocessing in comparisons to other techniques. It is capable to extract the hidden information of data in its own.

Convolution layer

This layer has a set of independent filters. Every filter in CNN is convoluted independently input image to produce corresponding feature maps. In the first layer of CNN, extraction of low-level features like edges and corners is carried while the higher-level layers extract the higher-level features. There can be many convolution kernels in the layers. The features are directly related to the kernel. Convolution of one input with one kernel produces one output feature.

Pooling layers

Pooling layer in CNN is used for reduction of resolution of features. The features extracted are not much affected by noise and distortion. A pixel value in the input image tends to have a similar value to its neighboring pixels which is output having redundant information. A pooling layer extracts feature value from a group of cells repeatedly. There are two ways of pooling. In both the types, the given input is divided into non-overlapping dimensional spaces. The first one is average pooling where average of given value is calculated and second one is the maximum pooling where the maximum value of the given input is calculated

Activation layer

As the signal layer flows from one layer to another layer, Output signals are strongly connected with past references activating more neurons that enable signals to travel in more efficient way for identification. There are many functions for activation. Rectified Linear Unit (ReLU) is known to have a fast training speed in comparison with other.

The activation function is an essential part of a neural network and is a non-linear function. This function determines whether and how much to fire up the neuron output with respect to the input. Rectified Linear Unit (ReLU) is used as the activation function and is given by

$y = \max(0, x)$ Equation 2.1

The ReLU reduces the model's training time. The linearity of the ReLU makes it a fast converging algorithm because the slope remains the same when x increases. The ReLU has a zero slope when neurons have negative values, neurons are stuck on the negative side and ReLu always outputs zero. Eventually, this property leads to many useless neurons in a neuron network, which lowers model classification accuracy. For the faster converging ReLU is used as activation function in this work.

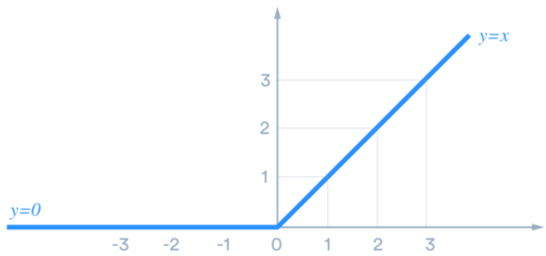


Figure 2-4 The graph of ReLU

Fully Connected

The last layers in CNN are fully connected. The neurons present in the preceding layers are connected to every neuron present in the subsequent layers. The entire possible path from input to output are considered.

2.3.2 Techniques Specifically for CNNs

There are various ways through which the CNN model can be used. The following techniques will be employed in this thesis for better performance.

Batch normalization

In a deep CNN with many hidden layers, the hidden layer parameters are dependent on its previous layer. Therefore, even a small change in the last layer's parameters can strongly influence the next layer's input distribution. It slows down the model training speed. Since there is a considerable amount of heartbeat images with limited computational resources, batch normalization is adopted for improving the training speed of proposed model. Batch normalization reduces training time. It standardizes layer inputs. It does normalization on the output of a previous layer by calculating the batch mean and variance, and this will shift and scale the output of the layer. The batch normalization can be applied before or after the activation function.

Shortcut

The depth of CNN has a lot of significance in the model's performance. Although in many times the higher number of depths leads to performance degradation. On the other hand, working with deeper CNNs leads the model to have difficulty in training. Residual networks came as a savior for this problem. There is a shortcut path to transfer the data from the former layer to the current activation function. This ensures that the model is well efficient at least as in the earlier stage.

2.4 Recurrent Neural Network (RNN)

Recurrent neural networks process the information based on the previously gained understanding. This behavior of RNN is not found in the traditional neural network. The sequence contains some information in its pattern also i.e., the way the sequence is made, there lies some pattern. These patterns hold a good amount of information. RNN can decode the underlying information of the sequence which cannot be performed by the feed-forward networks.

2.4.1 Long Short Term Memory Networks (LSTM)

LSTMs are variant of Recurrent Neural Network(RNN). RNN are capable of learning from the previous data but cannot store the information. This problem with RNN i.e. the vanishing gradient problem is solved in LSTM. It uses an additional cell known as forget cell that can capture long-term dependencies on a sequence of data. These cells are memory blocks of the network. RNN can capture only the recent information. LSTM uses the gates namely input gate, forget gate and output gate. In Figure 2-11, let W_n and U_n be the weights of inputs and recurrent connections respectively, and b_n be the bias. Let f be forget gate, input gate be i, output gate be o and cell be c.

Given for input x_t, the LSTM unit at time t is updated as follows:

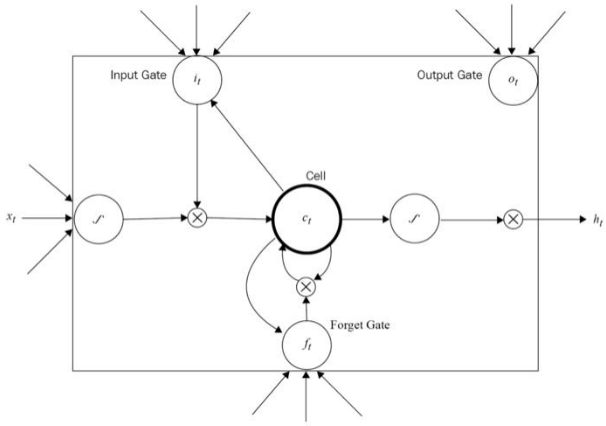


Figure 2-4 Architecture of LSTM with its memory [16]

$f_t = \sigma (W_f x_t + U_f h_{t-1} + b_f)$ Equation 2.2

$i_t = \sigma (W_i x_t + U_i h_{t-1} + b_i)$ Equation 2.3

$o_t = \sigma(W_0 x_t + U_0 h_{t-1} + b_o)$ Equation 2.4

$c_t = f_t \circ c_{t-1} + i_t \circ \tanh (W_c x_t + U_c h_{t-1} + b_c)$ Equation 2.5

$ht = ot \circ \tanh (ct)$ Equation 2.6

where σ is used to signify the sigmoid function and the operator ◦ represents the element-wise product. Applying LSTM in a heartbeat sequence helps to capture mutual-relationships of different waves and peaks in the heartbeat. Using LSTM one can be used to explore temporal dependencies between heartbeats.

2.4.2 Bidirectional Long Short Term Memory (BiLSTM)

Bi4STM is extended version of LSTM that uses two independent LSTMs to have a better performance on sequence or temporal data. It involves duplication of the LSTM networks. The first LSTM is trained on the input sequence normally while the second LSTM is trained in reverse order. At every time stamp, outputs from both LSTMs are aggregated allowing the BiLSTM network to have both past and future information about any sequence at a time stamp.

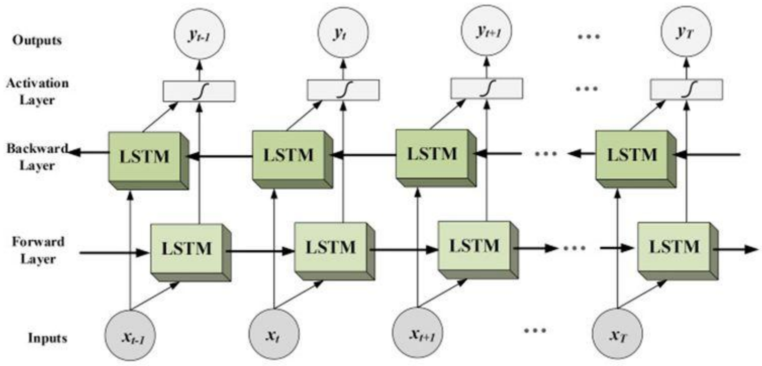


Figure 2-5 Architecture of BiLSTM [16]

2.5 Cost function

Softmax can be defined as a type of logistic regression which normalizes input data into an array of probabilities. The sum of these output probabilities is always one. Softmax activation function is used in multi-class classification problems. Softmax provides the list of probabilities for each data belonging to each class label. The data is then classified to the class-label having the highest probability.

One of the variant of sigmoid function is the softmax function. Its nature is non-linear and is mostly used multi-class classification problem. Softmax function is used as the activation function is the output layer. Softmax function maps the inputs into the probabilities of outputs belonging to each class. The class having the maximum probability for a given data is given the incoming data.

$$\sigma(zj) = \frac{e^{zj}}{\sum_{j=1}^{n} e^{zj}} \text{ for } j = 1,2,3. . ,n \qquad \text{Equation 2.7}$$

The cost function measures the performance of the model prediction accuracy. There are many types of cost functions, and focal loss is chosen as the model's cost function.

Focal loss is the variant of cross-entropy loss(CE) to address the class imbalance problem.

$CE(\hat{y}) = - \log(\hat{y})$ Equation 2.8

The performance of a classifying model can be evaluated based on cross-entropy loss. When trying to classify, model outputs a prediction probability between zero and one. The loss increases with the probability diverging from the actual label when trying to predict the output.

Focal loss assigns more weights to minor class data and less weights to major class data. The scaling factor moves towards zero when there is increase in the classification. The focal loss reduces the impact of the values if the classification began to execute accurately. The loss function can be described as the distance between the guess/prediction to the actual value of data point. The focal loss [17] is dynamically scaled to cross-entropy and can be calculated by following equation:

$FL(\hat{y}) = - (1- \hat{y})^{\gamma}.\log(\hat{y}), \gamma \geq 0$ Equation 2.9

where $\hat{y}$ represents the modulating factor and γ represents the focusing parameter. Modulating factor are weights that helps to deal with the class imbalance problem. It causes to decrease the impact of dominating or common classes of ECG beats. The modulating factor controls the shape of the curve. The normal class or the dominating class will have a high value of the modulating parameter which makes the fall in the loss for easily classified examples making the model to be focused on data that are difficult to classify at the time of training [18]. Whenever there is miss in the classification γ becomes small and the value of the modulation factor is close to 1 which in turns leaves the loss barely affected.

2.6 Adam optimizer

Adam optimizer is the optimization technique that replaces the classical stochastic gradient descent procedure. The optimization follows iterative method to update new weights while training. The process is simple to follow and is computationally efficient. It consumes less memory and also has a good performance on non-stationary objectives.

"Stochastic gradient descent maintains a single learning rate (α) for all weight updates and the learning rate does not change during training. While in Adam, a learning rate is maintained for each network weight and is separately adapted as learning unfolds. The method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradient." [19].

It combines two methods:

- Adaptive Gradient algorithm
- Root mean square propagation

Adam [19] configuration parameter:

α = this is the learning rate and an important hyperparameter while training the neural network. It is the parameter which tells the network how the weights should be updated. A large value of learning rate results in faster initial learning and smaller values results in slow learning rate during training.

β_1 = exponential decay rate for first moment estimates.

β_2 = exponential decay rate for second moment estimates.

epsilon: It is tiny amount for avoidance of division of zero. An alteration to this is the decay of learning rate using with the Adam optimizer as:

$$\alpha_2 = \frac{\alpha_1}{\sqrt{t}} \qquad \text{Equation 2.10}$$

At each epoch, the logistic regression parameters' value are as follows:

$\alpha = 0.001, \beta 1 = 0.9, \beta 2 = 0.999, \text{epsilon} = 1e\text{-}8$" [19] Equation 2.11

CHAPTER 3 LITERATURE REVIEW

Machine learning has been extensively used in the field of the medical domain. The ECG classification has also been the main attraction for machine learning and the deep neural networks in the past years. There is difference in various types of the heartbeats and thus the heartbeat classification problems are mostly based on pattern recognition paradigm. From the raw ECG signals, important features were extracted including RR-intervals, coefficients of wavelets and amplitudes of different waves present in the ECG signal [20]. Here morphological features refer to the shape and amplitude of the signal produced by device. It causes direct impacts on the final classification performance. In raw ECG signals, a heartbeat is denoted as a high-dimensional time-series sequence, which is difficult for a classifier to interpret and discover important information to distinguish different heartbeat types.

Heartbeat was represented by 30 feature- matrix and morphological descriptor of the heartbeat was represented and SVM classifier was used to classify the heartbeat [4]. However, pattern-based classification models have trouble on abnormal heartbeat detection as S-type heartbeat has similar type of pattern as that of N-type. These model accuracy is highly influenced by the features extracted, the relationship between these features and effectiveness of the classifiers. [21]. The feature engineering based arrhythmia was mostly based on RR-intervals, higher-order statistics, wavelet, signal energy coefficients etc.

In [5] features like inter beat, intra –beat intervals, amplitude morphology, area morphology and morphological distance were taken in account for the classification. The features were well taken care of giving the performance of the model a good accuracy. SVM was implied for the classification process. The feature engineering is the main core for the classification of the heartbeat in the machine learning domain.

Extensive work has now been done in heart beat classification by deep neural networks (DNNs). A DNN is a computational model consisting of multiple processing layers, which can automatically learn the high-level representations of the raw ECG recordings without extensive data preprocessing. The classification of S-type heartbeat still is a problem in DNN based models. The use of the synthetic heartbeats has now been used for training the model and evaluating the performance of the model. Arrhythmias classification model was proposed based on Stacked Bidirectional LSTM and two deep CNN. [6] This model did not require complex feature extraction and selection process and has comparatively good accuracy than the traditional ones.

One dimensional CNN was used for classification of ECG signals [7]. The method performed good on disease detection but the model turned out to be patient-specific. The unseen patient’s data could not be applied. This model could not be used in real world scenarios. Deep CNN neural networks classification model was used to classify segmented heartbeat which then produces the probabilities for each heartbeat in certain class [2]. These models somehow gave over-optimistic results as data was not partitioned properly. The same data was used in the training as well as the testing phase which leads the model to learn about the patient-specific patterns. The S-type heartbeats don’t differ much from the N-type heartbeat so the information extracted by the model did not have sufficient information for classification. A high misclassification rate was still the problem.

5-layer CNN with residual connections was used for the classification proposed. This method used a batch-weighted loss function for imbalanced data. This model included the rhythmic information by neighbor heartbeats. [8].The problem of data leakage was seen to be solved. However, this method gave arose to the problem on classification of normal heartbeat. The normal classes were treated as the diseased classes. This caused the classification model to have a higher recall value. This was not threatening but somehow unrealistic as it gives the patients unwanted burden of expensive additional test.

In [9], a Deep Learning technique was implemented with heartbeat segmentation and was able to classify arrhythmia from the ECG signals. The most important characteristic of this framework was the arrangement of the crucial elements like QRS complexes, P wave and T wave. The model was able to give a better output in the intra-patient heartbeat classification. Different temporal features were extracted from the raw signals and were able to do the classification. The neighboring vectors of the ECG signal were also the part of consideration in the classification task. The vectors in this work represented the different amplitude of the heartbeat. The cardiac cycle was depicted through the vectors.

Further, in [22] feature extraction as well as classification of the heartbeats was performed in single CNN. After the CNN model was trained with the data of a person, the model was prepared enough to define person's long ECG record. Minimal preprocessing was only needed for the classification purpose.

In [23] Continuous Wavelet Transform (CWT) and CNN were used to classify the ECG signals. From CWT, the time-frequency components of the ECG signals were obtained and CNN used these time-frequency components to extract features for the classification purpose. The R-interval features were also considered for the classifications propose.

CHAPTER 4 METHODOLOGY

The block diagram of the research methodology is as follows:

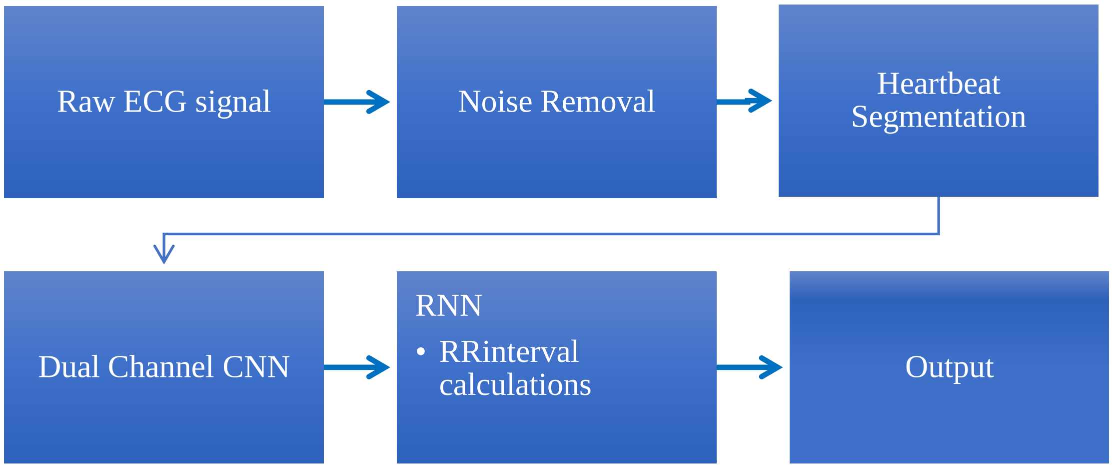


Figure 4-1 Block diagram of the research methodology

4.1 Data Collection

The dataset used in this work was MIT-BIH Arrhythmia database [24]. It is an open source data and is hosted by Physionet. The MIT-BIH database consists of forty-eight hours two-leads ambulatory ECG recordings. These recordings are from forty-seven patients where twenty-two are female patients and the rest are male patients. The recordings are approximately of thirty minutes' duration.

There are 16 classes of the arrhythmia in MIT BIH data which can be further merged into five super classes by the AAMI standard. "A patient's heartbeats are categorized into five classes as mentioned by the Advancement of Medical Instrumentation (AAMI) standard (IEC 60601-2-47:2012): normal (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beat (F), and unknown beat (Q)]. The standard also emphasizes that S and V are the two most critical arrhythmia categories. The classification of different types of heartbeats in MIT-BIH data along with the number of the samples in each class is shown as in the Table 4-1.

Table 4-1 AAMI standard grouping of heartbeats with the number of samples in MIT-BIH database

MIT-BIH class	**MIT-BIH number**	**AAMI groups**	**Number of samples**
Normal beat	1	N : Normal Beats	90548
Left bundle branch block beat	3		
Right bundle branch block beat	2		
Atrial escape beats	34		
Nodal (junctional) escape beat	11		
Atrial premature beats	8	S: Supraventricular ectopic beats	2779
Aberrated atrial premature beats	4		
Nodal (junctional) premature beats	7		
Supraventricular premature beats	9		
Premature ventricular contraction	5	V: ventricular ectopic beats	7234
Ventricular escape beat	10		
Fusion of ventricular and normal beat	6	F: fusion beats	803
Paced beat	12	Q: paced beats or unclassified beats	8012
Fusion of paced and normal beat	38		
Unclassified beat	13		33

MIT-BIH data are imbalanced data set. The normal dataset has high number of datasets and F-type heartbeats have lower number of datasets as compared to the other types of the heartbeats. Figure 4-2 shows the ratio of different heartbeats in the MIT-BIH datasets.

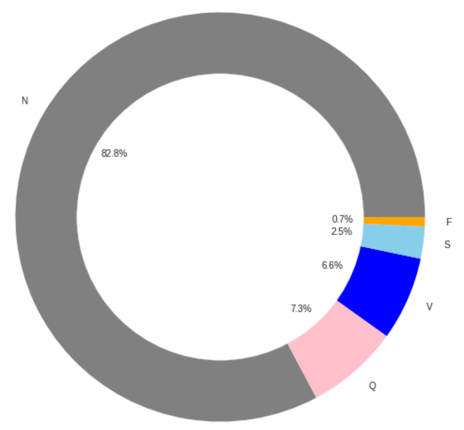


Figure 4-2 Distribution of the data in different classes of heartbeats

4.2 Preprocessing of Data

4.2.1 Noise Removal

The raw ECG signals contain some noise. These noises are baseline wander, power line noise, muscle artifacts, electrode motion etc. The baseline wandering, power line noise and white Gaussian noise are the main noise present in the ECG signal. This noise has negative impact on the accuracy of the classification model. To remove baseline wandering median filter were used while the high frequency noise are eliminated using the low pass filter. In this work, Butterworth loss pass filter was used as it is known to have the maximally flat pass band and contains no ripples. [25] Baseline wandering is the noise where the signal doesn't have the same axis. The power line interference is result of the electromagnetic field by the appliances nearby.

Baseline wanders is the noise that causes the heartbeats to move up or down from the base axis instead of aligning in the same axis. In the above figure the first diagram shows that the individual heartbeat is not in the same axis. There are many reasons behind the baseline wander like movement of the cables and the patents at the time of recording, loose electrodes. The frequency of the baseline wander is usually below 0.5 Hz. [11]

Baseline wandering was removed by median filter. Two median filters were used. First the signal was processed by 200ms median filter. Then the resultant from the first filter was again processed by the 600ms median filter. The final output was subtracted from the original signal in order to receive the corrected signal.

After that, Butterworth filter of order 3 and cutoff frequency 0.1 was deployed. This process was done for the removal of the background noise from the baseline corrected signals. The response of the Butterworth filter is expected to have zero roll off response in the stop–band and a flat frequency response in the pass-band. This is the main reason for choice of the Butterworth filter in removing the background noise.

4.2.2 Heartbeat Segmentation

For segmentation, every recording was segmented from R peak locations based on the annotations. The heartbeats were segmented on the basis of R peak. 90 samples were taken before each R peak and 110 samples were taken after each R peak. The whole was represented as one heartbeat segment. The samples were taken on the basis of the sampling frequency. The segmentation was enough long for the representation of the repolarization of ventricular while avoiding the neighbor heartbeats.

4.3 Dual Channel Convolutional Neural Network for feature extraction

The work was based on convolution neural networks for feature extraction process. The network accepted the segmented ECG as input and gave the extracted feature as output. The convolution neural networks capture both frequency pattern and temporal pattern segmented heartbeats.

CNN has been used more nowadays as it can effectively capture key pattern and learn features from these patterns. The key patterns here are P-waves and QRS- complexes of ECG signal. Convolutional layer is made up of an input layer, multiple convolution layers, pooling layers, and dense Layers. The first layer in CNN is the convolutional layer where most of the computation is done. This is the important layer in the convolution neural network. Given input data, a convolution is known as a linear combination of each point with its neighbors. It is usually followed by a ReLU activation to enable the network to learn non-linear patterns from the input data. The pooling layer is used to reduce the size of the learned representations to reduce the number of network parameters. The purpose of the dense layer is to provide an overall regulation of the previously learned representations.

The residual network (ResNet) was first introduced in [26]to solve the network degradation problem: with a network getting deeper, its accuracy gets saturated and then degrades rapidly. Figure 4-3 shows a sample ResNet block, which consists of two stacked layers and an identity mapping (the shortcut connection). The shortcut layer present in the residual network ensures that the model's performance will not get degraded as the network goes deeper even if it can't perform better than without layer. This network applies identity mapping which is the input to some layer is passed directly to the some other layer as a shortcut. This shortcut is known as skip connections. As in the figure below there is skip connection of x which is both present as input to the first layer and the merging layer.

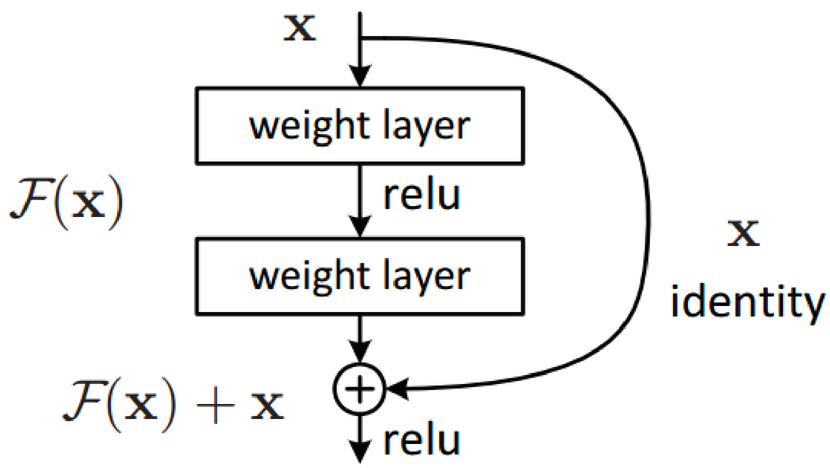


Figure 4-3 Residual learning: a building block [26]

The input is denoted by x and the desired mapping by H(x).Residual mapping can be expressed as:

$F(x) := H(x) – x$ Equation 4.1

Therefore, the desired underlying mapping H (x) can be recast into:

$H(x) := F(x) + x$ and represented by the stacked layers plus with an identity mapping

Representation Learning

This architecture is inspired from the paper [27]. Both the frequency and temporal features for extracting feature in signal processing was taken. Two CNNs were employed with small and large filter sizes to capture temporal pattern and frequency pattern in segmented ECG signals. These features are time-invariant. Information from these two CNNs will be added together before pooling operation.

4.4 Recurrent Neural Network for Classification

RNN is a class of artificial neural networks where the outputs from the previous steps are fed into the current step as the input. RR intervals are also added to this layer as input. Normal heartbeats do not only have morphological differences with the arrhythmias but also have difference in the neighboring RR intervals. [23] The RR intervals are the intervals from the present R peak to either past R peak or future R peak. Pre RR, post RR, ratio RR and local RR are the most common RR intervals that have been in use for the process of arrhythmia detection. Pre RR interval here means the interval time difference obtained in moving to the current R peak from the previous R peak. The post RR interval is the interval difference in between the present and following R peak. Local heartbeat is calculated from the average of ten previous RR while the ratio RR is the ratio of pre RR and post RR [5]. Robust scalar was used for normalization of the RR intervals. RR intervals were mainly used to classify the S-type heartbeat from the N-type heartbeats. These features were calculated as:

rris = np.diff(r_peaks) Equation 4.2

pre_rr = rris[index - 1] Equation 4.3

post_rr = rris[index] Equation 4.4

ratio_rr = pre_rri / post_rri Equation 4.5

local_rr = np.mean(rris[np.maximum(index - 10, 0):index]) Equation 4.6

Here, rris refers to the difference between the two consecutive peaks in the QRS complexes.

The BiLSTM here is responsible for the classification of the different types of heartbeats. It used the softmax as the activation function at its outputs which provides the probability for each data to belong in different classes. Class having the maximum value of the probability value will be assigned to the data. Bidirectional LSTM (BiLSTM) can propagate signal backward as well as forward in time making it to learn the temporal pattern underlying in the data. Unlike in LSTM, BiLSTM process both the past and future information of sequence. The N-type and S-type heartbeats are similar in shape and possess similar kind of morphological features. The only difference between them is S-type are premature so they do not have same time pattern as that of the S-type. BiLSTM learns and extracts hidden temporal dependency between heartbeats by processing the input RR-interval sequence in both the directions. The *softmax* function at output layer classifies the different types of heartbeats.

4.5 Focal Loss for handling the imbalanced dataset

The model used the focal loss as its loss function to handle the class imbalance problem. Focal loss is used to empower the weights of the minority class. More weight is given to the minor class so that there will be easier classification of the minor class as compared to that of the major class.

The following flow chart shows the working of the focal loss giving the idea how the weight is updated in the process of training.

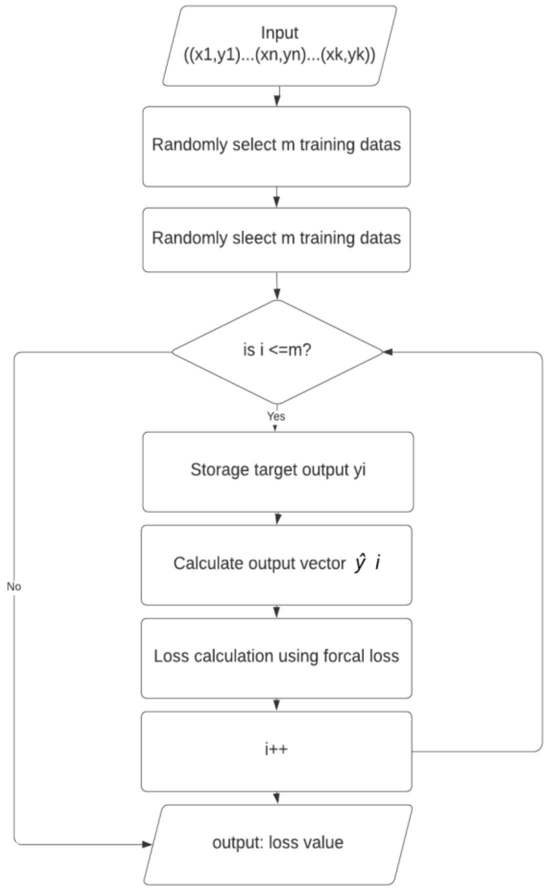


Figure 4-4 Block diagram to calculate loss value using focal loss

Whenever there is misclassification and $\hat{y}$ is small, the modulating factor has value about 1 which causes the loss to be robust from affection.

4.6 Evaluation matrices

Performance of the classification of the two-stage neural networks can be the indicator of the system. Accuracy, precision, recall and f1 score were used to evaluate the performance of the classification framework.

"Accuracy can be defined as the ratio of correctly classified correctly classified data to the total data.

Accuracy = (True Positives + True Negatives)/(True Positives +False Positives + False Negatives +True Negatives)

However alone accuracy is somehow biased in the context of imbalanced dataset.

Precision is the proportion of the correctly classified positive data to all the positively classified data present in a data set.

Precision = True Positives/(True Positives + False Positives)

Recall can be defined as the ration of correctly classified positive labeled data to all the positive data.

Recall = True Positives/(True Positives + False Negatives)

F1-score is the harmonic mean of precision and recall and considers both. F1 Score is best if there is some sort of balance between precision (p) & recall (r) in the system. Oppositely F1 Score isn't so high if one measure is improved at the expense of the other.

F1 Score = 2(Recall * Precision) / (Recall + Precision)"*

4.7 Receiver Operating Characteristic Curve

The receiver operating characteristics (ROC) curve is a performance measurement for classification problem. The ROC curve has two parameters the true positives and the false positives. An ideal ROC curve is shown by the classifier A in Figure 4-7 having true positive rate near 1 and false positive rate to zero. The ROC curve is generated by plotting the true positive rates against the false positive rates. There is another important parameter in the ROC curve that is AUC. AUC stands for the area under curve. This represents the measure of separability.

In Figure 4-5 there are three classifier A,B and C. Classifier A is the ideal classifier that has the sensitivity equal to 1 and FPR equal to zero.

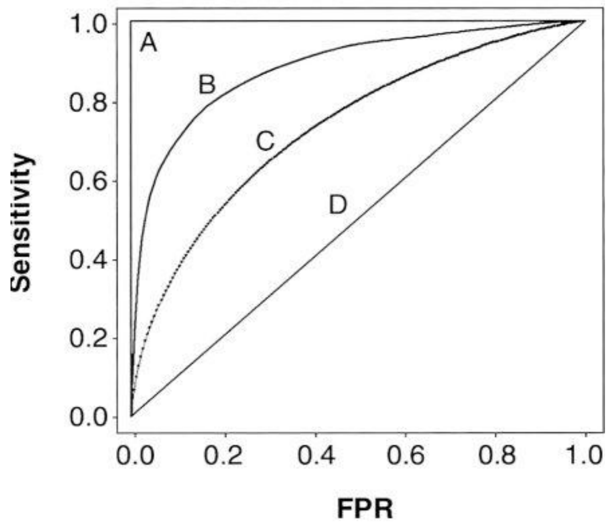


Figure 4-5 Receiver operating Characteristic Curve

In order to find the better among the classifier B and classifier C one has to find the AUC of the classifier B and C. Though ROC gives a visual representation of the classifier ability to perform, one can be sure from the AUC curve about the precision of the classifier.

CHAPTER 5 EXPERIMENTAL SETUP, RESULT AND ANALYSIS

5.1 Experimental Setup

5.1.1 Hardware Requirements

Deep learning requires high computational power. Deep learning technique comprises of multiple calculation. Therefore the experiment of this thesis was performed in the Google Colab. Colab is a product of Google Research which is a free platform for developing the codes requiring zero setup. One of the big advantages when working with Colab is that it provides free access to computing resources including GPUs. While working with machine learning and deep learning heavy computation's execution time can be made faster with the help of the parallel execution. Therefore, Google Colab is a good platform while working with deep neural networks. The specifications of Google Colab are:

GPU : Nvidia K80/T4

GPU Memory: 12GB

No. CPU Cores:2

Available RAM: 12GB

Disk Space:358 GB

5.1.2 Software Requirements

Coding part of this work was done in Python 3.6. Different libraries of the python were installed for the purpose of data manipulation and data analysis. Keras, Numpy, Pandas, Scikit, Wfdb are some of the libraries that were used during this execution of the program of this thesis.

5.2 Data Preprocessing

The dataset used in this work was MIT-BIH Arrhythmia database. The recordings of MIT-BIH datasets are sampled at 360 Hz with 11-bit resolution over a 10-mV range. Modified Limb II is used as the first lead for all the recordings except for 114 where V5 is used as the first lead and Modified limb II is used as the second lead. The second lead is usually V1, but depending on subjects it can be V2, V4 or V5. V leads don't have complete information of the different ECG waves, [12] first lead was used in this work.

5.2.1 Data Partitioning

The data was portioned under the inter-patient paradigm. If the same patient's data is to be used in the training phase and testing phase the paradigm is known to be inter patient paradigm. If the separate patient's data is used in testing that what was used for the training, it is known to be inter patient paradigm. Intra patient paradigm has the problem of over fitting and can't be use in the real world implications.

The partition of the data was done as accordance with the inter patient paradigm. Pandas library of the python language was used to extract the data from the source. In whole, 108909 data was extracted from the 16 classes making 5 major classes.

The records were splitted into a training set and a validation set which constitutes 70 percentages for the training while remaining 30 percentages for the validation. The following table shows the way for portioning the patient for the classification process.

Table 5-1 Training and Testing Datasets in inter patients partitions

Data Set	**Recordings (Patients ID)**
DS1	100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 111, 112, 113,114, 115, 116,118, 119, 121, 122, 123,124, 200, 201, 202, 203, 205, 207, 208, 215, 220, 223, 230
DS2	117, 209, 210, 212, 213,214, 219, 221, 222, 228, 231, 232, 233, 234

Here, DS1 is training dataset of the neural networks. DS2 is the testing dataset of the neural networks.

Different types of the heartbeats present in the training set and testing set is given by the Table 5-2. The major classes of the heartbeats were classified from the classification of AAMI. [4]

Table 5-2 Number of samples for each label in training and testing dataset

Heartbeat type	**Training Dataset**	**Testing Dataset**
N	71051	19497
S	1048	1731
V	4941	2293
F	777	26
Q	6235	1800

5.2.2 Noise Removal

Before the Network is trained, ECG signals were preprocessed for noise removal. ECG signal comprised of different types of noise. Baseline wanders and Gaussian white noise has a large impact on the quality of signal. These two noises are the main cause of deterioration of the signal.

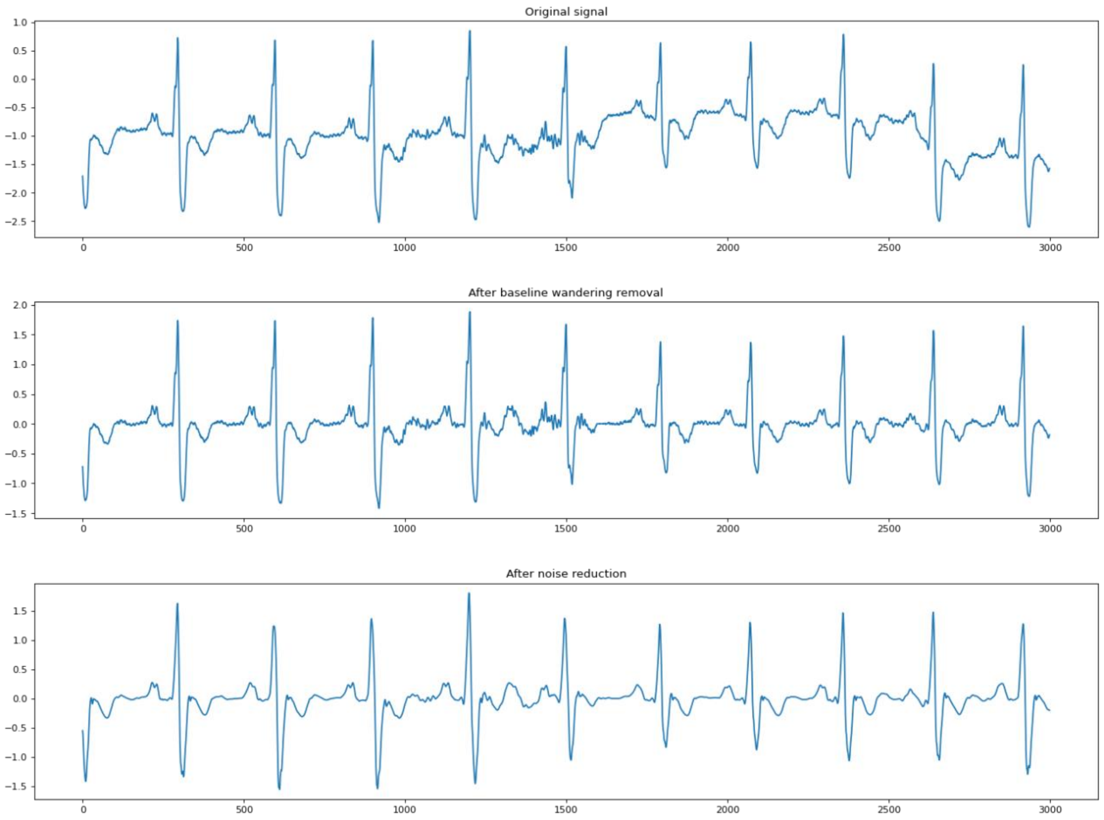


Figure 5-1 Noise Removal in the ECG Signal

5.2.3 Heartbeat segmentation

Heartbeats were segmented after noise removal. QRS complex detection was performed based on the location of the R peaks within the ECG strips. Annotation data was used to locate individual R peaks within these records.

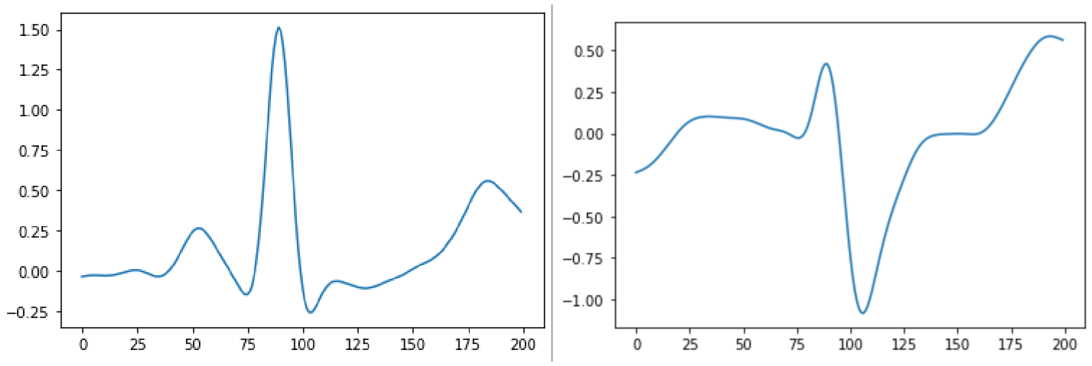


Figure 5-2 Examples of segmented heartbeat of N type and V type class respectively

As in the Figure 5-2, each segmented heartbeat only contained one R peak and it avoided the neighboring heartbeats features. One segmented heartbeat was of 200ms duration.

5.3 Training with neural networks

5.3.1 Convolutional neural network to extract features

The process of the feature extraction was done by the convolutional neural network. Dual channel CNNs were used for representation learning. These process employs two CNNs. The CNN block was able to extract the time as well as frequency related features at each epoch. CNNs used the filters that extract the key patterns of the heartbeats from the segmented heartbeats.

Figure 5-4 shows the flow diagram for the feature extraction process. As in the figure the features from the two CNNs were added before giving it as the input to the next layer.

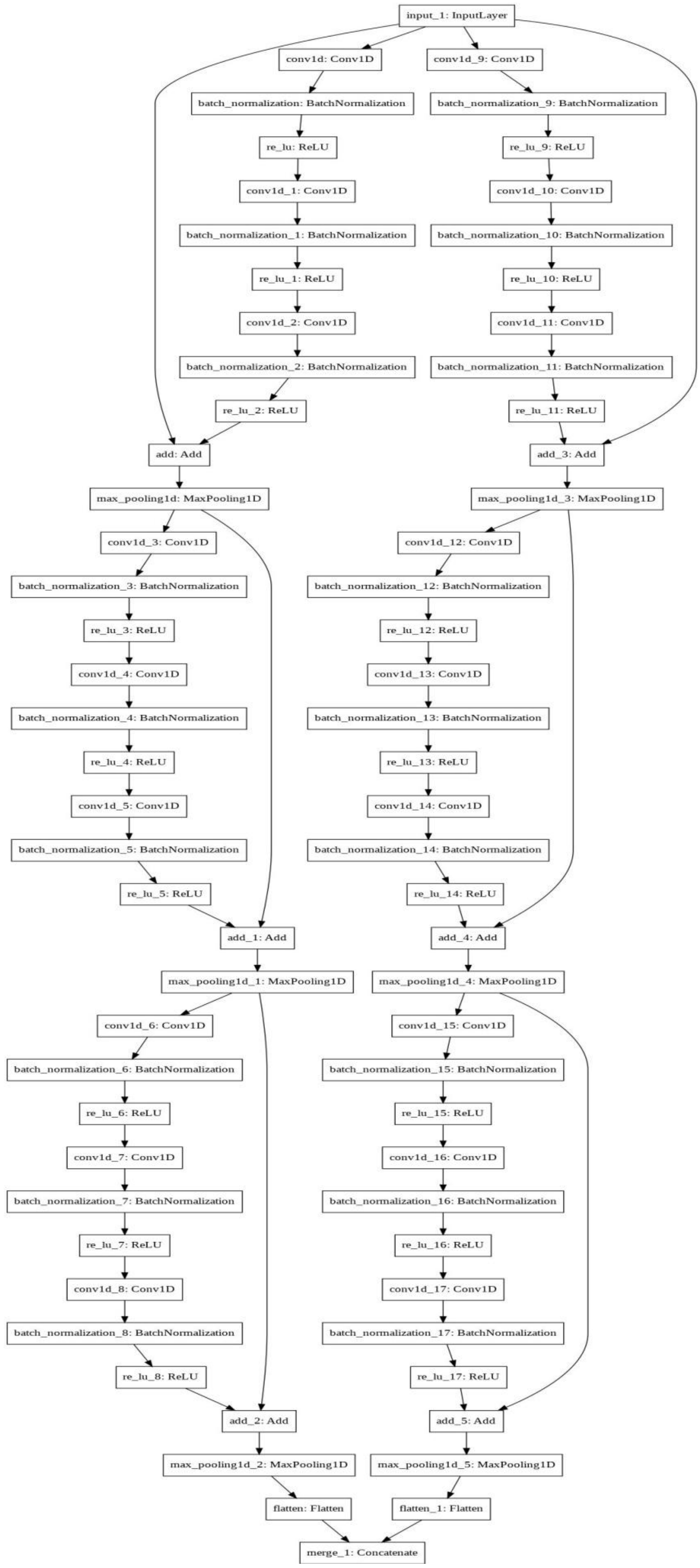


Figure 5-3 Flow model for the classification

Two CNNs had different filter size. The purpose of using two CNNs of different size is the trade-off between the temporal and frequency clarity in the process of feature extraction. CNN with smaller filter size (filter size =8) was used to capture time-domain features (i.e., the certain of ECG signal appears) while CNN with larger filter size (filter size =64) was used to capture frequency domain features (i.e. frequency components of the ECG signal). The frequency components of the ECG signals are relatable to the higher order statistical features.

The learned temporal and frequency information were added together. The entire block contained 18 convolutional layer, pooling layers, a concat layer. After each convolution, batch normalization was done. Batch normalization reduces the problem of overfitting. The issue caused by change in distribution of the data (covariate shift) in the network can also be resolved with the help of the batch normalization. The process was then followed by the ReLU activation. The ReLU activation functions allows the model to learn the non-linear patterns of the ECG signal . The concat layer concatenates the flattened output features from the two different CNNs. Pooling layer downsamples inputs using max operation.

Residual block of this network consists of three convolutional layers at each channel. These residual blocks were passed by the skip connection. This process call for the assurance that the network will not degrade as it gets deeper. Whatever the network learns from the residual block was added to the shortcut each time at the end of the every residual block. This process of the shortcut accelerated the process of training.

5.3.2 Recurrent Neural Network layer for classification

The output feature vectors from CNN layer was now transmitted to the BiLSTM layer. The important feature R-intervals was merged with the extracted features from the CNN to be passed through RNN layer as its input. The RNN was used to learn the temporal information. Bidirectional LSTM is the extended version of the traditional LSTM. The framework of the BiLSTM consists of two LSTMs which are stacked. One LSTM process input sequences in forward direction while the other LSTM process the sequence in the reverse direction. In this way, the bidirectional LSTM is capable of processing the pre-as well as post ECG temporal while LSTM processing only the pre ECG temporal information. The output of the RNN was softmax function. Softmax was implemented to compute costs in the model's training process, and tensorflow.keras was used to implement the softmax layer. The softmax function was used as the activation function in the output layer of neural network models that predict a multinomial probability distribution.

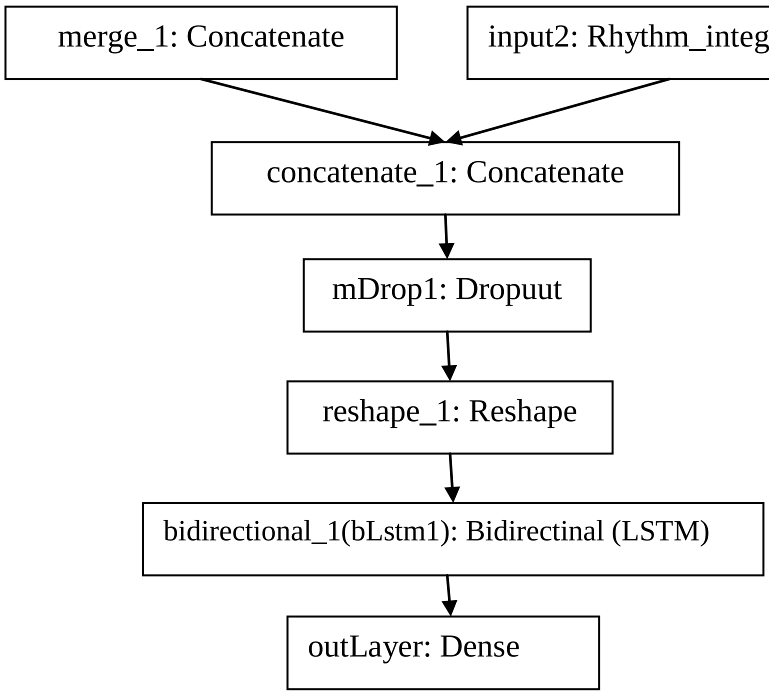


Figure 5-4 Rhythm integration and classification flow diagram

Figure 5-4 shows how bidirectional LSTM was used for the purpose of the classification. After the feature extraction, the extracted features were added with the rhythm information and were passed through the BiLSTM for the classification purpose. Reshaping was done to make the features transformed into time stamp. 64 hidden layers were used in the BiLSTM.

5.4 Network Parameter Configuration

To get the optimal result, the framework was evaluated with different hyper parameters. The parameters were updated according to the accuracy of the testing set.

- Batch size: Batch size refers to training samples that are sent to the network in one iteration on the process of training. If there is only one iteration per epoch than the data has been sent as a whole in the training.
- Epochs: The epoch can be defined as the complete pass of the training samples in the network. There are several iterations in one epoch i.e. for the whole data to be sent on the neural network it takes multiple steps. After the whole data has been passed on the network, it is said to have been trained for one epochs. It is one of the important hyper parameter in neural network. The number of the epochs is directly related to the accuracy of the network.
- Learning Rate: Learning rate is a hyperparameter in the neural network that tells the network how fast or slow to converge to local minima while training. If the learning rate is low the machine converges too fast to conclusion and if it is too high it converges too slowly to conclusion. If not optimized properly it can call for vanishing gradient and exploding gradient problem. The learning rate has a small positive value and is a configurable hyper-parameter.
- Dropout: Dropout is a regularization technique that refers to dropping out neurons during training. The dropout technique is usually done so as to resolve the overfitting problems.

Training the model:

Total labels=5

Filter size f1=8

Filter size f2 =64

Total validation samples=10% of training Samples

Epochs for training=50

Iterations = 591

Optimizer = Adam Optimizer

5.4.1 Learning rate optimization

The model is run at different learning models. Initially learning rate 0.1 is used. The graph shows a stabilized accuracy versus epochs curve. However the accuracy does not increase from the starting epoch. The model is not able to optimize its performance and unable to learn from the features it has extracted.

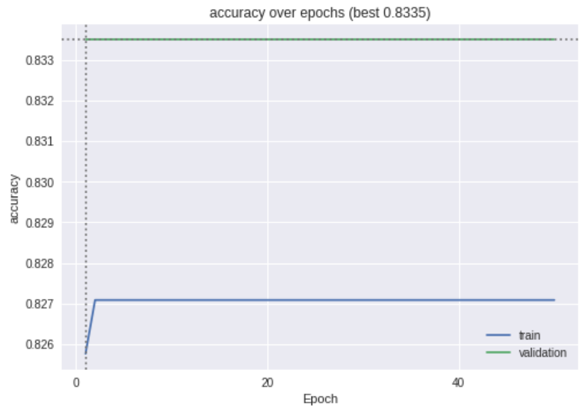


Figure 5-5 Accuracy over epochs curve at learning rate 0.1

With the learning rate 0.1 models seems to be divergent from the actual output. The classification model is left untrained.

A large learning rate (1e-6) is adopted after it giving the following result.

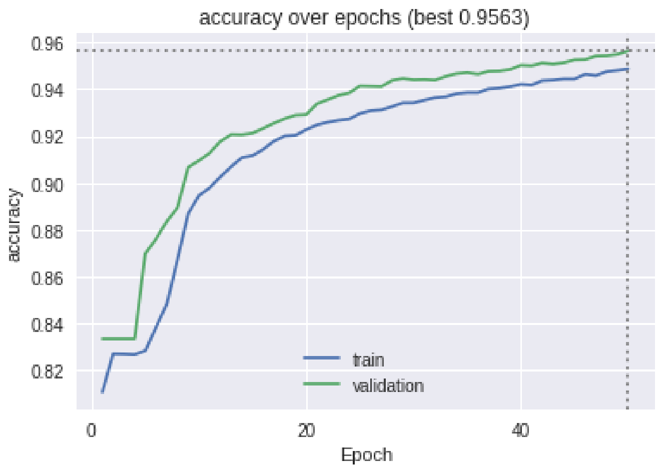


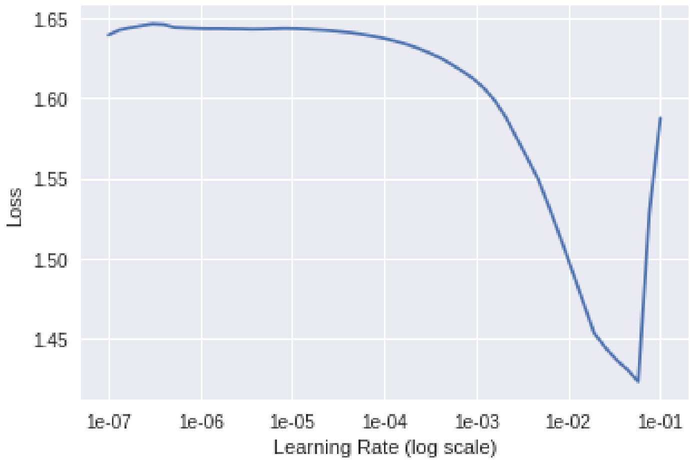
Figure 5-6 Accuracy versus epoch curve for learning rate (1e-6)

From the above graph it is clearly seen that the model is still trying to learn at 40 epochs. 50 epochs are still not enough for the model to classify the heartbeats. This learning rate needs much more updates before reaching the local minima.

Model with learning rate (1e-6) takes more epochs to converge. The solution for this is found on the cyclical learning rate. This requires a call back function to execute. In simple words, cyclical learning rate is way of setting and changing learning rates at the time of training. Learning rate is changed in a cyclical way for each batch and varies between the two bounds.

For a good learning rate range, learning rate should minimize the loss without making the loss explodes or diverges. This is done by increasing the learning rate after each batch and recording the loss. The learning rate before the loss becomes divergent is taken as an end learning rate. The starting learning rate is the learning rate after the loss just starts to decrease. This is done only for one epoch.

The plot of the framework for one epoch with the learning rate 1e-7 and 1e-1 is shown in fig.



s

Figure 5-7 Learning rate finder plot

From the above figure we can see that before 1e-05 learning rate the loss remains the same. After the learning rate is 1e-05 the loss slowly becomes to converge or decrease. Likewise at the right side after the learning rate is reached to 1e-3 the learning rate explodes. This causes the loss to be divergent from the local minima which is not good for training the model. The best learning rate is seen from the 1e-5 to 1e-3 which eliminates the vanishing as well as the exploding gradient problem. Thus the cyclical learning rate is chosen in between 1e-5 and 1e-3.

After finding a good learning rate, dropout was optimized. Different dropouts were evaluated and their respective accuracy percentage was calculated while keeping other parameters constant.

5.4.2 Drop out Optimization

Then, the different dropouts were evaluated on the framework and the following results are obtained.

Table 5-3 Performance evaluation of different dropouts

Dropout	**Accuracy of testing set(%)**
0	92.26
0.10	92.7
0.15	92.7
0.20	93.09
0.25	92.3
0.30	92.21

The performance of the models is high at drop out 0.20. Hence dropout value of 0.20 is chosen for the network.

5.4.3 Batch Size Optimization

Then the effect of the batch size on the performance of the model was evaluated. Different batch sizes on the networks are sent and their metrics were recorded. The following table shows how accuracy changes with the change in the batch size.

Table 5-4 Performance evaluation of the different batch sizes

Batch sizes	**Accuracy of the testing set (%)**
32	81.20
64	92.87
128	93.26
256	92.11
512	91.72

It is observed that the model performs at a higher accuracy when the batch size is 128. Therefore the batch size of 128 is carried out to the classification.

5.4.4 Classification without shortcut layer

The model is evaluated at first without the shortcut layer and the following graphs were observed at the output.

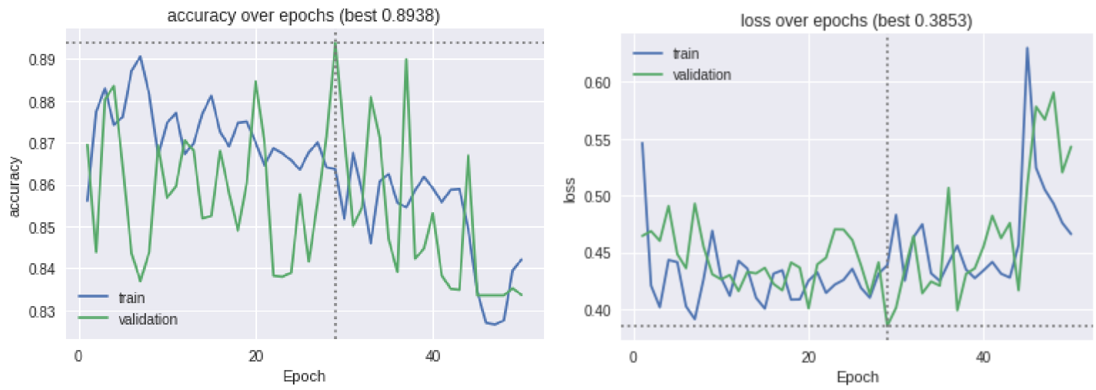


Figure 5-8 Accuracy and loss versus Epoch respectively curves without shortcut

The model itself deteriorates its parameter when passing through various layers. No metrics is seen to be increasing leading to the ambiguous results. The use of the different convolutions not only put zero meaning to the classification framework, but leads to the deterioration of the information gained at each level. Therefore a short cut layer was implied in the model. This ensures that even if the model does not perform better after every step, it will not deteriorate its performance.

5.4.5 Classification without RR interval in consideration

The progress in training was shown in the graph below:

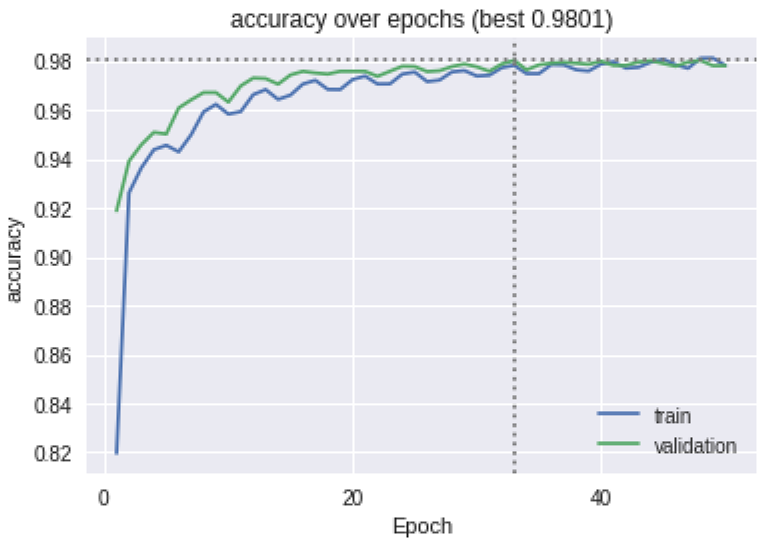


Figure 5-9 Accuracy over the epochs

The figure shows how the accuracy increased with the epochs. There is seen that after 37 epochs the accuracy does not seem to considerable increase so data training is limited to 50 epochs. This plot shows the accuracy of the model while training. The accuracy of the model does not merely define the performance of the model because there exists variations in ratio of existence of different types of samples belonging to each class. This calls for the plot of the loss versus epoch curve as well.

Loss function shows how the loss decreases with the epochs. Loss is the parameter that defines how much the predicted data is deviated from the original data. Lower loss tells higher precision of the data. The following graph shows how the loss decreases with the increase in the epochs.

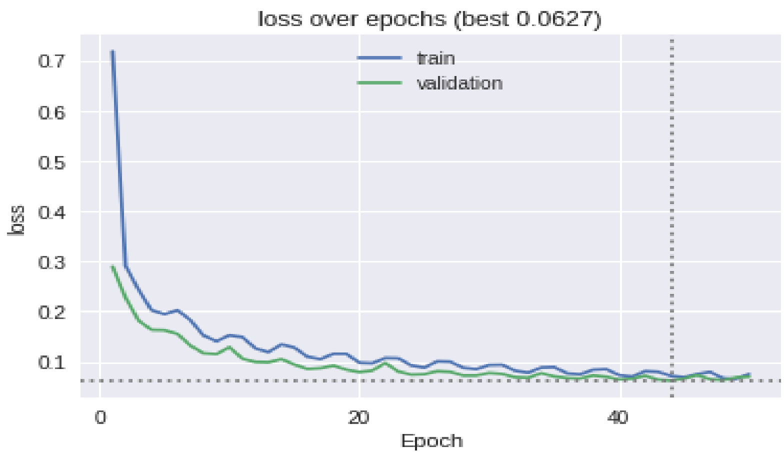


Figure 5-10 Loss function of the model

After 40 epochs, there is only a small amount of change in the loss while the loss decreases at a good rate during the first 20 epochs. The accuracy stops at the 37th epoch while the loss continues to decrease in the 40th epochs too.

At first the features extracted from the CNNs were only passed through the network for the classification. The R-R interval features were not passed through the recurrent neural network. The model gives the following confusion matrix as the output.

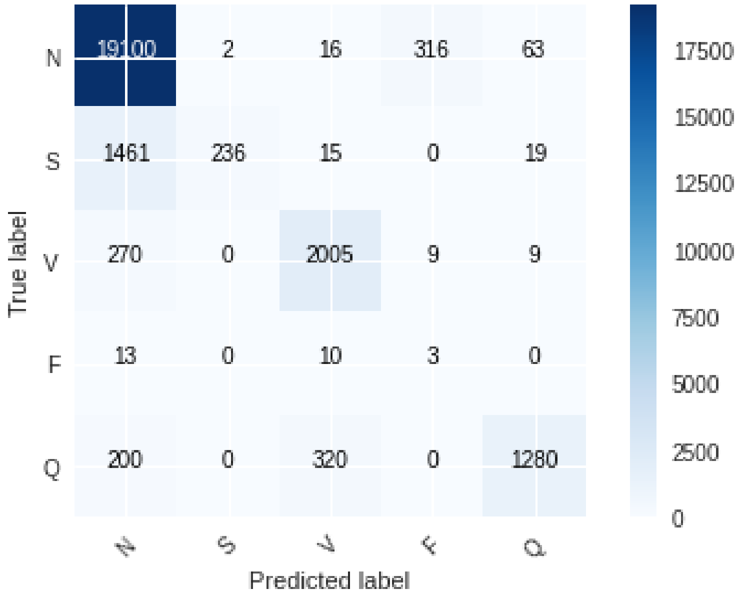


Figure 5-11 Confusion Matrix without the concatenation of RR interval.

This shows that the network was capable of classifying other types of the heartbeats with a good precision excluding the S-type heartbeat. S-type has only the time interval distance difference with the N-type heartbeat, otherwise they exhibit same pattern. This may be the reason behind the lower accuracy of the S-type heartbeat. After this the model was sent the RR interval information concatenating with the information from the CNN extracted features.

5.4.6 Classification under the Cross-Entropy loss

The RR intervals were concatenated along the CNN extracted features. The model had a significant accuracy for the S-type heartbeats along with the improvement of accuracy of other types of the heartbeats which can be seen in the figure below:

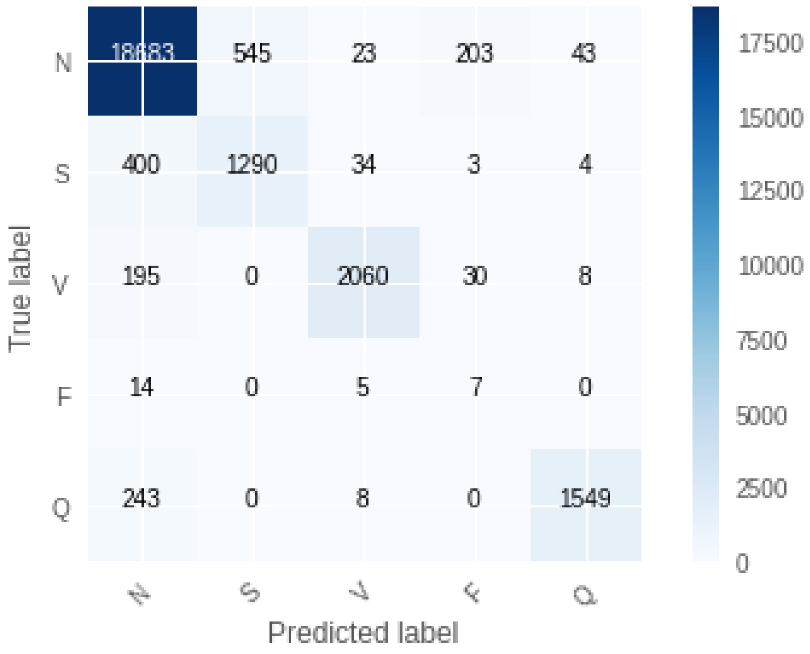


Figure 5-12 Confusion matrix with BiLSTM with RR intervals and cross entropy

5.4.7 Classification with focal loss

The model was again evaluated with the focal loss as the loss function. The focal loss is well known to solve the class imbalance problem. Imbalance in the data lets the data to have high accuracy even if the model does not perform well. It can be seen that about 83% of the data is in the normal class. If the model classify every data as the normal one the model will have the accuracy of the 83.75%, which is a good accuracy rate.

The other evaluation matrices will prove the accuracy not to be worthy in case of the class imbalance. The cross entropy loss function is the function that helps to minimize the error while updating the weights of the neurons. The problem with the cross entropy loss function is that the cross entropy loss functions are less affected by the minority class. It puts the same weights for both minority and majority class. This problem can be handled by the focal loss. With the use of the focal loss as the cost function, the following confusion matrix was obtained.

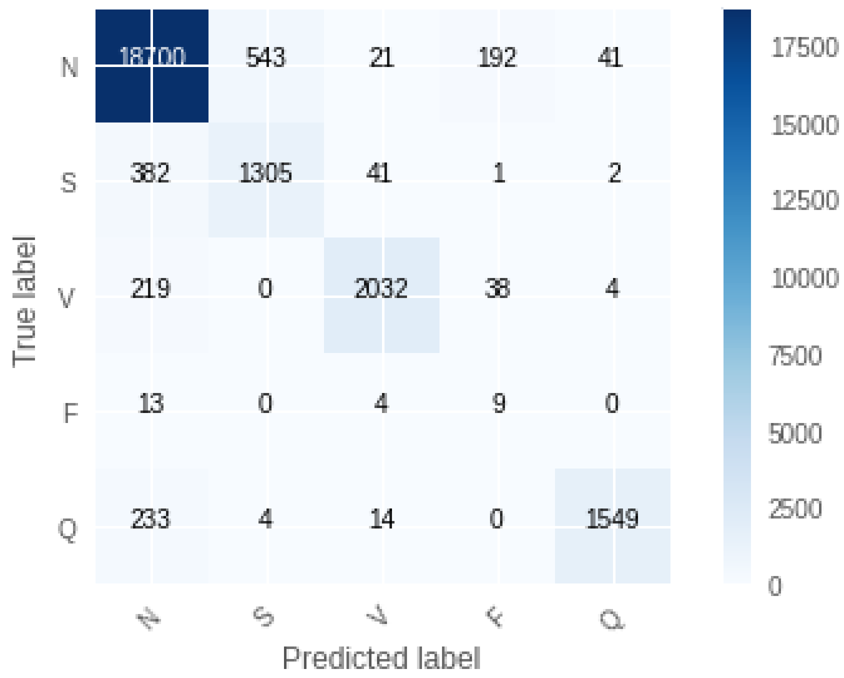


Figure 5-13 Confusion matrix with BiLSTM with RR intervals with focal loss

The overall accuracy of the system seems not much affected. However, true positive rates have been increased with the use of the focal loss in the minority class. The main objective of using the focal loss as the cost function is to highlight the minority class. Performance of cross-entropy and focal loss was compared to each other which shows result as follows:

Table 5-5 Comparison of performance of focal loss versus cross entropy loss

Methods	**Accuracy**	**Mac-roF1**	**N**		**S**		**V**		**F**		**Q**	
			P	**R**	**P**	**R**	**P**	**R**	**P**	**R**	**P**	**R**
FL	93	72	96	96	70	75	96	89	4	35	97	86
CE loss	93	71	96	96	70	75	97	90	3	27	97	86

The overall accuracy of the model remains unchanged. However there is increasement in the recall in the F-type with a good margin while using the focal loss. This result shows that the focal loss is well efficient in handling the class imbalance problem. It can classify the objects of the minority class well enough than the cross entropy as the loss function.

5.5 Analysis of the performance of the research model

The performance of the model was evaluated from accuracy, precision, recall and f1-score.The final result of the framework is shown by Table 5-6.

Table 5-6 Classification Report of Arrhythmia

Evaluation matrices Class	**precision**	**recall**	**f1-score**	**support**
N	0.956	0.959	0.958	19497
S	0.705	0.754	0.728	1731
V	0.962	0.886	0.922	2293
F	0.037	0.346	0.067	26
Q	0.970	0.861	0.91	1800
Accuracy			0.931	25347
Macro avg	0.726	0.761	0.718	25347
Weighted average	0.9399	0.930	0.935	25347

The accuracy of the model was 93%. As accuracy can't alone determine the efficiency of the model other parameters for evaluation was also taken for consideration. F-type exhibits the least accurate class on the test set. F-type heartbeats are the combination of N-type and V-type heartbeats. An F-type heartbeat contains the features from both of the N-type and V-type heartbeats. The F-type class has a lower number of the data as in comparison with the other class. Inadequacy in training data as well as the fusion from two classes of heartbeats is the reason for the least accuracy for F-class heartbeats.

Figure 5-14 shows the ROC curve for the different types of the heartbeats. As the classification of the F-type has lower precision and recall F-type has lower area under curve (auc).

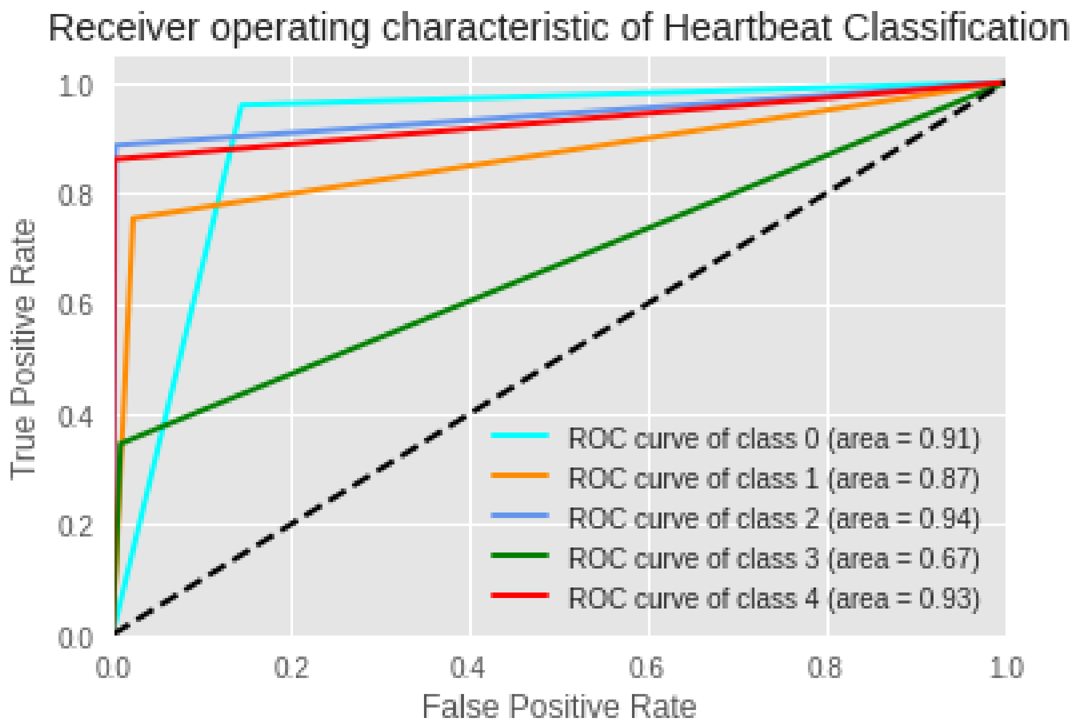


Figure 5-14 ROC Of multiclass heartbeat classification

V-type heartbeat has a good auc curve on comparison with the other types of the heartbeats. This means that the V-type heartbeat is less misclassified by the model than other types of the heartbeats. V-type heartbeat exhibits different morphological feature on comparison with the other types of the heartbeat. It is the reason for V-type having the good precision over other types of the heartbeats.

The above figure shows the receiver operating characteristic for different types of heartbeats. Class 0 refers to the N-type heartbeat, class 1 to the S-type, class 2 to the V-type heartbeat, class 3 to the F type and class 4 to the Q type heartbeat. The classification of the V-type heartbeat has most been accurate from the ROC curve.

5.6 Performance comparison of the research model with the existing works

Different deep learning algorithms have been used in the field of arrhythmia detection.

Table 5-7 Performance comparison of the method with existing works

Method	**Accuracy**	**N**			**S**			**V**		
		P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
This method	93.08	95.6	95.9	**95.8**	**70.5**	75.3	**72.8**	**96.2**	88.6	**92.2**
Acharya [2]	70.7	95.0	71.5	80.6	5.3	6.7	5.91	27.6	89.2	42.2
Zhang [5]	86.0	**98.8**	86.5	92.2	34.6	**77.9**	47.9	92.7	88.5	90.5
Shan [4]	94.7	94.5	**97.3**	**95.8**	27.6	37.4	31.8	70.2	**94.2**	80.4

Results shown in the Table 5-7 were produced in the testing set in the same environment of the research methodology.

Here the P stands for the precision and R stands for the recall. Comparison of five types of major arrhythmias with the inter-patient paradigm is not found in most of the works related. Thus, the performance of classification of the N-type, V-type and S-type has been only compared with the classification of the existing frameworks. From the Table 5-7, for the research model, the f1 score is highest in every classification. Dual Channel CNN had effectively extracted the important features of ECG signals and the use of the surrounding information of the RR intervals along with Bidirectional LSTM had contributed positively in the phase of classification. The use of the two filter sizes CNN extracted low level features and high level features at the same time. The concatenation of these features led a better learning pattern for the machine. Thus the accurate classification of the V-type was increased. LSTM are basically used to capture the sequential information. The classification of S-type is mainly dependent on the sequential difference among its neighboring heartbeats. In this way classification accuracy of S-type was increased.

CHAPTER 6 DISCUSSION AND FUTURE WORK

One of the major issues in the heart is the heart arrhythmia. In particular automated classification of humans' heartbeat is the technique where patient can get early detection and is cost effective analysis of abnormality in heart rhythm. The timely treatment of the at-risk patient is dependent on the detection of the arrhythmia which is a challenging task.

The main motive of this thesis was to build a model which can differentiate among the different types of the heartbeats. The framework in this thesis revolved around a model that can recognize the underlying essential patterns that defines the different types of heartbeats. This framework used in this work had been able for classifying the test data after training for 50 epochs.

The result of this classification framework was dependent on the various parameters. At the learning rate 0.1, the classification framework converges too quickly leaving the model to be untrained. At the learning rate of 0.000001, the model did not converge enough after 50 epochs but the results possess a better accuracy than in the former one. A cyclic learning rate oscillating between learning rate of 0.001 and 0.00001 is chosen so to get a faster conversion without exploding the loss. Then the batch size and dropout were found accordingly. Making one variable fixed and other changeable, the best value for the model was found.

On training the model without the shortcut function the model gives the ambiguous result. The loss curve and the accuracy curve against epoch curve were evidence that the model found it difficulty in increasing its performance. Input was passed to the model without the RR information of the ECG data. The model was well capable to identify other types of the heartbeats excluding the S-type heartbeats.

Macro-averaged is calculated as an average of all classes while the weighted average considers the number of the samples of each class in the data and average is calculated based on the average weights by the class size.

For the detection of the S-type heartbeat the RR-interval was passed through RNN layer along with the extracted feature of the ECG signals and hence the model did well with the classification. The class imbalance problem persists in MIT BIH data, focal loss was used instead of the traditional cross entropy loss. Performance of the both cost functions were compared against each other. The accuracy of the model did not seem to be fluctuating on the use of any of the two cost functions. However, the true positive rates of the minority class rates increased while using the focal loss. This signifies that the use of the focal loss can cause the model to detect the minority class objects.

The result of implying this algorithm points that it was possible to classify among different types of heartbeats from the ECG signal on this model. The least accurate data found out was the F-type heartbeat. It has relatively fewer numbers of data. This could be the cause for the lower accuracy fort that particular heartbeat. Again F-type are the fusion type heartbeats i.e. the heartbeat has features belonging to more than 1 class. This means that the F-type heartbeat contains feature of both the N-type heartbeats and S-type heartbeats. This makes the F-type to be categorized in N-type and V-type. According to the AAMI standard the F-type arrhythmia is no life threatening. Thus the accuracy of the F-type is considerable. The recall of the S-type heartbeat is also low as compared to the other type of heartbeats. The main reason behind this is the similar QRS complex of normal heartbeat and the S-type arrhythmia. That is the model is still not that efficient to capture all the S-type arrhythmia. This must be improved. Data augmentation might help to resolve this issue as the larger size training data might help the model to characterize the S-type well different from the N-type data.

Limitations and future Enhancement

The model was still not able to perform a proper classification over the detection of the S-type. This can be made better with taking the whole ECG signal rather than just the segment of the heartbeat [9]. From Figure 5-2 we can see that the noise removal process had smoothed the signal. During the smoothing process crucial patterns have been lost. Thus any other way of smoothing the signal might be efficient in this classification process such as by CWT [23]. These techniques might help the signal to retain its key pattern during the feature extraction process. However the result shows a satisfactory performance over the existing models. The future enhancement of this work will include:

- Increasing the S-type heartbeat accuracy
- Generalization of data with the developed model
- Improving the classification performance
- Deployment of this model in smart appliances like smart watch.

CHAPTER 7 REFERENCES

[1] "Cardiovascular diseases(CVDs)," [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). [Accessed 24 03 2021].

[2] U. Acharya, S. Oh, Y. Hagiwarya, J. Tan, M. Adam, A. Gertych and R. Tan, "A deep convolutional neural network model to classify heartbeats," *Computers in Biology and Medicine,* vol. 89, pp. 389-396.

[3] J. H. Medicine, "The Johns Hopkins Arrhythmia Service," The Johns Hopkins Medicine, 2018.

[4] S. Chen, W. Hua, Z. Li, J. Li and X. Gao, "Heartbeat classification using projected and dynamic features of ecg signal.," *Biomedical Signal Processing and Control,* vol. 31, pp. 165-173.

[5] Z. Zhang, J. Dong, X. Luo, K. Choi and X. Wu, "Heartbeat Classification using Disease-specific feature selection," *Computers in Biology and Medicine,* pp. 79-89, 2014.

[6] F. Liu, X. Zhou, J. Cao, Z. Wang, H. Wang and Y. Zhang, "A LSTM AND CNN BASED ASSEMBLE NEURAL NETWORK FRAMEWORK FOR ARRHYTHMIAS CLASSIFICATION," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* pp. 1303-1307, 2019.

[7] S. Kiranyaz, T. Ince and M. Gabbouj, "Real-time-patient-specific ecg classification by 1-d convolutional neural networks.," *IEEE Transactions on Biomedical Engineering,* vol. 63, no. 3, pp. 664-675, 2015.

[8] A. Sellami and H. Hwang, "A robust deep convolutional neural network with batch-weighted loss for heartbeat classification," *Expert Systems with Applications,* vol. 122, pp. 75-84, 2019.

[9] S. Xu, M.-W. Mak and C. Cheung, " Towards end-to-end ECG classification with raw signal extraction and deep neural networks," IEEE J Biomed Health Inform , 2018.

[10] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng and H. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet:

Components of a new research resource for complex physiologic signals. Circulation," 2000. [Online].

[11] L. S¨ornmo and P. Laguna, "ECG Signal Processing," in *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, vol. 8, Academic Press, 2005.

[12] A. Randazzo, "The Ultimate 12-Lead Placement Guide (With Illustrations)," Prime Medical Training, 2016.

[13] Healio, "Introduction to ECG," Cardiology, 2019.

[14] Hilmy, Iwan and Tessy, "Deep Learning Algorithm for Arrhythmia Detection," International Electronics Symposium on Knowlege Creation and Intelligent Computing, 2017.

[15] R. Perez, A. Marques and F. Mohammadi, "The application of Supervised Learning through Feed-forward Neural Networks for ECG signal classification.," IEEE Canadian Conference on Electrical and Computer Engineering, 2016 .

[16] N. Donges, "Towards Data Science," 28 February 2018. [Online]. Available: https://towardsdatascience.com/recurrent-neural-networks-and-lstm-4b601dd822a5. [Accessed 17 April 2021].

[17] T. Lin, P. Goyal, R. Girschik, K. He and P. Dollar, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2999-3007, 2017.

[18] J. Gao, H. Zhang, P. Lu and Z. Wang, "An Effective LSTM Recurrent Network to Detect Arrhythmia on Imbalanced ECG Dataset," *Journal of Healthcare Engineering*, 2019.

[19] J. Brownlee, "Gentle Introduction to the Adam Optimization Algorithm for Deep Learning," May 2021. [Online]. Available: https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/.

[20] C. Salvatore, A. Ceresa, P. Battista, Go, M. ardo, A. Quattrone and I. Castiglioni, "Machine learning models for the differential diagnosis of vascular early diagnosis of Alzheimer's patient: A Machine learning Approach," *Front Neuroscience*, vol. 9, p. 2015, 2015.

[21] R. M. Cruz, R. Sabourin and G. Cavalcanti, "Dynamic classifier selection: Recent advances and perspectives.," *Information Fusion,* vol. 41, pp. 195-2016, 2018.

[22] K. Serkan, T. Ince and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," IEEE Trans Biomed Eng , 2015.

[23] T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu and C. Ou, "Automatic ECG Classification Using Continuous Wavelet Transform and Convolutional Neural Network," *Entropy,* vol. 23, no. 1, 2021.

[24] M. GB and M. RG, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Eng in Med and Biol,* vol. 20, no. 3, May-June 2001.

[25] G. Bianchi and R. Sorrentino, in *Electronic filter simulation & design*, MCGraw-Hill Professional, 2007, pp. 17-20.

[26] K. He, X. Zhang, S. Ren and J. Sun, " Deep residual learning for image recognition," *qing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 770-778, 2016.

[27] H. Dong, A. Supratak, W. Pan, C. Wu, P. Matthews and Y. Guo, "Mixed Neural Network Approach for Temporal Sleep Stage Classification," *IEEE Transactions on Neural Systmes and Rehabilitation Engineering,* 08 2017.

[28] M. X. Cohen, Analyzing Neural Time Series Data:Theory and Practice, MIT PRESS, 2014.

[29] A. Isin and S. Ozdalili, "Cardiac arrhythmia detection using deep learning," in *9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception*, Budapest,Hungary, 2017.

[30] N. Donges, "Towards Data Science," [Online]. [Accessed 30 05 2021].

[31] T. K.M, "Confusion Matrix," Encyclopedia of Machine Learning and Data Mining, 2017.