USE OF NEURAL NETWORK ALONG WITH OTHER CLASSIC MACHINE LEARNING ALGORITHMS FOR ECG CLASSIFICATION

By

Kunal Tamang

A thesis submitted as the partial fulfilment of the

requirements for the degree of

Master's in Data Science

Liverpool John Moores University

July 2022

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First and foremost, I sincerely want to thank my thesis supervisor Dr. Rituparna Datta for his continuous support, advice, and patience. I am very much grateful for his incredible support and encouragement as well as technical support throughout my research work.

I would also like to thank all the concerned authorities of upGrad for providing me enough time and guidelines to continue my research work. It is their kind support that made the research work much more enthusiastic rather than a hectic schedule.

Finally, I would like to thank my family, my wife, and friends for their understanding and encouragement.

ABSTRACT

Electrocardiogram (ECG) is widely used instrument to monitor the diagnoses of cardiovascular system. It takes a highly skilled experts with a lot of experience to diagnose the ECG signal which consumes a lot of time as well as medical resources. The problem of accurately classifying the ECG signal is very challenging. The results are limited because of shallow feature learning architectures. The problem also relies on looking for the most appropriate features for classifying these ECG signals appropriately. Therefore, Machine learning (ML) solutions as well as Artificial Neural Networks are prevalent to analyze and classify ECG data. Machine learning Algorithms are applied to build a model and the performance of these different models can be differentiated based on evaluation metrics. Artificial Neural Networks learn tasks by examples and composed of stacked transformation. It has recently been popular due to its success in variety of task and therefore, it has lot of potential in improving various clinical practices. It has been shown by various studies that with enough data, ANN can show the accuracy to the level of human-expert cardiologist. In this research, different ML models like KNN, XGBoost, and Random Forest and deep learning models like MLPNN and CNN will be trained, and their performance will be presented based on evaluation metrics. In our paper, we will also discuss about the problems encountered in ECG classification, ANNs classifiers, feature extraction tools, pre-processing techniques, ECG database and measures to evaluate the performance. Furthermore, this paper will represent the limitation, performance, and details of each of the classifier and present the detail analysis of best possible classifier based on evaluation index.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Expansion
AUC	Area under curve
CNN	Convolutional Neural Network
ECG	Electrocardiography
EDA	Exploratory Data Analysis
GPU	Graphics Processing Unit
KNN	K-Nearest Neighbor
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLPNN	Multiple layer perceptron Neural Network
MSE	Mean Square Error
PVE	Percentage of Variance Explained
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristics
SES	Simple Exponential Smoothing
XGBoost	Extreme Gradient Boosting

CHAPTER 1

INTRODUCTION

1.1 Background of the study

Classification of ECG signals has vital importance in heart disease diagnosis, however, there are lot of problems associated with it. One of the problems is the way by which normal ECG seems different for each individual person and the same disease may also shows different signs on ECG signals of different patients. Likewise, different disease can show similar effects on signals of normal ECG. The classification of ECG signal is a multiple class problem which consists of 5 classes as Normal beats (N), Supraventricular beats (S), Ventricular beats (V), Fusion beats (F) and Unknown (Q). Each ECG signal has several beats where each beat consists of P zone, QRS complex and T zone. The features of ECG signal are peaks which are labelled as P, Q, R, S, T, U and intervals represented PR, RR, QRS, ST, and QT and segments labelled as PR and ST. The below figure represents those features, and these features have their normal duration values.

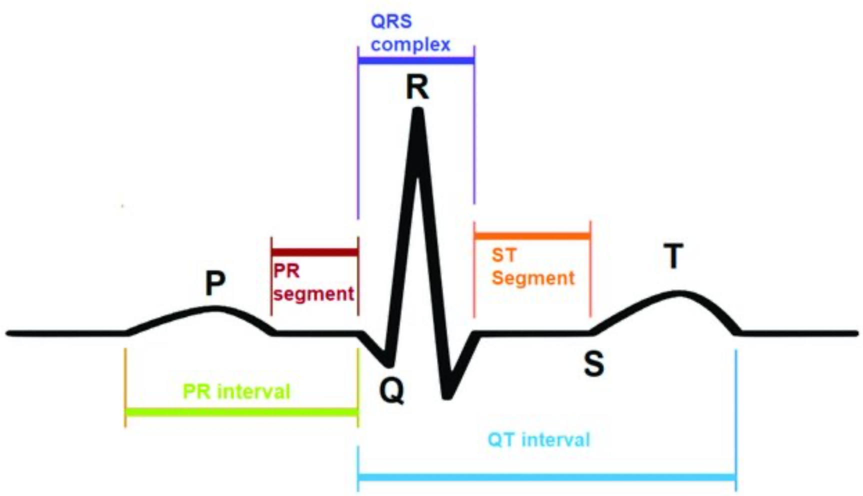


Figure 1.1 A model of an ECG signal (Ullah et al., 2021).

Researchers have employed various preprocessing techniques, feature extraction techniques i.e., Discrete Wavelet Transform (DWT), Continuous Wavelet Transform (CWT), Discrete Fourier Transform (DFT), Principal Component Analysis, etc. For classification, techniques such as Support Vector Machine (SVM) and Multilayer Perceptron Neural Network (MLPNN) has been applied. Researchers (Dallali et al., 2011) have used DWT as a pre-processing technique for extracting RR interval and normalized them using Z score. Likewise, A.

Daamouche et al. (Daamouche et al., 2012) have used SVM classifier as modeling technique for ECG classification. However, there are various challenges involved while classifying ECG signals which consists of standardization problem, feature variation, uniqueness in ECG patterns, lack of optimal models and tools for ECG classification, beats variation of single ECG, etc. (Singh et al., 2012).

1.2 Problem Statement

The purpose of this study is to use Neural Networks such as CNN and MLPNN for ECG classification and compare it with Machine learning models like KNN, Random Forest and XGBoost to demonstrate which classification techniques can efficiently improve the ECG diagnoses. Furthermore, this work will show the detailed analysis while selecting the input beats and presenting the output of the classifiers. This work can be fruitful for the beginners of ECG classification to know the limitations and advantages existing on different approaches and motivates to look further for better classifier.

1.3 Aim and Objectives

This research work is about understanding the patterns in ECG signals and studying different underlying ML models along with Neural networks to accurately classify the ECG signals. This work can be helpful for the early and accurate detection of arrythmia types to detect heart disease and choosing an appropriate treatment.

The research objectives are outlined below:

- To review the research work done in the field of ECG classification
- To utilize the feature engineering techniques to find the trend and correlation between different features.
- To apply ML models and Neural networks for ECG classification.
- To identify the best performing model for ECG classification.
- To illustrate the performance of different models based on evaluation index and describe the limitations existing on each of them.

1.4 Research Questions

This research aims at finding out the most suitable classification techniques for ECG classification based on database provided by the Massachusetts Institute of Technology (MIT-BIH database) having international standards and annotated information by multiple experts. Based on literature review, some of the underlying questions in the field of ECG classification are mentioned below:

- What are the possible influencing factors for ECG classification? What kind of preprocessing techniques and feature recognition results in better performance?
- What are the best performing models based on evaluation metrics for ECG classification?
- What are the scopes, opportunities, and limitations for future works in ECG classification?

1.5 Scope of the Study

For this research, we will apply different types of classification algorithms such as Logistic Regression, KNN, XGBoost and Random Forest (RF) for accurate prediction of the classes of ECG signals and applying the evaluation metrics like ROC curve and Accuracy to differentiate the performance of each model. Moreover, this research work also presents the artificial neural networks such CNN and MLPNN for ECG classification and compare it with classic machine learning algorithms giving a detailed description about each of these techniques. The evaluation of such models will also be based on measures such as Accuracy, F1-score, Roc curve, etc. The performance of each model will be presented in a tables and graphs to compare their effectiveness, accuracy, and limitations. This research work can give a good initial understanding of the problems in this domain of health care sector and possible solutions as well with the help of artificial intelligence. By having these basic understanding, once can look for more robust form of solutions and can further work for a more reliable and efficient algorithms.

1.6 Significance of the Study

This study covers a broad range of machine learning models and neural network in Arrhythmia classification. This research aims at presenting an overview of arrhythmia from medical perspective, ECG databases and evaluation metrics of ECG classifiers. This research can explore the factors responsible for arrythmia diagnoses and lead as a pathway to come up with models for better classification in future. This research will give the detailed analysis of machine learning techniques such as KNN, Random Forest and XGBoost along with Neural Networks such as Convolutional Neural Network (CNN) to classify the ECG signals for Arrhythmia diagnoses. This work will describe about the factors responsible for the difficulties encountered in ECG classification and the possible efficient way for better classification of ECG signals. In a nutshell, it will present a rigorous analysis of different issues encountered in ECG classification, available databases, various preprocessing techniques, various ML models and ANN for better classification, and performance measures based on evaluation metrics.

1.7 Structure of the Study

This final report comprises of six chapters. In Chapter 1, we introduced the context and motivation behind this thesis along with its objectives. In Chapter 2, different machine learning evaluation techniques, neural networks and theoretical concepts regarding ECG is explained in detail. It also provides a detail review about detecting abnormal event in ECG. Chapter 3 describes about the methodologies used for this research and the context explaining the reason, DNN architectures, brief explanation of dataset used, the experimental setup required consisting of hardware and software, and about different data preprocessing techniques and training parameters. In Chapter 4, we will discuss about different types of machine learning algorithm we have implemented in our research and the hyperparameter tuning for different parameters to obtain optimal performance of models. In Chapter 5, we will do the experimental evaluation of the performance of our model based on performance metrics like accuracy, F1-score, confusion matrix, ROC curve, etc. Finally, Chapter 6 will mention about the conclusion that we can draw from our experiment and performance metrices. Furthermore, the future scope of our research and enhancement that can be applied for better classification will be discussed in detail.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

According to World Health Organization (WHO), around 17 million people die every year, and it is associated with the irregularity in heart palpitation rate. The report presented by United Nations in 2018 shows that the aging population is increasing every single day and by 2030, the adults aged between 60 and above is predicted to rise to 1.4 billion which is 56% more from the current population of 901 million. Furthermore, by 2050 this is estimated to reach up to 2.1 billion. This problem of increasing aged population is a challenge to the both the health and economic sector in the world (NATIONS, 2018). The cardiovascular system of an individual weakens with growing age making us prone to diseases and the left ventricle muscle wall and arteries thickens and shrinks depending on age factor, which leads to problems in arteries blood vessels (Acharya et al., 2017). Therefore, understanding electrocardiogram (ECG) signals morphology and detecting and interpreting any heart disease accordingly is extremely important. In cardiology, ECG stands as a very important diagnostic tool to analyze and determine cardiac conditions of different patients specially while dealing with abnormalities such as arrhythmias and life-threatening heart diseases (Isin and Ozdalili, 2017). Arrhythmia is defined as an abnormal or irregular rhythm of heartbeat. It happens due to an ineffective blood pump from heart throughout the body (Johns Hopkins Medicine, 2018).

In this section, we will talk about the anatomy and function of human heart, Arrhythmia with its different types and causes, diagnostic test for Arrhythmia, different previous classification approaches along with the description of ML algorithms and Deep learning neural networks applied to diagnose Arrhythmia cases.

2.2 Anatomy and Conduction system of Human Heart

The mechanism of heart is to pump blood all over the body continuously through blood vessels in our circulatory system and thus, provides oxygen and nutrients in tissues as well as removes carbon dioxide and metabolic waste. The heart is surrounded with pericardium and is located at the center of chest having apex towards the left. It has a size to almost that of clenched fist and weigh around 230-280 grams in case of women and for men it weighs around 280-340 grams. It consists of four chambers (Figure 2.1) where the upper left and right chambers are called atria and lower left and right chambers are called ventricles. The term right heart denotes right ventricle and right atrium together while the left ventricle and the left atrium together is termed as left heart. The right heart function is to oxygenate the blood by pumping deoxygenated blood to lungs and the function of left heart is to take care of cellular respiration by pumping oxygenated blood to all body tissues.

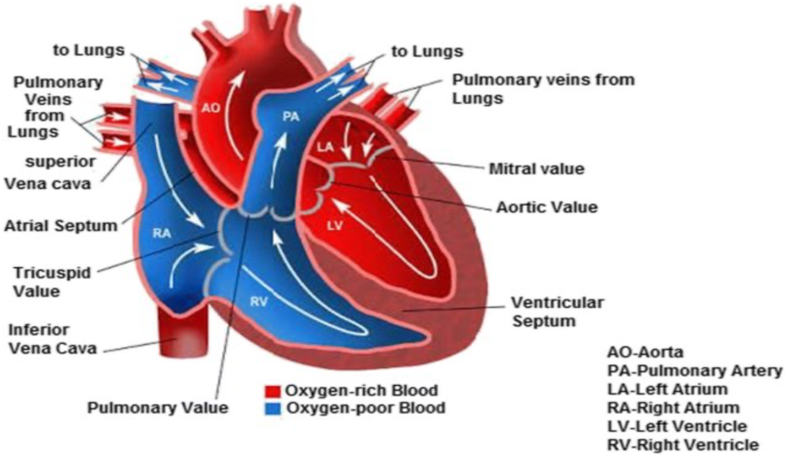


Figure 2.1: Anatomy and physiology of heart (Prabhu et al., 2020).

The heart consists of four valves i.e., pulmonary valve, tricuspid valve, aortic valve, and mitral valve which controls the blood flow direction through the heart. The left atrium and the left ventricle are connected by the mitral valve and the tricuspid valve is there to connect the right atrium and ventricle. There exists pulmonary valve between the pulmonary artery and right ventricle, whereas aorta valve exists in middle of left ventricle and the aorta. There lie four valves in a plane named as atrioventricular (AV) plane. Figure 2.2 depicts the AV plane along with all four valves. A dense connective tissue surrounds these four valves to keep the structure of the valves as well as to isolate the atria and ventricle.

The human heart works continuously for the distribution of oxygen throughout whole body and there are two pathways for blood circulation: the systematic circuit and the pulmonary circuit. The pulmonary artery passes the deoxygenated blood from right ventricle to the lungs in pulmonary circuit. This deoxygenated blood receives oxygen in the lungs and passes through the pulmonary vein towards left atrium. Moreover, oxygenated blood travels through aorta into capillaries and arteries from left ventricle and thus, provides oxygen to tissues in our body. Then, deoxygenated blood returns from tissues to heart and enters right atrium through the venae cavae. This cycle of contraction and relaxation of heart muscles during blood circulation is termed as Cardiac Cycle. The conducting system of the heart includes autorhythmic cells and conducting fibers which generates and conducts electrical impulses through the heart. The conduction system of the heart is represented by given figure below.

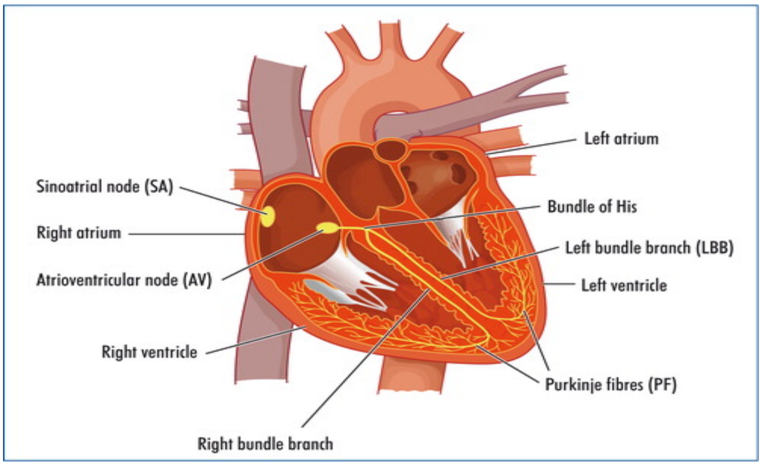


Figure 2.2: Conduction system of heart (Sampson and McGrath, 2015).

The conducting system of the heart consists of Atrioventricular (AV) node, Sino-atrial (SA) node, left and right Bundle branches, a bundle of His, and Purkinje fibers. It consists of SA node located at the junction of the right atrium and superior vena cava is a natural pacemaker of heart. It releases an electrical impulse and initiates cardiac cycle which results in propagation of an electrical wave towards the AV nodes through atria. The electrical wave creates a contraction on both atria by depolarizing atrial myocytes to pump blood towards ventricles. The electrical impulse from SA node is delayed briefly after reaching the AV node situated at the junction of ventricles and atria to allow atria contraction and pump all the blood to ventricles. The tricuspid valve and mitral valve close as soon as atria pumps blood to ventricles. After then, the atria start refilling, and the electrical impulse travels into the Purkinje fibers and Bundle branches through the Bundle of His and AV node. Finally, the ventricular wall cells get contracted as the electrical impulse spread throughout them. Then the left ventricle pumps blood into the aorta and blood is pumped to lungs by right ventricle.

The modification in ion concentration around cell membrane causes the electrical wave to pass from one cell to another in all directions. Sodium ions (Na+) gets into the cell through ion channels once a stimulus is received by cardiac cell from the conduction system which makes the inside cell more positive. Also, these different set of specialized channels are utilized by Calcium ions (Ca2+) to flood inside the cell. The depolarized state of the cell membrane is maintained by the release of calcium which initiates the myocyte contraction. The potassium ions (K+) flows towards extracellular space to balance the increase in membrane voltage. After depolarization, the cell cannot respond to an electrical impulse anymore for a short period which is termed as refractory period. Following depolarization, each cell must repolarize getting back to its relaxed state by re-equilibrating the ion concentrations. This change in potential of electrical membrane is referred as action potential (AP) and the duration for these changes is termed as Action Potential Duration (APD). Figure 2.3 clearly shows the action potential related to each specialized cell. It clearly shows how these waveforms and duration of action potential differs based on different regions of heart. These can be detected by applying electrodes on the skin and the recorded signal here is termed as Electrocardiogram (ECG). The detailed description of ECG will be discussed in following section.

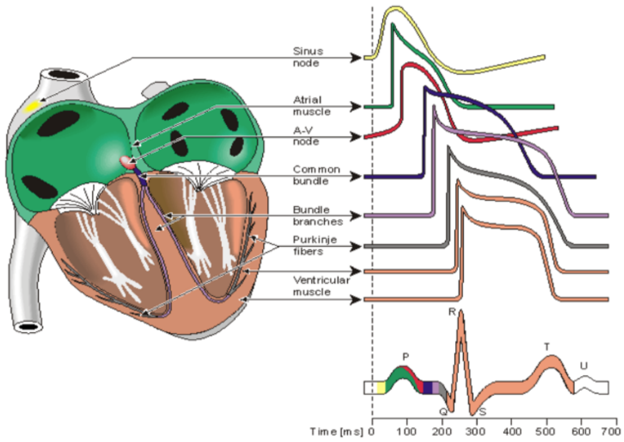


Figure 2.3: Action potential for each specialized cells in heart (Naderi et al., 2010).

2.3 Electrocardiogram

ECG is a prominent tool amongst physician for the analysis and diagnosis of cardia diseases since it depicts the electrical activity of heart based on the insertion of electrodes on skin. This signal's characteristics can be analyzed to understand electrical conduction, rhythm, and anatomy of the heart. Electrophysiologic signals (bio signals) are extracted and analyzed based on digital methods that involves filtering, amplification, processing, digitization, and storage to understand physiological and biological mechanisms. The cardia signal gets generated due to polarization/depolarization of cardiac muscles and the electrical impulse from specialized myocardial cells activates them. There are different bio signals: Electroencephalogram (EEG) records the brain signal, the ECG for recording cardiac signal and the Electromyography (EMG) for muscle activity.

The cardiac signal is cyclic since the heart muscles pumping blood process performed in sequence. ECG signal that consists of series of successively occurring events is only considered as healthy signal since the signal behavior may include random components because of the influence of nervous system. These events can be observed by inspecting ECG signal's amplitude and shape, various segments and intervals and the relationship between them. These waves are identified by the letters P, Q, R, S, T and U and the QRS complex is associated with ventricular contraction. Segments exists on sequential waves and is considered as duration in milliseconds. Intervals refers consecutive waves and segments and their relevant time duration. Figure 2.4 clearly depicts a ECG beat signal consisting of interval and segments.

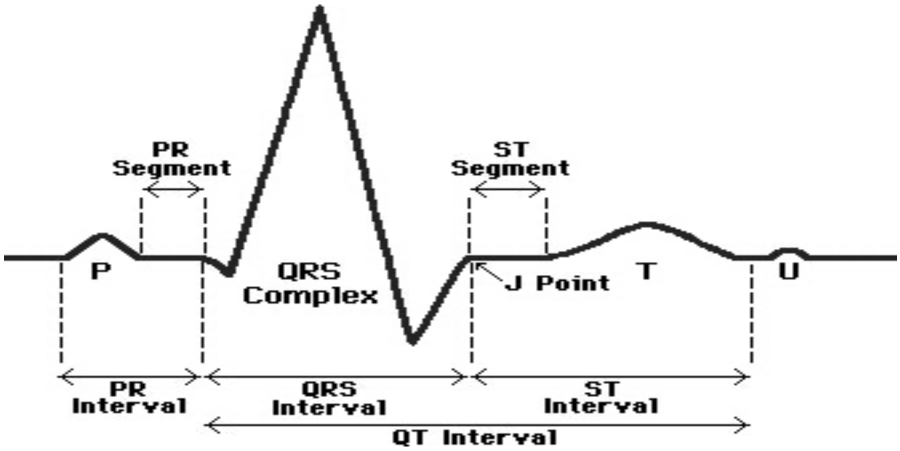


Figure 2.4: Normal ECG waveform (Papechen and Sebastian, 2016).

These waves reflect heart's physiological events and helps to diagnose the cardiac health and abnormalities of a person. The initial P wave refers to depolarization of atrial myocardium existing ahead of atrial contraction. The QRS-complex shows the depolarization of ventricular myocardium and is responsible for blood ejection out of heart. The T wave rises after the end of ventricular ejection and depicts the repolarization of ventricular myocardium which shows the recovery and announcement of new cycle. It is uncertain that U wave has any connection with physiological mechanism of heart however, few theories relate it with repolarization and depolarization.

The heartbeat rhythm is measured with the use of standard 12-lead ECG channels of recordings. Each channel is placed on different parts of patient's body which produces records from different angle. The patient's arms and Legs is attached to six leads and are considered as Limb Leads: Lead I, II, III, augmented Voltage Foot (aVF), augmented Voltage Right (aVR), and augmented Voltage Left (aVL). The other six leads are considered as Precordial leads and are placed on chest region (Healio, 2019) named as V1, V2, V3, V4, V5, and V6. There can be wrong interpretation or inaccurate results if these leads on patient's body is placed incorrectly and hence, it is very important to identify abnormal ECG signals correctly before considering treatment to patients (Acharya & Shu, 2017). For instance, even a 20-25mm deviation in the placement of leads can produce remarkable changes on ECG recording, especially on ST segment (Walker, 2019).

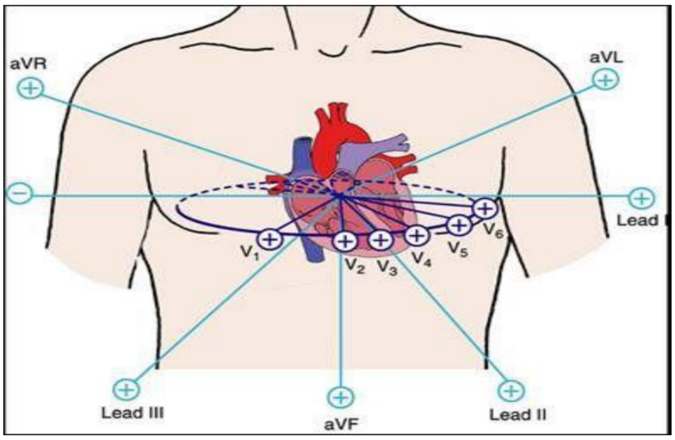
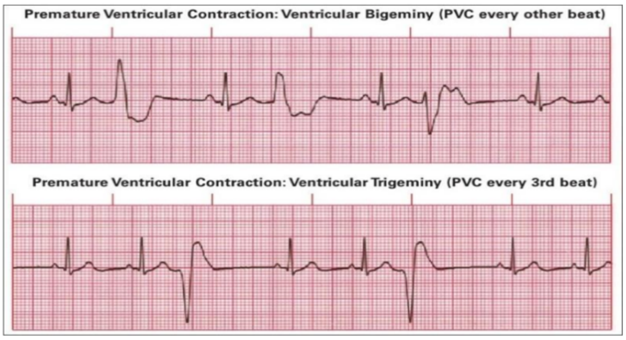


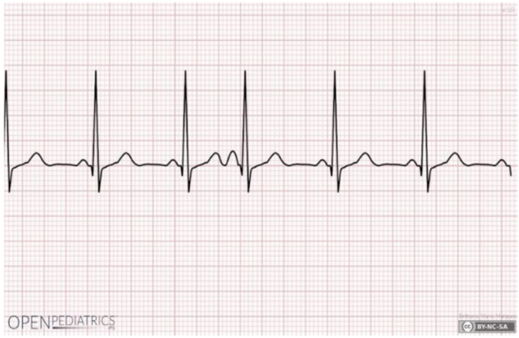
Figure 2.5: 12-Lead ECG Electrode Placement (Randazzo, 2016).

2.3 Arrhythmia

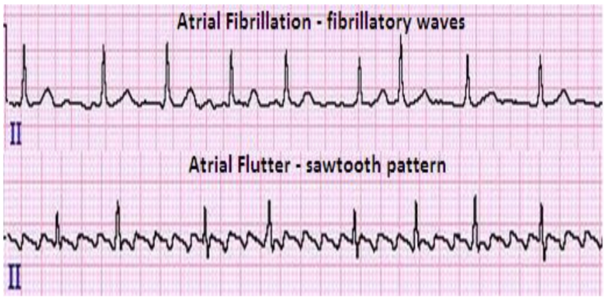
Arrhythmia refers to an abnormal or irregular rhythm of heartbeats. A healthy human has a normal heart rate at range of 60 – 100 Bpm. Those suffering from arrhythmia can feel the palpitation of heart, fluttering and racing heartbeat as well as skipping beats. The symptoms include short breath, chest pain, lightheadedness, fainting, or sometimes even no symptoms. Irregular or Extra heartbeat, Tachycardia (rapid) and Bradycardia (slow) are the examples of arrhythmia.

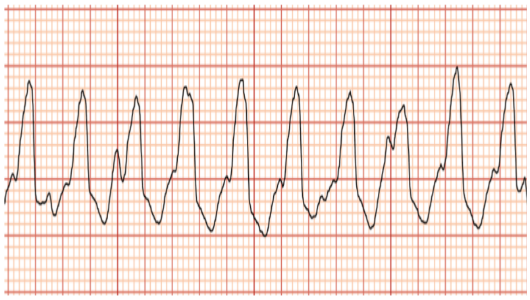
An Irregular heartbeat shows fluttering sensation along with forceful and extra heartbeat. The examples are Atrial premature beat or Contraction (APC) and premature ventricular contraction (PVC). The heart rate of Tachycardia arrhythmia falls on range of 100 to 350 Bpm. Supraventricular tachycardia, ventricular tachycardia, atrial tachycardia, and ventricular fibrillation are the examples of tachycardia arrhythmia. Some of the life-threatening arrhythmias are ventricular fibrillation and tachycardia. The heart rate for Bradycardia arrhythmia is at range of 60 to 0 Bpm which falls well below the normal range. Left and right bundle branch blocks, tachy-brady syndrome and AV heart blocks are the examples of Bradycardia arrhythmia.





(1) **(2)**

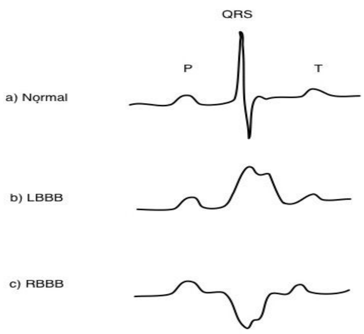




(3) **(4)**

Figure 2.6: Premature Ventricular Contraction (1) (Thomas V. , 2016), **Premature Atrial Contraction (2)** (OpenPediatrics, 2017), **Atrial Tachycardia (3)** (Lofgren, 2018),**and Ventricular Tachycardia (4)** (Schaeffer, 2017).

.



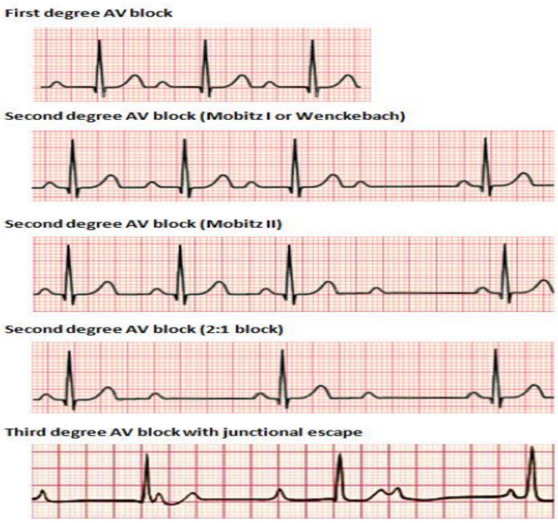


Figure 2.7: Left: Normal, LBBB and RBBB (Richard N. Fogoros, 2019), **Right: AV Heart Blocks** (Lecturio Medical Education, 2017).

Any age group of population can suffer from Arrhythmias, but it is more prevalent with increase in age although having no prior sign of heart disease. Although an individual of any age can suffer from Arrhythmias, the risk usually increases with increasing age even with the absence any heart disease signs. The different types of Arrhythmias are Atrial Tachycardia, Atrial Fibrillation and Extra heartbeats with an estimate of 2.7 million to 6.1 million people living with Atrial Fibrillation in USA and is expected to increase to 12.1 million by 2030 (OMICS, 2019).

There are many factors responsible for causing arrhythmia. Some of the factors responsible for arrhythmia are smoking, hypertension, hypertension, work stress, excessive consumption of alcohol and caffeine (Nordqvist, 2017). A health survey of year 2017 in Finland reported a complain of chest pain from fourth grader student leading to arrhythmia due to excessive consumption of energy drinks containing caffeine while playing video games (yle UUTISET, 2018).

2.5 Diagnostic tests for Arrhythmias

As we know, Electrocardiogram (ECG) is a test to show rhythm of heart's electrical response represented in the form of waveform signals and therefore, applied in checking electrical activities of the heartbeat. The wires connect the machine with legs, chest and arms and are considered as periodic signals as they consist of sequence of waves due to fluctuation of needles on ECG recordings. These waves are P wave, QRS Complex and T-wave. Holter Monitoring is one of test which records 24 to 48 hours of patient's heart continuously while the patient goes with their daily life routines by connecting wires from a portable ECG device to and individual chest (Johns Hopkins Medicine, 2018). Similarly Exercise stress test is another kind of test where a 12-lead ECG is used while being on a treadmill for an exercise (Thomas V., 2016)

2.6 Arrhythmias classification approaches

There have been a lot of different approaches for studying Arrhythmias which evolved over time, and we will be discussing few of them.

To begin with, we can see an application of mixture of experts (MOE) which is hybrid of two different classifiers (Hu et al., 1997). MOE combines the diverse estimates linearly and is formed from two different classifier: general classifier and patient personalized classifier. A collective decision through gate network weighs the classification results obtained from both classifiers based on voting output. From each side of R point, 14 points is taken, and a 29 sample is extracted from QRS time series reducing it to 9 points for the sake of principle components using Karhunen–Loeve (KL) transform. Likewise, artificial neural network of ECG signal in the time domain were anayzed for identification of arrhythmias (Özbay and Karlik, 2001). Two linear discriminant analysis (LDA) classifier was used as a supervised classification approach proposing a hybrid paradigm (De Chazal et al., 2004). There were four features selected from R-R interval and three features were selected from intra-beat segments to calculate seven interval features. There were 19 morphological amplitudes extracted that included 10 samples in the middle of QRS onset and QRS offset for each beat time series. In another research, there were correlation coefficient used in ECG signal for QRS complex and efficient arrhythmia detection algorithm were applied where the correlation coefficient and RR interval were utilized to calculate similarity of arrhythmia (Chiu et al., 2005). The use of SVM and MLP on six different classes of MIT-BIH database using inear discriminant (LD) based feature showed a very high accuracy for different classes (Song et al., 2005). They extracted seventeen features by applying wavelet transform and it was reduced by 4 features by LDA. The SVM classifier turned out to be superior to MLP classifier. K-nearest neighbor algorithm was introduced to classify two types of heartbeat features (Christov et al., 2005). They performed first QRS recognition method to compute 26 morphological features (MF). They considered 165 samples after R and 90 samples before R with a total of 256 sample time series including R-point and extracted a heartbeat time-series of 710 ms at sampling frequency of 360 samples per second. There were 10 features from time-frequency (TF) domain and one residual energy feature were calculated by applying a matching pursuits approach on 256 sample time series. The classification results were based on the application of MF and TF separately for N, LBBB, RBBB, PVC and PB classes.

Later in 2006, a hybrid paradigm was again proposed based on LDA classifier (De Chazal and Reilly, 2006). A classifier on an individual basis was employed for tuning a general (global) classifier which was trained initially. This scheme used purely inter-patient classifier and turned out to outperform those solutions. They introduced weights on training data and addressed the how the imbalance in class distribution affects the model. In this experiment they used 26 features from their previous work mentioned above. To improve their precious work mentioned above they trained 500 beats from training set as well as training set beats. A stacked neural network was proposed to classify four non-AAMI classes (Güler and Übeyli, 2007). They took 256 sample time series from each beat and wavelet coefficients was computed by using Daubechies wavelet transform. For each beat, 19 element feature vectors were formed by extracting standard deviations, average mean, average powers, and adjacent sub-bands mean ratio. There were four outputs and 30 hidden neurons chosen in second stage neural network. Furthermore, the data division scheme was used like that mentioned by De Chazal et al. and AAMI classes to reduce unbalanced classes effect and presented a hierarchical SVM (Park et al., 2008). They segmented 181 time series to apply feature extraction technique on each beat by taking 90 samples on either side of R peak. The use of independent component analysis (ICA) based features for eight different classes of MIT-BIH can be seen proposing support vector machines (SVM) and probabilistic neural network (PNN) (Chou and Yu, 2008). A generalized discrimination analysis-based feature selection (GDAFS) technique was used proposing SVM classifier (Asl et al., 2008). They took R-R interval from two consecutive R points and extracted Heart rate variability (HRV) from ECG time series. There were 32 RR intervals in each segment depicted on time verses RR intervals. We can also see a use of hybrid MOE paradigm proposing a multi-dimensional swarm optimization (PSO) neural network where the training data included both randomly selected common beats training recordings and patient-specific beats which were included from the first 5 minutes of each recording as like that in AAMI practice (Ince et al., 2009). At scale 24, a discrete wavelet transform (DWT) was employed taking R-peak centered feature vector of 180 sample at a time duration of 500 millisecond and a sampling rate of 360 samples per second. We can also see an application of hidden markov models and wavelet transform (Gomes et al., 2009). This tends to outperform the conventional standard linear segmentation.

Four classes of MIT-BIH data were classified by using Levenberg - Marquardt (LM) algorithm to introduce recurrent neural network (RNN) (Derya Übeyli, 2010). Using Jacobi matrices method, 128 Lyapunov exponents were extracted for each ECG beat of 256-point time series.

Similarly, all 15 classes of MIT-BIH dataset were considered and a Support Vector Machine (SVM) based classifier was proposed (Ye et al., 2012). There were 200 points after R point and 100 points before R points taken and 300 sample time series used from each beat. For each beat, features were extracted separately from independent component analysis (ICA) and discrete wavelet transform (DWT). A discrete cosine transform (DCT) coefficients from the segmented beats of ECGs were proposed which were then subjected to principal component analysis for dimensionality reduction, and a probabilistic neural network (PNN) for automatic classification of ECG beats into five categories (Martis et al., 2013). A combination of decision rule and support vector-based ensemble was used for classification in AAMI framework (Huang et al., 2014). They used random projection-based features and applied fifteen support vector machines to tackle 15 sub-classes of 5 AAMI classes respectively. In another research, we can see the classification of non-ectopic originated beats from ectopic originated beats by proposing a radial basis function neural net (RBFNN) (Mateo et al., 2016). They used two class patient-specific classification and two activation functions.

Table 2.1: ECG Classification approaches mentioned in Literature.

Method	**Features**	**Classifier, classes**	**Train Data, Test Data**	**Training Scheme**	**Results Se, PPV**
Hu (1997)	Intervals Morphology (12)	SOM+LVQ (12) KL-PCA, AAMI	15 records Train 20 records Test	First 5-min Patient-specific	Overall Se% 82.6 Overall PPV% 77.7
Chazal (2004)	Amplitudes Intervals	Hybrid LD lead A +lead B	22 records train 22 records test	AAMI Inter Patient	86.8, 75.9, 77.7, 89.4 99.1, 38.5, 81.5, 8.6
Guler (2005)	Intervals wavelet	CNN Intra-patient	Intra-patient Four classes	Non-AAMI 360 Testing	Sp: 97.78% TCA: 96.94%
Chazal (2006)	Amplitudes Intervals	Hybrid LD general+specific	DS1+500 beats 22 records test	AAMI Patient-specific	94.2, 87.7, 94.3, 73.9 99.3, 46.9, 94.3, 29.1

Christov (2006)	Intervals 256 points	k-NN VCG MF	First 3 to 12 min 424 beats general	Non-AAMI Intra-patient	98.5 99.6
Ubeyli (2007)	wavelet-4 (70) Daub 2 (10)	RNN L-M training	small dataset intra-patient	Non-AAMI 4 class	Se% 94.44 Sp% 99.61
Park (2008)	RR intervals HOS and HBF	SVM 250 ms	Train 22 Rec (30 min) Test 22 Rec (30 min)	AAMI Inter Patient	86.3, 82.6, 80.9, 54.9
Babak (2008)	Intervals, Morphology	SVM classifier	small dataset intra-patient	Non-AAMI 6 class	98.9, 98.9, 98.5, 98.5, 100, 100
Ince (2009)	wavelet TI-DWT intervals, PCA	MD PSO ANN 5-min train set	20 records train 24 records test	AAMI patient-specific	Se% 97, 84.6, 63.5,61 Acc%:-, 97.4, 98.3, -
Ye (2010)	Amplitudes Wavelet	SVM 15	Non Chazal 10 subset intra	Non-AAMI 15 classes	Se: 99.91% 99.66%
Eduardo (2011)	Intervals 6 methods	LDC,OPF,SVM	Chazal scheme 22 + 22	AAMI Inter Patient	99.6, 0, 48.0, 48.7, 0
Llamedo (2011)	Intervals DWT scale-4	LDC projection pursuit	22+22 Inter Patient	AAMI Tca: 78	77.5, 76.4, 82.9, 95.3 99.4, 41.3, 88.0, 4.0
Lannoy (2012)	Intervals, HOS Morphology	wCRF +$L1$ MI ranking	Chazal scheme 22+22	AAMI Inter-patient	79.8, 92.6, 85.2, 84.5 BCR% 85.4
Martis (2013)	Amplitudes DCT, PCA	PNN,FNN, SVM Six features	Non Chazal 10 subsets of 44 rec	AAMI intra-patient	Se 98.69% Sp 99.91%,

Huang (2014)	Amplitudes, RR Random proj (51)	Ensemble SVM 15 svms+Rule	Chazal scheme 22+22	3 class AAMI Inter-patient	se 99.2, 91.1, 93.9, ppv 95.2, 42.2, 90.9,
Zhang (2014)	Intervals Morphology	RBF SVM classifier	Inter-patient 22+22	AAMI Inter-patient	88.9, 79.1, 85.5, 93.8
Das (2014)	RR intervals 500 ms beats	RBF SVM S-transform	20 records train data 24 records testing	AAMI first 5 min intra-patient	98.1, 74.0, 91.4, 67.0 97.9, 73.6, 91.8, 80.9
Mateo (2016)	Amplitudes correlation	MLP raised-cosine RBF	small set Ectopic	Intra Patient non-AAMI	Se% 99, 72 Sp% 99, 87

2.7 Machine learning models for ECG Classification

2.7.1. Support Vector Machines (SVM)

SVM is one of the popular classification methods found in literature till date. The main advantage of SVM classifier is that it reduces the local minima and curse of dimensionality in traditional machine learning. It has simple mathematical form, and it is a fast-learning algorithm with good generalization capability and very suitable for high dimensional data. SVM maps the datapoints in higher dimension and creates best set of hyperplanes in feature space between different classes of training dataset as shown in figure below.

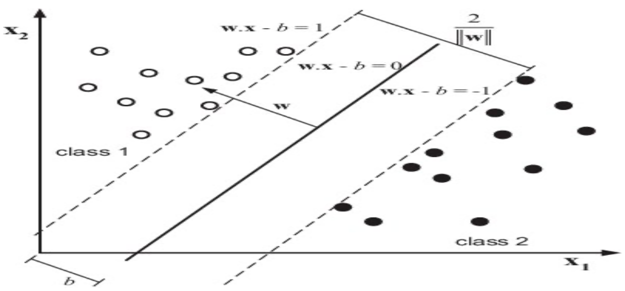


Figure 2.8: SVM classifier showing hyperplane (Papadonikolakis and Bouganis, 2012).

The largest distance between support vectors is given by best hyperplane consisting of datapoints on the edge of class. This leads to increase in accuracy even with small dataset used for training since less sample training points are used. SVM chooses the best suited hyperplane by increasing the margin in hyperplane although large number of hyperplanes can exist. The dataset splits into two classes by linear hyperplane by SVM classifier and if it cannot be split then the dataset in transformed into higher dimension. The SVM algorithm can be understood as the following optimization problem:

If P = (xi, yi), i = 1,2,…..l is a training set where $xi \in R^n$ represents the input feature vectors containing n attributes of a training dataset and $yi \in \{-1,1\}$ is the desired output. Then the optimization problem can be given as:

$$\min_{w,b,\xi} \left\{ \frac{1}{2} w^T w + c \left(\sum_{i=1}^{l} \xi_i \right) \right\} \tag{2.1}$$

Subject to:

$$y_i(w^T \phi(x_i) + \mathrm{b}) \geq 1 - \xi_i, \quad \xi_i > 0 \forall_i \tag{2.2}$$

Where $\phi(x)$ is a nonlinear mapping function to transform the input feature vectors x into a higher dimension, b is bias, ξi is slack variable and w is the weight vector.

The best hyperplane is searched by solving the Lagrangian which is transformed into dual maximizing problem on $Q(\alpha)$ as follows:

$$\max Q(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \tag{2.3}$$

Subject to:

$$\sum_{i=1}^{l} \boldsymbol{\alpha}_i y_i = 0;$$

$$0 \leq \boldsymbol{\alpha}_i \leq C, \quad for\ i = 1,2, \dots\dots, l$$

having,

$$k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

is kernel function and $\alpha = (\alpha 1, \alpha 2, \dots, \alpha_l)$ is a vector of non-negative LaGrange multipliers. There are many research done with the use of SVM classifier. There were different approaches like fuzzy theory proposed with SVM variations combined for better SVM classification, combination with classifiers ensemble, restricted fuzzy SVM and genetic algorithms combination along with least squares SVM. There were significant improvements reported by using SVM in hierarchical manner with maximum voting strategy (Huang et al., 2014). For capturing data from SVM, a new approach was initiated by introducing new kernel function (Moavenian and Khorrami, 2010). In this work, SVM and a Multilayer Perceptron Artificial Neural Network (MLP-ANN) performance were compared using same methodology where MLP performed better based on evaluation metrics while SVM showed efficiency in both training and testing execution time. To minimize the effect of imbalanced classes, balancing techniques like Synthetic Minority Over-sampling Technique (SMOTE) can be implemented at training phase.

2.7.2 Random Forest

Random forest (RF) is another famous supervised machine learning technique which was proposed first by L. Breiman and primarily is an ensemble of decision trees for training and predicting outcomes. Because of its two origins of randomness i.e., bootstrap data sampling and random attribute sub-set selection, it is also termed as stochastic method and identified as parametric algorithm regarding number of trees in forest. The randomness tends to avoid the problem of overfitting while being in training process. The model depends on lots of different parameters like number of trees, maximum split and maximum depth, etc. where the splitting properties are picked at each internal node from a random set of k characteristics. The best split tends to built trees without trimming and are taken with these attributes chosen randomly. Because of it high speed, it is used particularly in sectors having larger number of situations or attributes among many classification challenges.

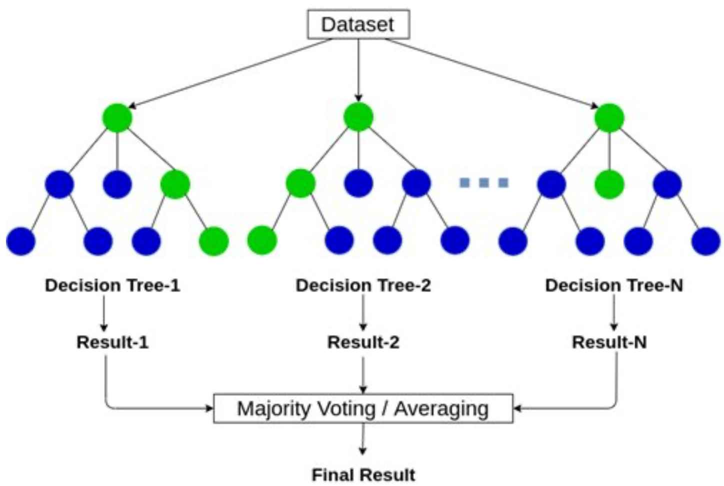


Figure 2.9: Random Forest model architecture (Hao et al., 2021)**.**

2.7.3 K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a commonly used for classification tasks and one of the renowned and simplest supervised machine learning technique. Since it is only based on memory and does not use any model for fitting, it is also known as non-parametric lazy algorithm. It classifies feature vectors in feature space in accordance with labels of closest training samples. For the collection of k-nearest neighbors, distances like Hamming, Euclidean and Minkowski distance are calculated between unknown feature vector and all vectors in training set. With the help of votes from neighbors an unknown feature vector is assigned to class having most of the closest k samples. The class having largest number of votes is considered as prediction. There are many recent ECG classification studies where KNN has been widely used. There are two parameters of KNN classifier: the threshold value to evaluate unusual neighbors and the value of K which shows the number of nearby neighbors.

$$D(X, Y) = \left(\sum_{i=1}^{n} |x_i - y_i|)^p\right)^{\frac{1}{p}} \tag{2.4}$$

where p is the order and $X = (x_1, x_2, ..., x_n)$, $Y = (y_1, y_2, ...,y_n) \in R^n$.

2.7.4 XGBoost

XGBoost stands for “Extreme Gradient Boosting” and hence, it works on gradient boosting framework. It is an open-source library and used by many winning teams in machine learning competitions. Boosting is used to convert weak learners to strong ones and helps in reducing bias and variance in supervised machine learning algorithms. It consists of boosted regression trees and the hyperparameters are updated in gradient boosting method. This algorithm ensembles multiple trees which reduces the variability during classification by summing over the scores for a class corresponding to a leaf predicted by each tree.

$$\widehat{y_i} = h_\theta(x_i) = \sum_{k=1}^{K} f_k\,(x_i),\ \ f_k \in \mathcal{F} \tag{2.5}$$

f_k is a regression tree and acts as a function which maps features to score whereas $\mathcal{F}$ acts as a space consisting of all regression trees.

All the constituting parameters are denoted by Θ and are given as:

$$\Theta = \{f_1, f_2, f_3, \dots \dots . f_k\}$$

The objective function is given as:

$$\mathcal{J}(\Theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \; f_k \in \mathcal{F} \tag{2.6}$$

Through addictive training, the gradient boosting learns Θ. It adds a new function f_k after each round starting from constant prediction. The Θ is completely constructed after ‘k’ rounds. In each round, $\mathcal{J}(\Theta)$ is minimized and f_k is then chosen. Since the gradient of loss function is used while computing for minimization, this method is termed as “gradient boosting”.

2.8. Deep Learning Methods for ECG Classification

Deep learning is branch of machine learning applied in areas such as image classification, computer vision, health care, self-driving cars and so on. It learns massive data and considers as biological brain inspired learning. The key driving factors of Deep learning are computing power, algorithms, and big data (Hemsoth, 2017). It has high accuracy although hard to interpret and it is very important to understand the mechanism behind the decisions made through Deep learning. In Deep learning, self-learning is implemented throughout the process of feature selection, feature extraction as well as classification (Acharya & Shu, 2017) minimizing the drawbacks of conventional machine learning algorithms which uses features like frequency and time domain as well as beat to beat interval without even considering raw QRS morphology to detect anomalies. This allows to respond rapidly to new stimuli as the way the neuron in brain adapts and function based on human experience. A raw input data is fed to a neural network (NN) and weights are adjusted to reach certain goal.

This chapter provides a better understanding of deep learning approaches like artificial neural networks, Feedforward neural networks, optimization and training, deep learning architecture as well as Convolutional Neural Networks.

2.8.1 Feedforward Neural Network

Feedforward Neural Network can be considered as simplest form of Single layered Perceptron and was introduced first by Frank Rosenblatt in 1957. It consists of input, output and hidden layers and the information travels from left to right in forward direction which is due to approximation of function “f” (Kumar, 2019). In this process, weight vector w is acted to an input vector x which passes through activation function f and transformed to response y. This is mathematically represented as:

$$y = f \sum_{i=1}^{n} (w_i \, x_i) \tag{2.7}$$

Feedforward Network is very important since it forms a path to other models. To handle more complex approximation function, the number of hidden layers must be increased between input and output layer. Furthermore, to minimize or amplify the input vectors we can control the weights w by changing it either to positive or negative. Similarly, adding biases on input layer is performed to make sure the output is 1 (Nielsen, 2019). The weights are varied or given zero value to select input vector according to its importance. The drawback of feedforward neural network is that it does not perform well while dealing with time series or sequential data analysis.

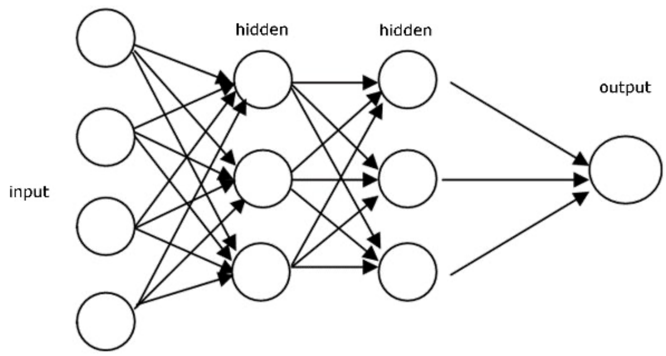


Figure 2.10: Structure of feedforward neural network (Reza, 2020).

2.8.2 Recurrent Neural Network

A Recurrent Neural Networks (RNN) is a memory-based approach while learning data where previous state affects the current state. The weights are altered in each iteration which generates the desired output and then in next step, it is fed to network's input. RNN is mostly used for machine translation, speech recognition, image recognition, spelling correction, text generation and other pattern-based recognition approaches.

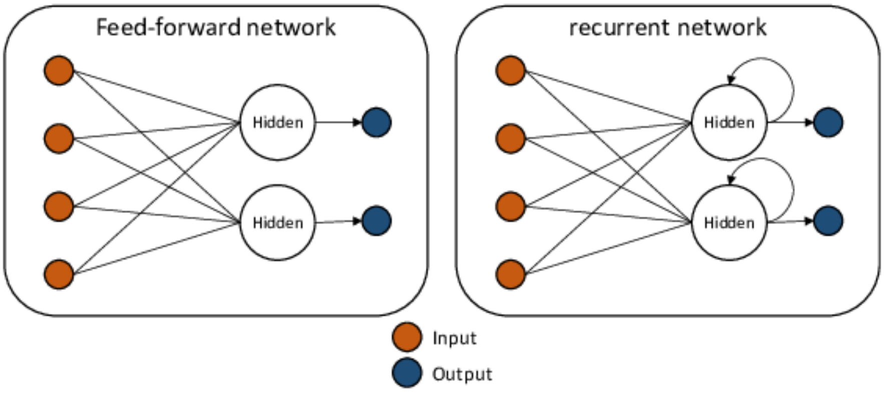


Figure 2.11: An illustration of feed-forward and recurrent network
(Alfarraj and Alregib, 2018)**.**

This kind of sequential data such as handwriting recognition and speech recognition led to the foundation of artificial neural networks like Recurrent Neural Network. RNN can use their internal memory to classify these arbitrary sequences of inputs forming a directed cycle of units connections (Singh et al., 2018). The limitations of CNN over RNN are that we have to cut the beats into pieces having fixed length which results in the lack of performance of classification. The facility of improving the performance of RNN classifier by providing handicraft features to classifier makes it better while comparing with CNN.

2.8.3 Convolutional Neural Network

Convolutional Neural Network are neural networks which uses convolution instead of general matrix multiplication in at least one of their layers and designed especially for image recognition problem and 1-D time series data in computer. They possess one or more convolutional layer and applied to data with grid-like topology (Goodfellow et al., 2017). At around 2012, CNN received a popularity for image classification in ImageNet challenge (Krizhevsky et 2012) although it was first introduced for handwriting recognition (LeCun et al., 1998). There are three layers in CNN network i.e., convolutional layer, pooling layer and fully connected layer. The feature extraction layer in CNN is the combination of pooling layer having sequence in repeated pattern and convolutional layer having reLu activation function (Gibson & Patterson, 2017). Through convolutional filter convolutional layer generates feature map and passed to pooling layers. Finally, the number of classes are determined from fully connected layers where neurons from previous layer is connected to current layer. This type of model is mostly applicable for handwriting classification, image classification, object recognition as well as for diagnostic purpose in medical field. It is mostly useful for datasets where large number of parameters and nodes are to be trained. On the other hand, 1D CNN contains two types of layers i.e., CNN layer which consists of 1D convolutions as well as pooling layer and fully connected layers which are like that of layers of Multi-layer Perceptron (MLP). The configuration for such network is based on hyperparameters like filter size of CNN layer, number of hidden CNN layers and the pooling and activation operators.

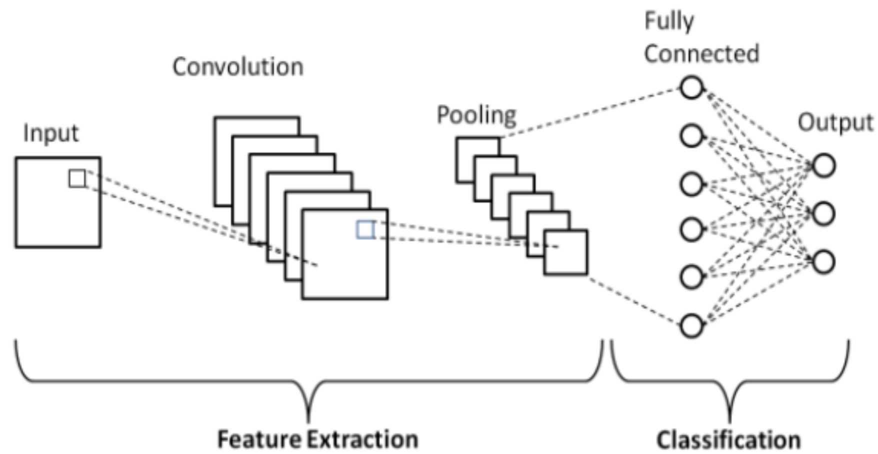


Figure 2.12: Schematic diagram of basic Convolutional Neural Network (CNN) architecture (Phung and Rhee, 2019)**.**

2.8.4 Activation Function

Activation Function is a decision maker at output layer which maps the value to output in a range of 0 to 1 and -1 to 1. It can either be nonlinear or linear and makes sure that the gradient is large throughout hidden layer process and introduces the non-linearity in hidden layer of network. The activation function is supposed to perform backpropagation and hence it is expected to be differentiable (Stewart, 2019).

Activation Function is a very important part of Neural Networks which can either be linear or nonlinear. It is used to map value to output and function as a decision maker. It is used to make the result in a range of 0 to 1 or -1 to 1. The activation function introduces the hidden layer to describe non-linearity of the network and make sure for the gradient to be large. They need to be differentiable for the task of backpropagation (Stewart, 2019). We can expect the output vectors to be that of linear function without activation function. A straight line describes a linear function, and it fits the graph in right proportion and passes through the center which is mathematically represented as:

$$f(x) = Wx \tag{2.8}$$

The choice of activation function depends on the kind of neural network and classification problem presented. Sigmoid function is one of the most used activation functions. It has S-shape on graph and it wide application when the output requires to be posterior probability. It possesses the characteristics for vanishing gradient and slow convergence problems. It has a problem of zero-centeredness where the achieving optimization is quite difficult (Nielsen, 2019).

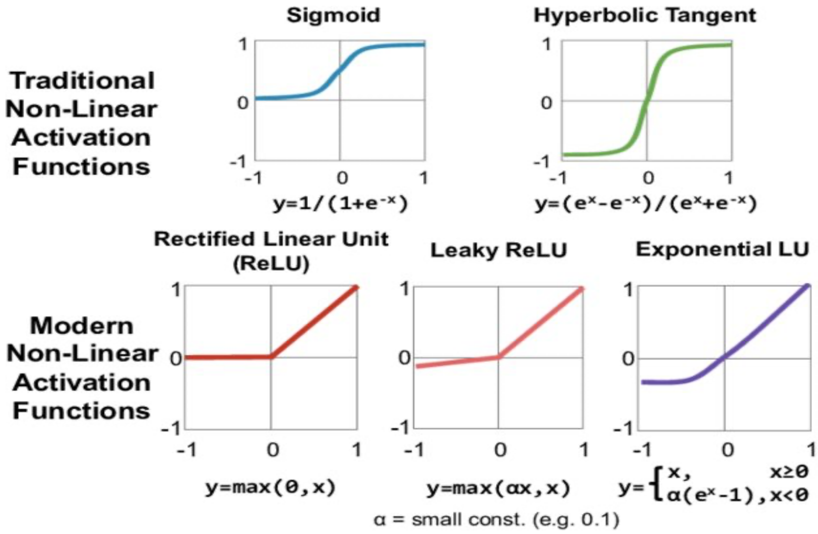


Figure 2.13: Various forms of non-linear activation functions (Sze et al., 2017)**.**

The determinant of activation function is needed for the classification task in neural network. One of the examples of activation function is Sigmoid function. It possesses S-shape on the graph. The limitation of sigmoid function is slow convergence and vanishing gradient. Sometimes, achieving optimization also seems to be a problem using sigmoid function (Nielsen, 2019). It is mathematically represented as:

$$f(x) = \sigma(x) = \frac{1}{1+e^{-x}} \tag{2.9}$$

Hyperbolic tangent (tanh) is another activation function which also forms S-shape on the graph and cannot solve the vanishing gradient problem. However, it can solve the zero-centeredness in sigmoid function and ranges from -1 to 1 (Gibson & Patterson, 2017).

$$f(x) = 2\sigma(2x) - 1 \tag{2.10}$$

Rectified linear function (ReLU) is a non-linear activation function. Its gradient value is either zero or constant so can solve the vanishing gradient issue (Gibson & Patterson, 2017). It is mathematically represented as:

$$f(x) = \max(0,x) \tag{2.11}$$

2.8.5 Cost Function

Cost Function is an error which is computed from the difference of the actual network's output y and desired output ŷ. It tells about the error that need correction in a network. The two major types of cost functions are Mean squared error (MSE) and cross entropy (Stewart, 2019). Mean square error doesn't seem to perform well for modern deep neural network since it learns slowly and suffers mostly from saturation. It is mathematically represented as:

$$Lsq = \tfrac{1}{2} \sum (y - \hat{y})^2 \tag{2.12}$$

Cross entropy on the other hand is mostly used for classification task. For instance, for categorical prediction SoftMax output is used in multi-class classification problem. Similarly, sigmoid output is used for binary classification resulting either 0 or 1. Cross-entropy tends to make the network learn faster and hence, improves its performance with sigmoid output (Stewart, 2019). It is mathematically represented as:

$$L_{ce} = -(y.\log(\hat{y}) + (1-y) \log (1- \hat{y})) \tag{2.13}$$

$$\mathrm{L_{ce}} = -\sum_{\mathrm{i}} \mathrm{y_i} \log \mathrm{y_i} \tag{2.14}$$

2.8.6 Gradient-based Optimization

Optimization process is very important to train a neural network since it minimizes cost function by altering inputs thorough a brute force approach. For instance, when neurons are assigned with the cost function Gradient descent can be applied to update each neurons weight. However, there can be Curse of Dimensionality (Rumelhart et al., 1986) as weights are updated and numbers in network increases. Based on the amount of data used in gradient calculation, there are three different variants: Stochastic gradient descent, Batch gradient descent and Mini-batch gradient descent. The weight of each neuron is update by Stochastic gradient descent which can cause error because of frequent updates but it works faster than batch gradient descent. On the other hand, Batch gradient descent updates all neurons only after their evaluation although its high consumption of computation requires huge amount of memory. Mini-batch gradient descent is commonly used in deep learning and formed by combining both stochastic and batch gradient descent having batch range of 50 to 256 (Ruder, 2016). It is mathematically represented as:

$$\theta_{y+1} = \theta_y - \varepsilon \nabla_\theta f(\theta) \tag{2.15}$$

Where, θ denotes parameter/weight, ϵ as learning rate and $\nabla_\theta f(\theta)$ denotes partial derivatives of the cost function.

2.9 Discussion

This section showed various recognition and classification approaches published in the field of ECG classification. We came through a brief introduction on those topic and discussion of various Machine Learning algorithms along with Deep learning architectures to understand the proposed method and their performances. ECG beat classification itself is very challenging problem. The most recent challenges faced are large variations in ECG waveform morphologies not only from different patients' groups but also on same patient. These tends to be unlike and at the same time alike for different types of beats. Several systems were proposed like SVM, neural network, ML algorithms, linear discriminants, etc. with different preprocessing techniques where all of them have some limitations. For instance, gradient-based training algorithms gets easily confined in local minima and converges slowly. Therefore, addressing all these issues and comparing the performance of each of separate approaches from simple ML algorithms to complex neural networks and suggesting limitations of these approaches will be interesting research to be accounted for the future scope in the field of ECG classification problem.

2.10 Summary

These research reports presented so far clearly shows that cardiovascular disease is one of major cause of death in today's world especially for aged population. The cardiovascular system is prone to disease as the person gets older since the arteries and muscle wall of our heart shrinks due to age factor and therefore, ECG is an important diagnostic tool to understand the cardiac condition of patients. As we know that it takes a lot of hard work even for an experience specialist with years of experience to differentiate between normal and arrhythmic heart, it is very important to apply an automated analysis based on machine learning and deep learning techniques so that the diagnosis method becomes more efficient. Applying machine learning techniques and deep neural network on MIT-BIH arrhythmia database to extract beat signal and comparing those to evaluate and present the best performing model is very important to understand the mechanism behind this developing a lot of research questions and possible solutions. Experimenting with different preprocessing techniques along with different models can further add to improve the performance of the model

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

For this research, different types of machine learning algorithm and deep learning neural network are employed. The main purpose of this research is to compare different ML algorithms and neural networks and evaluate the best performing model based on performance metrics. It also aims at applying different preprocessing and dimensionality reduction techniques along with suitable normalization to further improve the performance of models.

In this section, different types of classification algorithms used in this research are introduced with the mechanism behind the way each works and the pros and cons of each of these models. The dataset used is MIT-BIH Arrhythmic dataset and is explained briefly in this section. The preprocessing techniques like class balancing including Normalization for this research is explained to give a clear picture of how the data will be handled before applying it to any ML models. Furthermore, ML algorithms like KNN, Random Forest, and XGBoost along with Deep Learning neural networks like Multilayer Perceptron neural network (MLPNN) and Convolutional neural network (CNN) are explained briefly. This research uses evaluation metrics like Accuracy, precision, recall and F1 score derived from confusion matrix for evaluating the performance of different models.

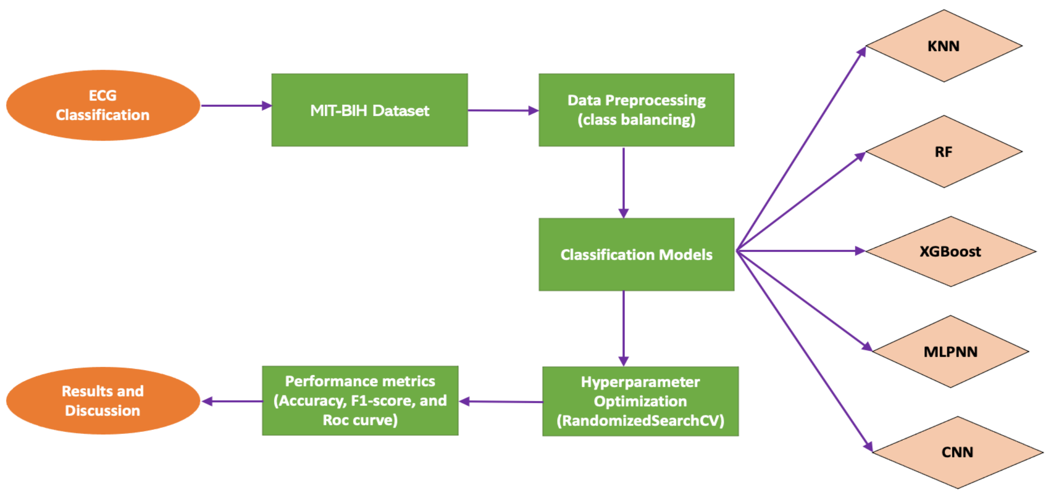


Figure 3.1: Flowchart of ECG classification.

3.2 Dataset

This thesis uses open-source MIT-BIH Arrhythmia database containing 48 records from 47 patients which is hosted by PhysioNet. From a mixed population having 60% inpatients and 40% outpatients at Boston's Beth Israel Hospital 4000 24-hour ECG recordings were collected where 23 recording labelled between 100-124 records were chosen randomly. Similarly, remaining 25 recordings labelled between 200-234 records were selected to include clinically significant and less common arrhythmias that could possibly not be well-represented by a small random sample (Physionet, 2005). For this to achieve, ECG holder recordings was used and each records contains two lead channels MLII and V1-V5 having band-pass filtered at 0.1-100 Hz for 30 minutes long duration. Additionally, these recording were interpreted by at least two cardiologists (Physionet, 2005). Therefore, these labelled ECG signal made analysis and preprocessing easy to perform. The performance of an arrhythmia detection algorithm should be based on five major categories of heartbeat according to Association for the Advancement of Medical Instrumentation (AAMI) standard to test and report cardiac rhythm performance. These five categories of heartbeats are "N" which is for beats in sinus node, "S" for supraventricular ectopic beats (SVEBs), "V" for ventricular ectopic beats (VEBs), "F" which is fusion of normal and VEBs and "Q" which includes the one that doesn't fall under N, S, V, and F categories or an undefined heartbeat as well the one originating from the use of pacemaker operated under skin of patient. Q beats are comparative less meaningful for the study of arrhythmia classification and hence we exclude this from our study. This practice is also followed in other studies like state-of-the-art. The detailed topology of the dataset is shown below.

Table 3.1: Categories in Arrhythmic Dataset

<table>
<tr><td>Categories</td><td>N</td><td>S</td><td>V</td><td>F</td></tr>
<tr><td rowspan="5">Definitions</td><td>Normal beat</td><td>Atrial premature beat</td><td rowspan="2">Premature ventricular Contraction</td><td rowspan="5">Fusion of Ventricular and normal beat</td></tr>
<tr><td>Left bundle branch block beat</td><td>Aberrated atrial premature beat</td></tr>
<tr><td>Right bundle branch block beat</td><td>Nodal (junctional) premature beat</td><td rowspan="3">Ventricular escape beat</td></tr>
<tr><td>Atrial escape beats</td><td rowspan="2">Supraventricular premature beat</td></tr>
<tr><td>Nodal (junctional) escape beat</td></tr>
<tr><td>Annotations</td><td>90,585</td><td>2,781</td><td>7,325</td><td>802</td></tr>
</table>

3.3 Data Preprocessing

The imbalance dataset can create a problem during training and a model might overfit and predict the most represented category almost all the time and therefore, results in accuracy paradox. For example, if we have a dataset with two labels A and B and 90% records belong to A, then there is a possibility that classifier might predict class A lot of times and results in 90% accuracy although this cannot be defined as good classifier. There are different approaches which can be applied to reduce this: using abnormalities detection algorithms, applying class weights, and performing data balancing. The class weights if applied using logarithmic function of the proportion of ratio of total instances divided by class instances represented mathematically as:

$$\text{Class weight} = \min\left(\log\left(\frac{\mu * \text{totalinstances}}{\text{classinstances}}\right), 1\right) \tag{3.1}$$

Where, μ is a tuning parameter.

This approach helps to remove overfitting at early stage of training. Another approach h is by applying data balancing. Some of the commonly used techniques for data balancing are Under-sampling which is by deleting the instances present in over-represented classes, Over-sampling which is by repeating training of classes that are under-represented, Generating Synthetic Samples which is simply generating new samples based on available samples and using algorithms like decision trees to handle imbalanced data. Furthermore, data normalization can be applied to direct model to learn absolute values of all instances. To make the final scale in a range of -1 to 1 the mean values of record can be subtracted from record values and divided by the absolute value of the record.

3.5 Experimental setup

It requires substantial amount of computing power for deep learning. The heavy calculation of convolution and backpropagation along with crucial amount of display memory requires GPU parallelization in deep neural network. To compute one pass of data millions of bytes of data are often required which depends on how complex the model is. The number of simple matrix calculation machine learning can also be efficiently performed. It is possible to do optimization in mini batches using mean values of whole batch where the batch size depends on the amount of memory. The experiment is conducted by using desktop computer with Intel Core i5-8250U and 1.8 GHz central processing unit (CPU), 8 GB RAM and Nvidia GeForce 920MX GPU.

The library used are the ones commonly used in deep learning framework like Tensorflow, Keras, and PyTorch. The experiment is implimented using Python 3.6 on Anaconda Jupyter. The libraries intalled are Numpy to deal with arrays, pandas to manipulate data and Scikit Learn to analyse data.

3.6 Classification Algorithms

3.6.1 K-Nearest Neighbor

K-Nearest Neighbor (KNN) is one of the simplest forms of classification algorithms. It works by finding the minimum distance between training data and reference values in feature space. It was developed to apply the discriminant analysis. It is lazy learning algorithm since there is no need of generalization from training data sets and they are only needed during testing phase. The algorithm computes the K closest neighbors for an instance whose class in unknown and a voting with those neighbors is done to assign the class to that unknown instance. The training phase of this algorithm is a lot faster than testing phase. Thus, it is a critical task to choose the value of K and distance metric to compute the nearest distance. There are two phases of KNN: Training and classification phase. The training examples in training phase are vectors in a multidimensional feature space where the class labels and feature vectors are stored from training samples. In the classification phase, *K* is a user-defined constant where a test point is classified by assigning a label among the K training samples to the most recurrent nearest to that query point. The value of K is chosen in a way to achieve highest correct classification rate. If x refers to true value of ECG parameter and y refers to measured value and N as the number of values, the distance metric is represented mathematically as:

$$\text{Distance } d(x, y) = \sum_{i=1}^{N} \sqrt{(x_i - y_i)^2} \tag{3.2}$$

3.6.3 Random Forest

Random Forest is a bootstrap aggregating ensemble method. Decision Trees are created on different datasets by sampling with replacement. It performs feature bagging which means that a random subset of features is selected for training each decision tree rather than utilizing all the features. Furthermore, best split is chosen on random subset of features while splitting a node during tree construction. Thus, the bias of the forest increases slightly but the variance decreases because of averaging and therefore, compensating the increase in bias leading to a better performing model.

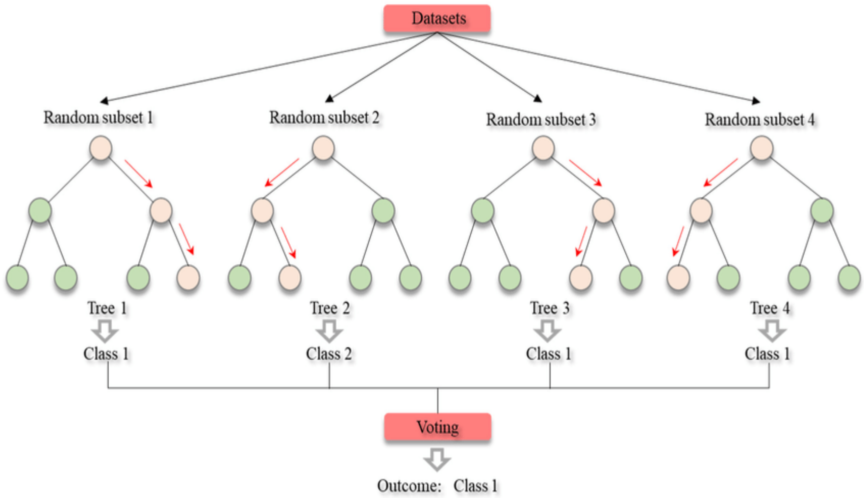


Figure 3.2: Random Forest Architecture (Yang et al., 2019)**.**

3.6.4 XGBoost

XGBoost is widely known for extraordinary speed and performance along with its ability to deal with missing data, large dataset and skewed class distribution which fulfills our requirement to choose this algorithm. XGBoost stands for Extreme Gradient Boosting and it shows machine learning tree models. It works as an ensemble of decision tree algorithm where the errors can be fixed made from earlier trees using pruning strategy. The addition of trees to a model is based on the improvement of the model. It is based on term Gradient Boosting where boosting refers to machine learning meta-algorithm to reduce bias and variance occurred in supervised learning so that the weak learners can be changed to stronger ones. In this process, the iteratively learning weak classifier is added to a stronger classifier. It consists of boosted regression trees where the hyperparameters learn through gradient boosting method. It functions by decreasing the classification variability from the ensemble of multiple trees.

3.6.5 Multilayer Perceptron

Multilayer Perceptron (MLP) is a main branch of feedforward artificial neural networks consisting of at least three layers of nodes. It utilizes backpropagation technique to train like that of supervised learning method and can distinguish data which are linearly not separable. For linearly separable data, there is linear activation function which can linearly map input to output whereas for non-linearly separable data, the algorithm uses non-linear activation function like sigmoidal or logistic function. This algorithm is widely used in field of image and speech recognition as well as machine translating software.

In this work, the MLPNN is composed of input layer having same size as that of feature vector, hidden layer and output layer corresponding to respective ECG beats and a sigmoid function is used as an activation function.

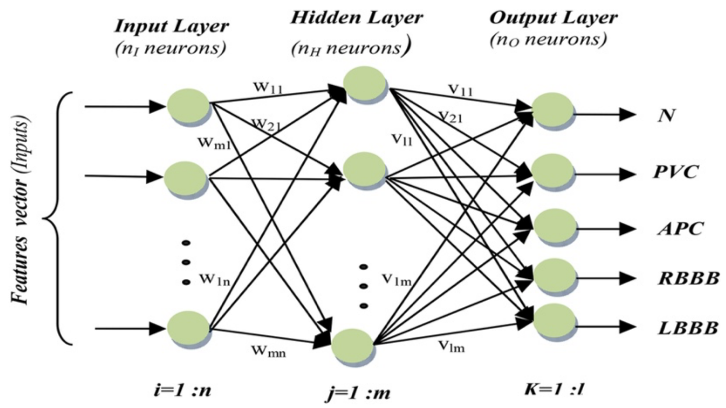


Figure 3.3: Structure of MLPNN (Bouaziz et al., 2019)**.**

There is weights and bias assigned to each layer during feature training and calculating output. The output from each neuron of hidden layer is calculated as:

$$y_j = f(\sum_{i=1}^{n} w(i,j)x_i), \qquad j = 1, 2, \ldots\ldots\ldots m \tag{3.3}$$

where x_i denotes input variable, y_j denotes the output of each neuron j in the hidden layer, f refers activation function, and w is the weights assigned to neuron in hidden layer, n is the number input layer neurons and m is the number hidden layer neurons.

3.6.6 Convolutional Neural Network

Convolutional Neural Network (CNN) is most common neural network for image classification. The inspiration behind CNN is a human visual system (Sze et al., 2017). They represent the state-of-art approach for object detection, pattern recognition and image classification. A deep CNN solution demonstrated as the champion of ImageNet Large Scale Visual Recognition Challenge 2012 competition (Krizhevsky et al, 2012). It distinguishes itself from other pattern recognition since it combines both feature extraction and classification. This network consists of five different layers: input layer, convolutional layer, pooling layer, fully connected layer, and output layer. These layers are divided as feature extraction and classification. The feature extraction consists of Input layer, Convolution layer and Pooling layer and the classification consists of fully connected layer and output layer. The size of an input images is specified by input layer which is then convolved by sharing weights with multiple learned kernels by convolution layer and then, the image size is reduced maintaining the information by pooling layer. The output of feature extraction layer is termed as feature maps. The classification result is the output of classification layer.

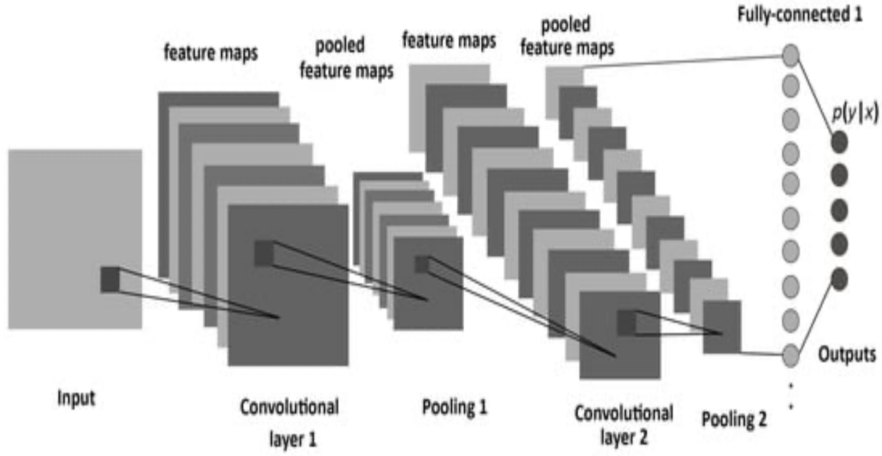


Figure 3.4: Schematic diagram of basic convolutional neural network (CNN) (Albelwi and Mahmood, 2017) **.**

3.7 Performance metrics

The percentage accuracy along with precision, recall and F1 scores are the measures to evaluate the effectiveness of CNN model for classification problem. A confusion matrix statistic having true-positive data (T_p), false-positive data (F_p), false-negative data (F_n) and true-negative data (T_n) is used to calculate these measures. The metrics are defined as follows:

$$Overall\ Accuracy = \frac{T_{p1}+T_{p2}+\cdots+T_{pN}}{\Sigma} \quad (3.4)$$

$$Recall = \frac{T_p}{T_p+F_n} \quad (3.5)$$

$$Precision = \frac{T_p}{T_p+F_p} \quad (3.6)$$

$$F1\ score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (3.7)$$

To up-sample the data to make it uniform from an uneven class distribution of our dataset is important so that the model is not biased towards a particular class. Also, data augmentation can improve model accuracy. Accuracy can be reliable metric if false-positives and false-negatives shows similarity whereas for large difference in false-positives and false-negatives precision and recall measures can be more reliable which ultimately produces F1 score measure. Based on these values of accuracy, precision, recall and F1-score, we can also show the confusion matrix to know the number of correct and incorrect predictions made by our classifier.

Table 3.2: Confusion Matrix

	Total	Positive	Negative
Actual	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Predicted

In this table, True Positive values refers to predicted positive values which are positive in actual, True Negative values refers to predicted negative values which are negative in actual, False Positive values refers to predicted positive values but actually are negative and False Negative values refers to predicted negative values which are actually positive.

3.7 Summary

This section gave us brief explanation of different kind of ML algorithms and Deep Learning neural networks this will be implemented for this research. The effectiveness and drawback of each algorithm from lazy ML algorithms like KNN to complex convolutional neural network can provide better understanding for classification problem of ECG heartbeat. The knowledge of MIT-BIH Arrhythmic database with their different categories and the preprocessing techniques like random sampling for imbalanced dataset gave better understanding of how the research is proceeded. Furthermore, the tools required to carry out research were presented that consists of required hardware, software, and libraries like Keras, pandas, numpy, scikitlearn, etc. The brief explanation of classic ML algorithms like KNN, Naïve Bayes, Random Forest, XGBoost along with deep learning framework like MLPNN and CNN added to understand the pros and cons of each of them and how the comparison of each of these models based on evaluation metrics can contribute to further advancement in future goals. The explanation of evaluation metrics like accuracy, precision, recall and F1 score added the information about how the models needs to be evaluated to inspect the model performance.

CHAPTER 4

MACHINE LEARNING ALGORITHMS IMPLEMENTATION

4.1. Machine Learning Models

We have implemented different machine learning models like KNN, Random Forest and XGBoost. For this task, we used different python libraries like pandas to work with data frames, numpy for mathematical calculation, sklearn for training and testing of dataset and implementing ML models and Keras to work with neural networks. All the detail explanation of each model, training and testing of data as well as hyperparameter optimization is discussed in following sections.

4.1.1 K-Nearest Neighbor

K-Nearest Neighbor is one of the commonly used algorithms for classification. Before implementing our KNN, we did some data preprocessing so that our algorithm performs well while evaluating performance metrics. Our dataset has 5 different classes with 72471 samples for class 0, 2223 samples for class 1, 5788 samples for class2, 641 samples for class 3 and 6431 samples for class 4. As we previously discussed, these classes refer to Normal beats (N), Supraventricular beats (S), Ventricular beats (V), Fusion beats (F) and Unknown (Q). The dataset is highly imbalanced and therefore, we used resampling technique to first balance our dataset where we upper sample the classes with minimum number of samples up to 5000 and down sample the classes with higher countdown to 5000. This is plotted and shown in figure below.

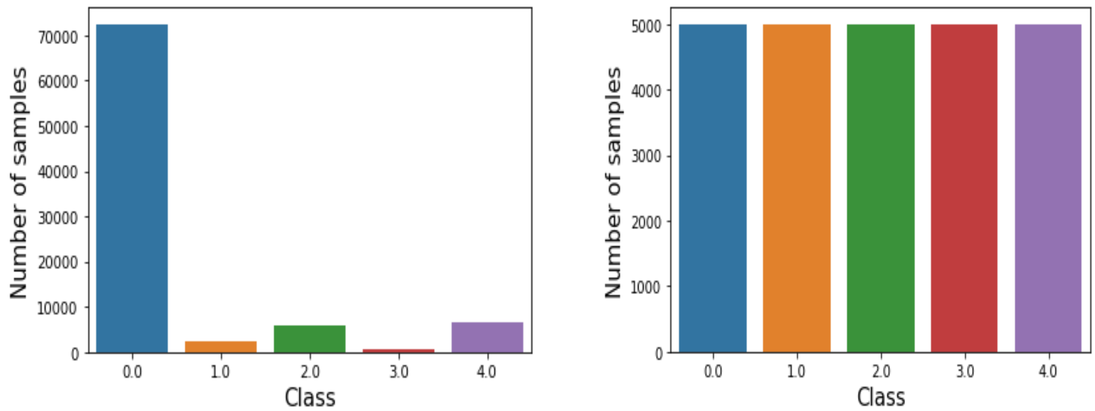


Figure 4.1: Number of samples for each class in original training dataset (left) and balanced training dataset (right).

In our work, we have used Euclidean distance as a parameter to measure the distance between nearest neighbors. The hyperparameter optimization process involves looking for the best value for the number of nearest neighbors. To find the optimal number of number of neighbors, we found the loss function and plotted with reference to number of neighbors. We took the range of neighbors from 1 to 100. We got 1 nearest neighbor as the optimal value of number of neighbors parameter when implemented RandomizedSearchCV to do the same.

4.1.2 Random Forest

Random Forest is an ensemble technique which uses ensemble of decision trees for training and predicting outcomes. It uses the method of bagging also called as bootstrap aggregation where it creates different set of training subset from training set sample with replacement and the majority vote is used to predict the final output. Here, individual decision trees are implemented for each sample parallelly to generate the final output based on major voting for classification. For our work, we used Randomized Search CV for hyperparameter optimization. There are different parameters like 'max_depth', 'n_estimators', 'max_samples', 'min_sample_split' in Random Forest that needs to be optimized. Here, 'n_estimators' refers to the number of trees in the forest and we set it in a range of 1 to 100 with 5 steps. The criterion we applied is the default one which is 'gini' for Gini impurity. The 'max_depth' parameter denotes the maximum depth of the trees and we set it in a range of 1 to 100 with 5 steps.

4.1.3 XGBoost

XGBoost is one of the widely used algorithm for multiclass classification problem. We are applying same balanced dataset to all our algorithms. One vs all scheme is used for solving our classification problem. To apply this algorithm, we first binarize the training and testing output and then OneVsRestClassifier is used to fit the model in training set of input data. We used RandomizedSearchCV to optimize our parameters. The parameters we used here to optimize are learning rate, max depth, and n_estimators. Overfitting is one of the main problems with gradient boosting decision trees and hence learning rate needs to be optimized, also referred as shrinkage to slow down learning and make our algorithm effective for classification. We set our learning rate as 0.01, 0.1, 0.2 and 0.3 giving us the optimum value for learning rate as 0.1. The max depth refers to the maximum depth of trees and it is set from 1 to 30 with an interval of 5. The optimal value of number of trees for our model was 200.

4.2. Deep Learning Models

Deep Learning Models like Multi-layer perceptron and Convolutional Neural Network is applied for our research project. The concept of Deep Neural Network is applied in Deep Learning Models to perform the task of regression or classification. We have used tensorflow and keras to apply these neural networks and train our model for classification. The training of these models and their performance based on loss, accuracy and f1 score is used.

4.2.1 Multilayer Perceptron Neural Network

Multilayer Perceptron is one of main branch of feed forward artificial neural network. As we already discussed, this network consists of input layer, number of hidden layer and output layer. For our work, we have used two hidden layers with output layer applying sigmoid activation function for all three layers. The number of neurons we used for first hidden layer is 187 while for second hidden layer we use 64 neurons. The output layer consists of 5 neurons. We have used sparse categorical cross entropy for our loss function since our data is classified into five different categories of arrythmia disease. We have used accuracy as a measure of performance metrics to evaluate our model. We split our training set such that 10% of data is used for validation part while 90% is used for training part. The model is trained for 15 epochs to measure the performance based on accuracy.

4.2.2 Convolutional Neural Network

Convolutional Neural Network consists of sequence of different layers like convolutional layer, maxpooling layer and dense layer. The objective of convolutional layer is to extract high level features like edges of the input with padding to maintain or increase the dimension of input channel. This convolution operation is done by applying a kernel/filter to extract convolved feature of input. In our work, we have used three convolution layers with 32, 64 and 128 number of neurons having filter size of (3,), (3,) and (5,) dimension respectively. The activation function used is a relu function. The padding parameter is set to 'same' so that input and output size remains same. The MaxPool1D layer is used with same padding and stride value set to 2. A dropout layer is used then to activate only those neurons which plays significant role and updating the weight parameter of corresponding active neurons. Then we flatten these data to feed into fully connected neural network and layers here are referred as dense layers consisting of input layers, hidden layers, and output layers. There are two hidden layers with 512 number of neurons and 1024 number of neurons and both uses Relu activation function. The output layer consists of 5 neurons with softmax activation function. The model is compiled with ‘adam’ optimizer and taking ‘sparse_categorical_crossentropy’ as loss function. We have used accuracy as metric to evaluate the performance of model. The model is run upto 50 epochs for the optimal value of validation accuracy and error. We have further showed this with a plot of epochs with respect to training and validation accuracy and error. The above defined parameters are optimal parameter for our model and they are adjusted after lots of hit and trials in accordance with the plot and the accuracy/error with infer from each different trials.

CHAPTER 5

EXPERIMENTAL EVALUATION

5.1 Accuracy and F1 score

Accuracy and F1-score are reliable metrics to measure the performance of different classification models. We found the accuracy and F1-score for each different classes using the classification report for each individual models. Although accuracy is commonly preferred metric to measure the performance of different classification algorithms, F1-score is one of the suitable metrics for our problem since it denotes how the classifier stand out while dealing with minority classes and this is very important for our problem because of imbalanced dataset. The higher value of F1-score for a particular class denotes the higher ability of the model to classify that specific class.

Table 5.1: Classification Report of KNN.

Class	Precision	Recall	F1-score	Support
0	0.99	0.86	0.92	18118
1	0.24	0.82	0.37	556
2	0.72	0.92	0.81	1448
3	0.22	0.90	0.35	162
4	0.94	0.96	0.95	1608
Accuracy			0.87	21892
Macro Avg	0.62	0.89	0.68	21892
Weighted Avg	0.94	0.87	0.90	21892

Table 5.2: Classification report of KNN with hyperparameter optimization.

Class	Precision	Recall	F1-score	Support
0	0.99	0.94	0.96	18118
1	0.40	0.79	0.53	556
2	0.80	0.93	0.86	1448
3	0.41	0.81	0.54	162
4	0.96	0.96	0.96	1608
Accuracy			0.91	21892
Macro Avg	0.71	0.89	0.77	21892
Weighted Avg	0.96	0.93	0.94	21892

From above table, we can see the difference in the accuracy of KNN algorithm after hyperparameter optimization although the only parameter used for our optimization is the number of nearest neighbors through RandomizedSearchCV which gave us the optimal value of number of nearest neighbors as 1. The Precision, Recall, F1-score, and Accuracy values are the one we consider for the performance of our model. The Support column denotes the number of samples belonging to each class in testing dataset giving us the total test sample of 21892 consisting of all classes. The F1-score for class 0, class 2 and class 4 is comparatively higher showing the model being efficient for classifying majority classes, however, for minority classes, it's comparatively lower. This can be improved with hyperparameter optimization as shown in above table.

Table 5.3: Classification report of Random Forest Algorithm.

Class	Precision	Recall	F1-score	Support
0	0.99	0.96	0.97	18118
1	0.55	0.80	0.65	556
2	0.86	0.94	0.90	1448
3	0.51	0.81	0.62	162
4	0.95	0.97	0.96	1608
Accuracy			0.95	21892
Macro Avg	0.77	0.90	0.82	21892
Weighted Avg	0.96	0.95	0.96	21892

Table 5.4: Classification report of Random Forest with hyperparameter optimization.

Class	Precision	Recall	F1-score	Support
0	0.99	0.96	0.97	18118
1	0.52	0.81	0.63	556
2	0.84	0.94	0.89	1448
3	0.46	0.85	0.59	162
4	0.96	0.97	0.96	1608
Accuracy			0.95	21892
Macro Avg	0.77	0.90	0.82	21892
Weighted Avg	0.96	0.95	0.96	21892

In case of Random Forest Algorithm, we saw the higher accuracy of 95% which remain consistent even after hyperparameter optimization. However, there is slight decrease in F1-score for minority classes like class 1 and class 3. Overall, the algorithm performs well for our classification purpose.

Table 5.5: Classification Report of XGBoost with hyperparameter optimization.

Class	Precision	Recall	F1-score	Support
0	0.99	0.83	0.90	18118
1	0.41	0.79	0.54	556
2	0.82	0.91	0.86	1448
3	0.36	0.86	0.51	162
4	0.95	0.97	0.96	1608
Accuracy			0.90	21892
Macro Avg	0.71	0.87	0.76	21892
Weighted Avg	0.96	0.85	0.89	21892

For XGBoost algorithm, we optimized the learning rate and n_estimators which refers to number of trees for our model. This gave us the accuracy of 92%. However, the F1-score for minority classes like class 1 and class 3 are comparatively lower which signifies the inefficiency of model to deal with minority classes.

Table 5.6: Classification Report of MLP Algorithm.

Class	Precision	Recall	F1-score	Support
0	0.99	0.90	0.94	18118
1	0.33	0.82	0.47	556
2	0.76	0.93	0.84	1448
3	0.24	0.88	0.38	162
4	0.93	0.96	0.94	1608
Accuracy			0.89	21892
Macro Avg	0.65	0.90	0.71	21892
Weighted Avg	0.95	0.90	0.92	21892

Applying MLP Algorithm gave us the accuracy of 90%. The F1-score for minority classes is comparatively lesser than classes having large number of samples. This shows the inefficiency of model to correctly classify these minority classes.

Table 5.7: Classification Report of CNN Algorithm.

Class	Precision	Recall	F1-score	Support
0	0.99	0.95	0.97	18118
1	0.47	0.86	0.61	556
2	0.87	0.95	0.91	1448
3	0.50	0.88	0.63	162
4	0.95	0.98	0.97	1608
Accuracy			0.96	21892
Macro Avg	0.76	0.93	0.82	21892
Weighted Avg	0.97	0.95	0.96	21892

CNN Algorithm perform well compared to other algorithms if we consider the F1-score. The accuracy is 95% and it also classifies minority classes well comparatively as can be seen through the F1-score values.

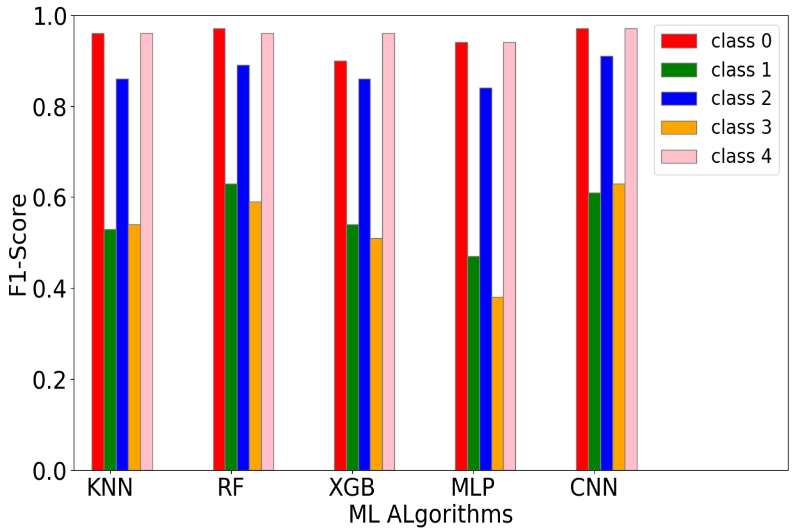


Figure 5.1: Comparison of F1-score among different algorithms.

This bar plot simply represents the summary of above table consisting of F1-score and Accuracy of our model. This clearly shows, how CNN and Random Forest classify each class well compared to other algorithms. For the majority classes i.e., class 0, class 2 and class 4, the F1-score is higher for all the five different algorithms, however, for minority classes the classification is poor. Even though, we still can see the minority classes i.e., class 1 and class 3 being classified well by RF and CNN algorithms while comparing with other ML algorithms.

5.2 Confusion Matrix

Confusion matrix shows the summary of correct and incorrect predictions made by classifier in tabulated form. It evaluates the performance based on accuracy, precision, recall as well as F1-score. We plotted the confusion matrix for each algorithm where all the y-axis labels denote the actual classes and x-axis labels denotes the predicted classes.

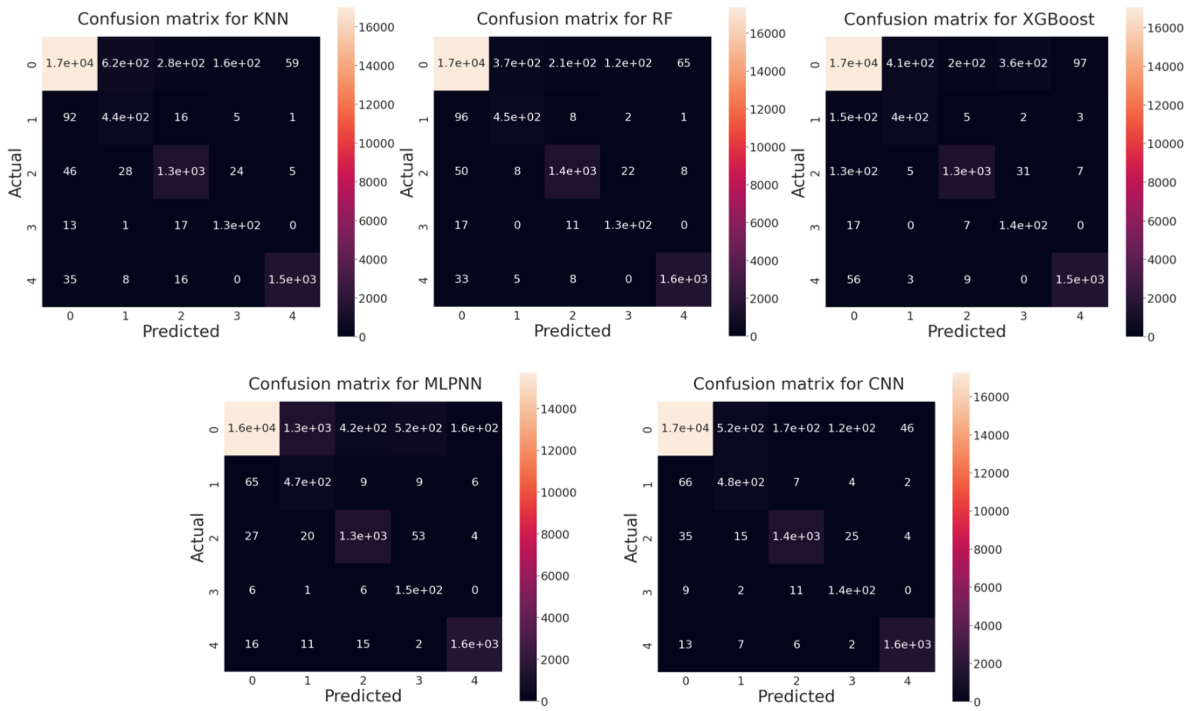


Figure 5.2: Confusion matrix for different ML Algorithms.

The above confusion matrix for different ML algorithms shows that all the algorithms classify the majority class very well but due to the class imbalance of the data most of the minority classes are misrepresented as class 0.

5.3 Loss Function

Loss functions are means of learning for machines. The performance of specific algorithm on given data is evaluated by loss function. A large value of loss function denotes deviation of predicted values from actual values. Therefore, optimization is done to so that the loss function learns to reduce the error in prediction. There is never a loss function which fits for all types of machine learning problems and algorithm. These loss functions are described based on the problem of regression or classification. For our classification purpose, we are using CrossEntropyLoss function while dealing with neural network model and defining error rate and accuracy to portray the optimal value of parameters chosen while optimizing hyperparameters.

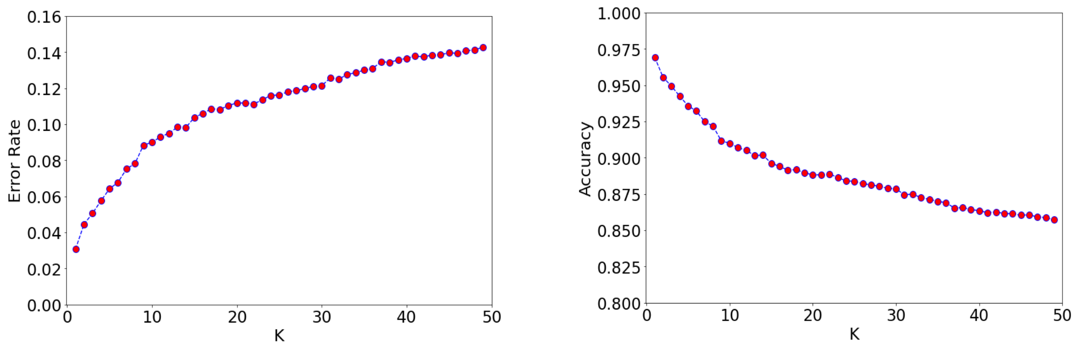


Fig 5.3: Variation of Error rate (left) and Accuracy (right) with respect to number of nearest neighbors in KNN algorithm.

The figure above depicts how the change in 'k' value of KNN algorithms affects the error and accuracy of a model showing us the optimal value of k equal to 1. In our case, we have used RandomizedSearchCV to get this optimal value of hyperparameters. There are many hyperparameters involved and modifying them according to the problem is very important for the better performance of model.

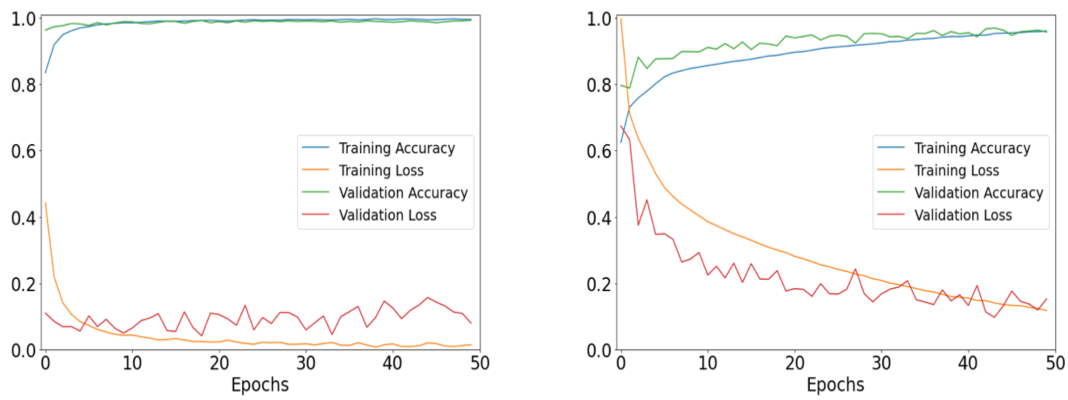


Fig 5.4: Loss and Accuracy of MLPNN (left) and CNN (right) with respect to Epochs.

In case of MLPNN and CNN Algorithms, we ran our model through 50 epochs to look for the minimum loss and high accuracy. The dataset is split into training and validation set where we seek for higher accuracy for validation set compared to training set to optimize the number of epochs.

5.4 ROC Curve

ROC curve is another way to evaluate the performance of our machine learning models at different classification threshold. In ROC curve, the x-axis represents the false positive rate and y-axis denotes the true positive rate. Here, we look for the AUC which refers to area under the ROC curve and represents the degree of separability. The higher the area under curve, better is the performance of model for classification. In our research, we made the ROC curve showing the performance of different models for classifying individual classes.

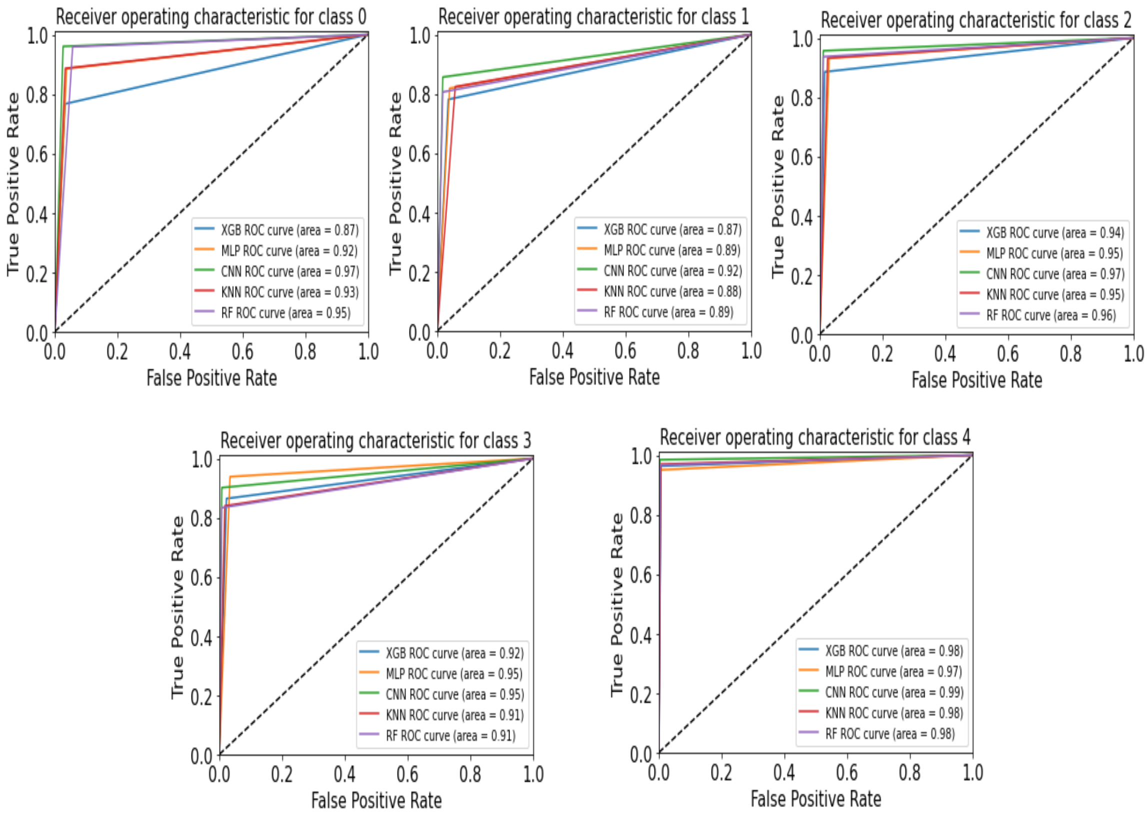


Fig 5.5: ROC curve for individual classes comparing different ML algorithms.

In the figure above, we can see the ROC curve for different ML algorithms and the corresponding areas covered by this curve. Considering all the classes, the performance of CNN algorithm is better compared to other algorithms. We can also see the performance of RF algorithm performing better for classification. However, the AUC for XGBoost, MLPNN and KNN are smaller than CNN and RF considering all classes. The results of the ROC curve are consistent with the confusion matrix and the accuracy and F1-score from above table.

5.5 Summary

The proposed different ML algorithms show how well they classify each different class of Arrythmia dataset. We proposed different evaluation metrics to measure the performance of each algorithm. Our dataset comprises of 5 different categories of heartbeats which were defined by 5 different classes i.e., class 0, class 1, class 2, class 3, and class 4 as given in dataset description above and these datasets were imbalanced having large number of samples for class 0, class 2 and class4 and comparatively smaller number of samples for class 1 and class 3. In our dataset, class 0 referred to Normal beats, class 1 referred to supraventricular ectopic beats, class 2 referred to ventricular ectopic beats, class 3 referred to fusion beats and class 4 referred to unknown beats which are described in detail in introduction section.

In our classification process, we observed that CNN and Random Forest algorithm shows high F1-score for each class as shown in table above and Fig. 5.1. We obtained an accuracy of 96% for CNN algorithm and 95% for Random Forest algorithm. The accuracy score of KNN turned out to be 91% whereas MLPNN gave an accuracy of 89% and XGBoost gave an accuracy of 90%. To obtain this, we applied several hyperparameter optimization for different algorithms as explained above. For example, the accuracy score of KNN before optimizing the 'k' value which denotes the number of nearest neighbors was 87% as shown in table above. We used RandomizedSearchCV to perform this operation of hyperparameter optimization. Furthermore, the graph of error rate/accuracy with respect to optimizing parameter showed the optimal value of parameter to be assigned to get less error and maximum accuracy as shown in fig. 5.3. Similarly, for neural networks, we can visualize the training and validation accuracy and loss with respect to number of epochs as shown in fig. 5.4 to obtain the optimal number of epochs for higher testing accuracy in our model. Beside this, the confusion matrix in fig. 5.2 reveals the number of correct predictions made by each classifier for different classes. In addition to this, ROC curve also stood out as a very important measure to evaluate the performance of different algorithms. The area under the curve (AUC) for each ROC curve in fig. 5.5 clearly shows how each model perform with respect to each other while classifying different classes of heartbeat. The AUC for CNN and Random Forest were much better compared to KNN, XGBoost and MLPNN which is consistent with other performance metrics described in our research.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

6.1 Introduction

In this section, the conclusion derived from the findings of our study about different ML algorithms approach towards ECG classification are described. The conclusion of our work is based on research questions, purpose, and results of the study. The implication of our findings and the recommendation will be explained in detail. The importance of our research work as well as the possible future improvements will be discussed.

6.2 Discussion and Conclusion

As discussed in earlier sections, the objective of our study is to use different ML algorithms for ECG signal classification and comparing them to illustrate the limitations of different algorithms. We have used Arrythmia dataset which consists of 5 different categories of heartbeats. These 5 different categories were described as Normal beats which were denoted by class 0, supraventricular ectopic beats denoted by class1, ventricular ectopic beats denoted by class 2, fusion beats denoted by class 3 and unknown beats denoted by class 4. The prediction of arrythmia from ECG dataset is a challenging task which involves classification of future ECG signals. We had to tune different hyperparameters for different ML algorithms through trial and error to get best possible results. For instance, learning rate had a major impact in the performance of model. The smaller value of learning rate leads to very slow learning of model, and we had to run through lot of epochs to reach the best performance. Similarly, high value of learning rate was unstable and made uncertain whether the global optima has been reached. Furthermore, other hyper parameters like 'max depth' of decision trees in RF and XGBoost model, 'k' in KNN model, etc. needed to be optimized to reach the best performance of models. We further applied different architecture of neural networks to consider the one which leads to higher values in performance metrics. Different performance metrics like F1-score, Accuracy, confusion matrix and ROC curve were used to compare the performance of individual algorithms and the results were consistent with each other. In our work, considering overall performance, CNN algorithm performed very well giving us high accuracy, F1-score for different class, and high area under the curve in ROC curve. We represented these things both in graphical and tabular form to deal with actual numbers being correctly and incorrectly classified as well as to visualize the overall performance of model in graphical representation.

6.3 Importance of Study

The diagnosis of cardiovascular disease is still a difficult problem to deal with and hence, classification of ECG signal is an important task, and many new models are still researched and introduced for better classification of these signals. Our study is an introductory step for those interested in learning the application of machined learning for ECG classification. We have described how different models perform with respect to each other and the limitations that could exists in their application. It is important to note that although the performance of different models is described by values of our performance metrics and are defined as optimal values for our research, these values can be modified leading to better performance by tuning hyperparameters and architectures along with introducing more balanced dataset so that our machine can learn without biasness giving less error. The architecture and hyperparameters optimization defined in our research can help newcomers to look for the part which can be improved so that they can work on understanding the problem in more detail and introduce more efficient results.

6.4 Future Work

The scope of ECG signal classification is very broad and there are lots of works still going on to solve these problems with new and advanced algorithms. As already mentioned, our work is a very brief introduction for solving the problem using machine learning techniques. However, this can give good initiation for understanding the problem and looking for more efficient and reliable solution. As we can see the highest accuracy among our models was that for CNN giving 96% accuracy. One can look for better architecture or combination of different models or some other neural networks like Recurrent Neural Network and many more to look if they can possibly reach close to 100% accuracy without overfitting the data.

REFERENCES

Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M., Gertych, A. and Tan, R.S., (2017) A deep convolutional neural network model to classify heartbeats. *Computers in Biology and Medicine*, [online] 89August, pp.389–396. Available at: https://doi.org/10.1016/j.compbiomed.2017.08.022.

Albelwi, S. and Mahmood, A., (2017) A framework for designing the architectures of deep Convolutional Neural Networks. *Entropy*, 196.

Alfarraj, M. and Alregib, G., (2018) Petrophysical-property estimation from seismic data using recurrent neural networks. November.

Asl, B.M., Setarehdan, S.K. and Mohebbi, M., (2008) Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artificial Intelligence in Medicine*, 441, pp.51–64.

Bouaziz, F., Oulhadj, H., Boutana, D. and Siarry, P., (2019) Automatic ECG arrhythmias classification scheme based on the conjoint use of the multi-layer perceptron neural network and a new improved metaheuristic approach.

Brien, J.O., (n.d.) Using Hidden Markov Models and Spark to Mine ECG Data. pp.4–5.

Celesti, F., Celesti, A., Carnevale, L., Galletta, A., Campo, S., Romano, A., Bramanti, P. and Villari, M., (2017) Big data analytics in genomics: The point on Deep Learning solutions. *Proceedings - IEEE Symposium on Computers and Communications*, Iscc, pp.306–309.

De Chazal, P., O'Dwyer, M. and Reilly, R.B., (2004) Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 517, pp.1196–1206.

De Chazal, P. and Reilly, R.B., (2006) A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 5312, pp.2535–2543.

Chiu, C.C., Lin, T.H. and Liau, B.Y., (2005) Using correlation coefficient in ECG waveform for arrhythmia detection. *Biomedical Engineering - Applications, Basis and Communications*, 173, pp.147–152.

Chou, K.T. and Yu, S.N., (2008) Categorizing heartbeats by independent component analysis and support vector machines. *Proceedings - 8th International Conference on Intelligent Systems Design and Applications, ISDA 2008*, 1, pp.599–602.

Christov, I., Jekova, I. and Bortolan, G., (2005) Premature ventricular contraction classification by the Kth nearest-neighbours rule. *Physiological Measurement*, 261, pp.123–130.

Daamouche, A., Hamami, L., Alajlan, N. and Melgani, F., (2012) A wavelet optimization approach for ECG signal classification. *Biomedical Signal Processing and Control*, [online] 74, pp.342–349. Available at: http://dx.doi.org/10.1016/j.bspc.2011.07.001.

Dallali, A., Kachouri, A., Samet, M., Dallali, A., Kachouri, A. and Samet, M., (2011) A Classification of Cardiac Arrhythmia Using WT, HRV, and Fuzzy C-Means Clustering Classification of Cardiac Arrhythmia Using WT, HRV, and Fuzzy C-Means Clustering. *Signal Processing: An International Journal (SPJI)*, [online] 5, p.101. Available at: https://www.researchgate.net/publication/237102471.

Derya Übeyli, E., (2010) Recurrent neural networks employing Lyapunov exponents for analysis of ECG signals. *Expert Systems with Applications*, 372, pp.1192–1199.

E., T., El-Sayed, A. and R., S., (2016) A Survey on Classification of ECG Signal Study. *Communications on Applied Electronics*, 65, pp.11–16.

Gomes, P.R., Soares, F.O., Correia, J.H. and Lima, C.S., (2009) Cardiac arrhythmia help - Diagnosis system using wavelets and Hidden Markov Models. *Proceedings of the 2nd WSEAS International Conference on Biomedical Electronics and Biomedical Informatics, BEBI '09*, September 2018, pp.60–63.

Güler, I. and Übeyli, E.D., (2007) Multiclass support vector machines for EEG-signals classification. *IEEE Transactions on Information Technology in Biomedicine*, 112, pp.117–126.

Hao, L., Kim, J., Kwon, S. and Ha, I. Do, (2021) Deep learning-based survival analysis for high-dimensional survival data. *Mathematics*, 911, pp.1–18.

Hu, Y.H., Palreddy, S. and Tompkins, W.J., (1997) A patient-adaptable ECG beat classifier using a mixture of experts approach. *IEEE Transactions on Biomedical Engineering*, 449, pp.891–900.

Huang, H., Liu, J., Zhu, Q., Wang, R. and Hu, G., (2014) A new hierarchical method for inter-patient heartbeat classification using random projections and RR intervals. *BioMedical Engineering Online*, 131, pp.1–26.

Ince, T., Kiranyaz, S. and Gabbou, M., (2009) A generic and robust system for automated patient-specific classification of ECG signals. *IEEE Transactions on Biomedical Engineering*, 565, pp.1415–1426.

Isin, A. and Ozdalili, S., (2017) Cardiac arrhythmia detection using deep learning. *Procedia Computer Science*, [online] 120, pp.268–275. Available at: https://doi.org/10.1016/j.procs.2017.11.238.

Kachuee, M., Fazeli, S. and Sarrafzadeh, M., (2018) ECG heartbeat classification: A deep transferable representation. *Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018*, pp.443–444.

Korürek, M. and Doğan, B., (2010) *ECG beat classification using particle swarm optimization and radial basis function neural network*. *Expert Systems with Applications*, .

Li, H., Yuan, D., Ma, X., Cui, D. and Cao, L., (2017) Genetic algorithm for the optimization of features and neural networks in ECG signals classification. *Scientific Reports*, [online] 7January, pp.1–12. Available at: http://dx.doi.org/10.1038/srep41011.

Ma'Sum, M.A., Jatmiko, W. and Suhartanto, H., (2017) Enhanced tele ECG system using Hadoop framework to deal with big data processing. *2016 International Workshop on Big Data and Information Security, IWBIS 2016*, pp.121–126.

Martis, R.J., Acharya, U.R. and Min, L.C., (2013) ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform. *Biomedical Signal Processing and Control*, [online] 85, pp.437–448. Available at: http://dx.doi.org/10.1016/j.bspc.2013.01.005.

Mateo, J., Torres, A.M., Aparicio, A. and Santos, J.L., (2016) An efficient method for ECG beat classification and correction of ectopic beats. *Computers and Electrical Engineering*, [online] 53, pp.219–229. Available at: http://dx.doi.org/10.1016/j.compeleceng.2015.12.015.

Moavenian, M. and Khorrami, H., (2010) A qualitative comparison of Artificial Neural Networks and Support Vector Machines in ECG arrhythmias classification. *Expert Systems with Applications*, [online] 374, pp.3088–3093. Available at: http://dx.doi.org/10.1016/j.eswa.2009.09.021.

Moody, G.B. and Mark, R.G., (2001) The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 203, pp.45–50.

Naderi, R., Yazdanpanah, M.J., Azemi, A. and Roaia, B.N., (2010) Tracking normal action potential based on the FHN model using adaptive feedback linearization technique. *Proceedings of the IEEE International Conference on Control Applications*, pp.1458–1463.

Özbay, Y. and Karlik, B., (2001) A recognition of ECG arrhythmias using artificial neural networks. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 2, pp.1680–1683.

Papadonikolakis, M. and Bouganis, C.S., (2012) Novel cascade FPGA accelerator for support vector machines classification. *IEEE Transactions on Neural Networks and Learning Systems*, 237, pp.1040–1052.

Papechen, P. and Sebastian, S., (2016) A New RS Based Encryption Scheme for Wireless Communication. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, 47, pp.2321–2004.

Park, K.S., Cho, B.H., Lee, D.H., Song, S.H., Lee, J.S., Chee, Y.J., Kim, I.Y. and Kim, S.I., (2008) Hierarchical support vector machine based heartbeat classification using higher order statistics and hermite basis function. *Computers in Cardiology*, 35, pp.229–232.

Phung, V.H. and Rhee, E.J., (2019) A High-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets. *Applied Sciences (Switzerland)*, 921.

Prabhu, V., Kuppusamy, P.G., Karthikeyan, A. and Sucharitha, M., (2020) A Novel Approach for non-invasive measurement of mean arterial pressure using pulse transit time Content courtesy of Springer Nature , terms of use apply . Rights reserved . Content courtesy of Springer Nature , terms of use apply . Rights reserved . pp.3775–3789.

Reza, M., (2020) Galaxy Morphology Classification using Automated Machine Learning, August.

Sampson, M. and McGrath, A., (2015) Understanding the ECG. Part 1: Anatomy and physiology. *British Journal of Cardiac Nursing*, 1011, pp.548–554.

Singh, S., Pandey, S.K., Pawar, U. and Ram, R., (2018) ScienceDirect Classification of ECG Arrhythmia Intelligence using Recurrent Neural Networks Classification of a ECG Arrhythmia using Recurrent Neural Networks Classification of Recurrent Neural Networks a ECG Arrhythmia using. *Procedia Computer Science*, [online] 132Iccids, pp.1290–1297. Available at: https://doi.org/10.1016/j.procs.2018.05.045.

Singh, Y.N., Singh, S.K. and Ray, A.K., (2012) Bioelectrical Signals as Emerging Biometrics: Issues and Challenges. *ISRN Signal Processing*, 2012, pp.1–13.

Song, M.H., Lee, J., Cho, S.P., Lee, K.J. and Yoo, S.K., (2005) Support vector machine based arrhythmia classification using reduced features. *International Journal of Control, Automation and Systems*, 34, pp.571–579.

Sze, V., Chen, Y.H., Yang, T.J. and Emer, J.S., (2017) Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proceedings of the IEEE*, 10512, pp.2295–2329.

Ullah, A., Rehman, S.U., Tu, S., Mehmood, R.M., Fawad and Ehatisham-Ul-haq, M., (2021) A hybrid deep CNN model for abnormal arrhythmia detection based on cardiac ECG signal. *Sensors (Switzerland)*, 213, pp.1–13.

Yang, J., Gong, J., Tang, W., Shen, Y., Liu, C. and Gao, J., (2019) Delineation of urban growth boundaries using a patch-based cellular automata model under multiple spatial and socio-economic scenarios. *Sustainability (Switzerland)*, 1121, pp.1–27.

Ye, C., Vijaya Kumar, B.V.K. and Coimbra, M.T., (2012) Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Transactions on Biomedical Engineering*, 5910, pp.2930–2941.

Zhang, D., Chen, Y., Chen, Y., Ye, S., Cai, W. and Chen, M., (2021) An ECG Heartbeat Classification Method Based on Deep Convolutional Neural Network. *Journal of Healthcare Engineering*, 2021.

APPENDIX A

RESEARCH PROPOSAL

USE OF NEURAL NETWORK ALONG WITH OTHER CLASSIC MACHINE LEARNING ALGORITHMS FOR ECG CLASSIFICATION

Student id no.: 980484

Under the supervision of

Prof. Rituparna Dutta

Research Proposal

Liverpool John Moore's University – Master's in Data Science

November 2021

Table of Contents

Abstract

Electrocardiogram (ECG) is widely used instrument to monitor the diagnoses of cardiovascular system. It takes a highly skilled experts with a lot of experience to diagnose the ECG signal which consumes a lot of time as well as medical resources. The problem of accurately classifying the ECG signal is very challenging. The results are limited because of shallow feature learning architectures. The problem also relies on looking for the most appropriate features for classifying these ECG signals appropriately. Therefore, Machine learning (ML) solutions as well as Artificial Neural Networks are prevalent to analyze and classify ECG data. Machine learning Algorithms are applied to build a model and the performance of these different models can be differentiated based on evaluation metrics. Artificial Neural Networks learn tasks by examples and composed of stacked transformation. It has recently been popular due to its success in variety of task and therefore, it has lot of potential in improving various clinical practices. It has been shown by various studies that with enough data, ANN can show the accuracy to the level of human-expert cardiologist. In this research, different ML models like KNN, Logistic Regression, Random Forest, etc. will be trained and their performance will be presented based on evaluation metrics. In our paper, we will also discuss about the problems encountered in ECG classification, ANNs classifiers, feature extraction tools, pre-processing techniques, ECG database and measures to evaluate the performance. Furthermore, this paper will represent the limitation, performance, and details of each of the classifier and present the detail analysis of best possible classifier based of evaluation index.

List of Abbreviations

Abbreviation	Expansion
AUC	Area under the ROC Curve
ROC	Receiver Operating Characteristic curve
EDA	Exploratory Data Analysis
GPU	Graphics Processing Unit
KNN	K-Nearest Neighbors
MAPE	Mean Absolute Percentage Error
PVE	Percentage of Variance Explained
MAE	Mean Absolute Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SES	Simple Exponential Smoothing
CNN	Convolutional Neural Network
MLPNN	Multiple layer perceptron Neural Network

1) Background

Classification of ECG signals has vital importance in heart disease diagnosis, however, there are lot of problems associated with it. One of the problems is the way by which normal ECG seems different for each individual person and the same disease may also shows different signs on ECG signals of different patients. Likewise, different disease can show similar effects on signals of normal ECG. The classification of ECG signal is a multiple class problem which consists of 5 classes as Normal beats (N), Supraventricular beats (S), Ventricular beats (V), Fusion beats (F) and Unknown (Q). Each ECG signal has several beats where each beat consists of P zone, QRS complex and T zone. The features of ECG signal are peaks which are labelled as P, Q, R, S, T, U and intervals represented PR, RR, QRS, ST, and QT and segments labelled as PR and ST. The below figure represents those features, and these features have their normal duration values.

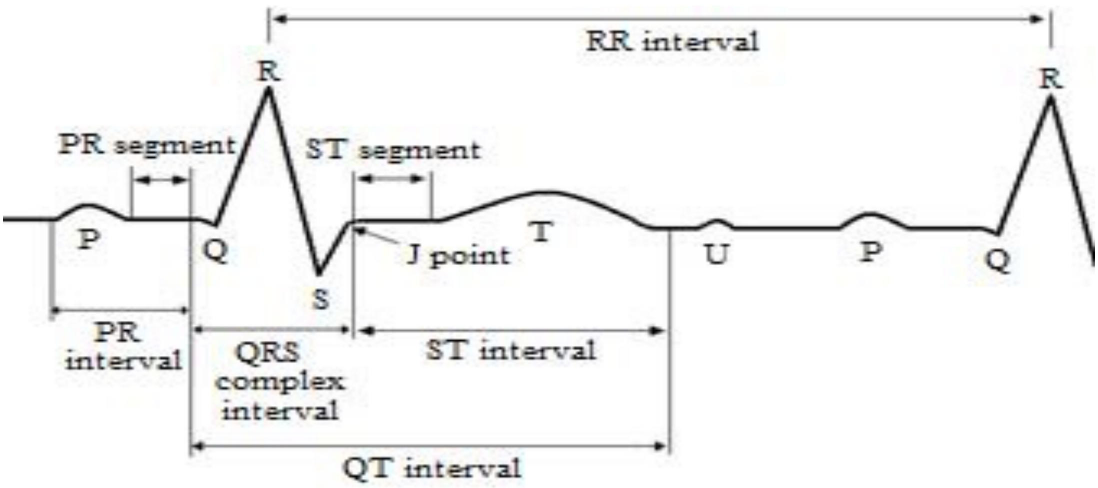


Fig. 1. Normal ECG waveform (E. et al., 2016)

Researchers have employed various preprocessing techniques, feature extraction techniques i.e., Discrete Wavelet Transform (DWT), Continuous Wavelet Transform (CWT), Discrete Fourier Transform (DFT), Principal Component Analysis, etc. For classification, techniques such as Support Vector Machine (SVM) and Multilayer Perceptron Neural Network (MLPNN) has been applied. Researchers (Dallali et al., 2011) have used DWT as a pre-processing technique for extracting RR interval and normalized them using Z score. Likewise, A. Daamouche et al. (Daamouche et al., 2012) have used SVM classifier as modeling technique for ECG classification. However, there are various challenges involved while classifying ECG signals which consists of standardization problem, feature variation, uniqueness in ECG patterns, lack of optimal models and tools for ECG classification, beats variation of single ECG, etc. (Singh et al., 2012).

The purpose of this study is to use Neural Networks such as CNN and MLPNN for ECG classification and compare it with Machine learning models to predict which classification techniques can improve the ECG diagnoses. Furthermore, this work will show the detailed analysis while selecting the input beats and presenting the output of the classifiers. This work can be fruitful for the beginners of ECG classification to know the limitations and advantages existing on different approaches and motivates to look further for better classifier.

2) Related Works

There are number of works associated with classification of ECG dataset based on big data tools as well as those using without big data tools for small size datasets.

To minimize the mortality rate in Indonesia caused by cardiovascular disease, tele-ecg system was invented and monitored using Hadoop framework (Ma'Sum et al., 2017). This system was the first real system for classification of heartbeats which used decision tree (DT) as well as random forest (RF). The server handled 60 requests and the accuracy was 97.14% in case of decision tree while 98.92% for random forest classifier model.

Similarly, Spark and Hidden Markov Models were used to speed up the analysis of ECG signals. This paper showed the potential for classification of heart beats by developing a fast classifier (Brien, n.d.).

SVM along with Neural Network were used to classify different type of heart beat and each type has its own model giving an accuracy of more than 90%, however, the need of unified model for classifying all these types together still remained (Li et al., 2017).

DNN (Deep Neural Network) was applied showing an accuracy of 99%, however, only limited to classification of two different types (Normal / Abnormal) with a dataset consisting of almost 85,000 records (Celesti et al., 2017).

Multitype classification has been studied by the use of Convolutional Neural Network (CNN) showing an accuracy around 93.4% (Kachuee et al., 2018).

In addition to this, the below table shows the brief description of related works in the field of ECG classification.

Table 1. Related Work of ECG Classification

Author	**Year**	**Feature Extraction**	**Classifier Model**	**Classes**	**Accuracy**
Park et al. (Park et al., 2008)	2008	HOS	Hierarchical SVM	5	86%
Soria et al. (Llamedo Soria and Martínez, 2009)	2009	Morphologic and RR-intervals	Weighted LD	5	90%
Llamedo et al. (Llamedo and Mart, 2011)	2011	RR-interval and Morphological features	Weighted LD	5	93%
De Lannoy et al. (De Lannoy et al., 2012)	2012	Morphology, HOS and RR-intervals	Weighted CRF	5	85%
Huang et al. (Huang et al., 2014)	2014	RR intervals and random projection	Ensemble of SVM	5	94%
Zhang et al. (Zhang et al., 2014)	2014	Inter-beat and intra-beat intervals, amplitude morphology, area morphology, and morphological distance	Combined SVM	5	87%
Hongquiang (Li et al., 2016)	2016	Combination of ApEn and WPD	SVM	5	97%
Honquiang (Li et al., 2017)	2017	Combination of PCA and KICA	SVM	5	97%

Zubair et al. (Zubair et al., 2016)	2016	Raw data	1D-CNN	5	93%
Acharya et al. (Acharya et al., 2017)	2017	Raw data	1D-CNN	5	94%
Plawiak (Pławiak, 2018b)	2018	Frequency components of the power spectral density of ECG signals	Evolutionary Neural system (based on single SVM)	17	90%
Plawiak (Pławiak, 2018a)	2018	Frequency components the power spectral density of ECG signals	Genetic ensemble of SVM classifiers	17	91%
Yildirim et al. (Yıldırım et al., 2018)	2018	Rescaling raw data	1D-CNN	17	91%
Plawiak et al. (Pławiak and Acharya, 2020)	2020	Frequency components of the power spectral density of ECG signals	Deep genetic ensemble of classifiers (DGEC)	17	95%

3) Research Questions

This research aims at finding out the most suitable classification techniques for ECG classification based on database provided by the Massachusetts Institute of Technology (MIT-BIH database) having international standards and annotated information by multiple experts. Based on literature review, some of the underlying questions in the field of ECG classification are mentioned below:

- What are the possible influencing factors for ECG classification? What kind of preprocessing techniques and feature recognition results in better performance?
- What are the best performing models based on evaluation metrics for ECG classification?
- What are the scopes, opportunities, and limitations for future works in ECG classification?

4) Aim and Objectives

This research work is about understanding the patterns in ECG signals and studying different underlying ML models along with Neural networks to accurately classify the ECG signals. This work can be helpful for the early and accurate detection of arrythmia types to detect heart disease and choosing an appropriate treatment.

The research objectives are outlined below:

- To review the research work done in the field of ECG classification
- To utilize the feature engineering techniques to find the trend and correlation between different features.
- To apply ML models and Neural networks for ECG classification.
- To identify the best performing model for ECG classification.
- To illustrate the performance of different models based on evaluation index and describe the limitations existing on each of them.

5) Expected Outcomes (Significance and Scope)

This study covers a broad range of machine learning models and neural network in Arrhythmia classification. This research aims at presenting an overview of arrhythmia from medical perspective, ECG databases and evaluation metrics of ECG classifiers. This research can explore the factors responsible for arrythmia diagnoses and lead as a pathway to come up with models for better classification in future. This research will give the detailed analysis of machine learning techniques such as KNN, Logistic Regression, etc. along with Neural Networks such as Convolutional Neural Network (CNN) to classify the ECG signals for Arrhythmia diagnoses. This work will describe about the factors responsible for the difficulties encountered in ECG classification and the possible efficient way for better classification of ECG signals. In a nutshell, it will present a rigorous analysis of different issues encountered in ECG classification, available databases, various preprocessing techniques, various ML models and ANN for better classification, and performance measures based on evaluation metrics.

For this research, we will apply different types of classification algorithms such as KNN, XGBoost and Random Forest (RF) for accurate prediction of the classes of ECG signals and applying the evaluation metrics like ROC curve and Accuracy to differentiate the performance of each model. Moreover, this research work also presents the artificial neural networks such CNN and MLPNN for ECG classification and compare it with classic machine learning algorithms giving a detailed description about each of these techniques. The evaluation of such models will also be based on measures such as Accuracy, F1-score, Roc curve, etc. The performance of each model will be presented in a table to compare their effectiveness, accuracy, and limitations.

6) Research Methodology

6.1) Introduction

Electrocardiogram (ECG) has been used widely since long time to monitor the condition of cardiovascular system. It measures the activity of our heart by displaying data on a paper and helps medical practitioner to detect abnormalities in our heart rhythm as well as other cardiac abnormalities. The shape of heart rhythm differs from a normal to abnormal one. An early detection of heart disease is very important due to high mortality rate of heart disease. It takes many years of experience for a cardiologist to detect abnormalities in ECG graphs and classify signal into different categories. In this scenario, machine learning techniques can be effective to diagnose efficiently and quickly for doctors. However, classification of ECG signals comes with lot of challenges because of several issues like standardization problem of ECG features, uniqueness of ECG patterns, varying ECG waveforms in patients, etc. To build suitable classifier to classify arrythmia on real-time is also one of the issues. We need right tools and techniques to perform complex signal analysis and prediction for which machine learning algorithms and artificial neural networks plays an important role. This research will work on MIT-BIH arrythmia database as a dataset and uses traditional ML algorithms along with neural networks

such as MLPNN and CNN to classify the dataset and present the performance of each of these models to identify the best classifier. The research work is divided into number of steps which are explained below:

1. Cleaning and Understanding Dataset

Understanding a dataset is a very crucial step in machine learning research which involves collecting data from source and knowing about its features. In our case, we have MIT-BIH arrythmia database which is publicly available where each records indicate the heartbeat of 47 individuals. We can apply data cleaning on this dataset by treating the missing values, dropping redundant columns, selecting appropriate features, etc.

2. Exploratory Data Analysis

Exploratory Data Analysis (EDA) will be applied that consists of multivariate analysis. This can give an idea of outliers and correlation existing between different features in dataset. We can construct pie charts, line graph, histogram, heat maps, etc. for visualization. This can further help in visual inspection of dataset and gives more clear insights about dataset.

3. Data pre-processing

Different pre-processing techniques like Principal Component Analysis (PCA) can be applied to dataset to change the data into correct format so that we can utilize this dataset for our model building. This step will also remove the missing values, or any sort of abnormalities present in a dataset.

4. Build and Tune Models

For this research, we will use ML algorithms such as KNN, Logistic Regression, etc. to build models and compare it with ANN. We will look for the best performing model by means of hyperparameter tuning and keep on improving the model performance.

5. Model Evaluation

Different types of evaluation metrics can be applied to evaluate the models on the basis of their performance. Most common types of evaluation metrics are Accuracy, Sensitivity/Specificity, Precision/Recall, F1 score, ROC/AUC Curve metric. The type of evaluation metric to be used depend on the kind of problem we solve. Here, we will come up with the table to show and compare the different models based on their performance.

6. Results and Conclusion

We will take the results of all the models and evaluate them together to indicate the best performing model and compare the results with actual statistics. The limitations of each model will be discussed, and description of the effective model will be presented considering the future scope of work that can lead to more effective results.

6.2) Dataset Description

This research paper uses MIT-BIH arrythmia database which consists of data gathered from 47 subjects and experimented by Arrythmia Laboratory of Boston's Beth Israel Hospital from year 1975 to 1979. It contains 48 half-hour excerpts of two-channel ambulatory ECG recordings. Each test consists of two 11 beat resolution ECG lead signals having range of 10 mv which are digitalized at 360 samples per second. These annotations were grouped into 5 separate categories in collaboration with Association for the Advancement of Medical Instrumentation (AAMI) EC57 standard (Moody and Mark, 2001).

The mapping between different categories and their descriptions are given in table below.

Category	Descriptions	Annotations
N	Nonectopic beat	Normal Left bundle branch block Right bundle branch block beat Atrial escape Nodal (junctional) escape (NE)
S	Supraventricular ectopic beat	Atrial premature (AP) Aberrated atrial premature (aAP) Nodal premature (NP) Supraventricular premature (SP)
V	Ventricular ectopic beat	Premature ventricular contraction (PVC) Ventricular escape (VE)
F	Fusion beat	Fusion of ventricular and normal (fVN)
Q	Unknown beat	Paced (P) Fusion of paced and normal (fPN) Unclassifiable (U)

Table 2. Mapping between different categories and description (Zhang et al., 2021).

6.3) Data pre-processing

The are many features extraction and selection techniques used by different researchers in ECG classification. To mention, some of the feature extraction techniques used are CWT, DWT, Pan-Tompkins algorithm, etc. (Korürek and Doğan, 2010). We can use many pre-processing tools like Outlier Treatment, up-sampling, down-sampling, Normalization, Feature Scaling, Train-Test Split, etc.

6.4) Models

For this research, the classic Machine Learning Algorithms we are using are Logistic Regression, K Nearest Neighbor (KNN), Random Forest and XGboost. Moreover, we will analyze our dataset using Artificial Neural Networks as well. We will use Multiple-layer perceptron neural network (MLPNN) and Convolutional Neural Network (CNN). These models are widely used in any industry or research problems. Although the choice of algorithms is as mentioned, we can make a use of more advanced algorithms with the progress in our research. We can expect for these models to perform well in our research based on literature review.

6.5) Evaluation Index

The evaluation of this research will be based on tools like Accuracy, F1 Score, Precision/Recall, Sensitivity/Specificity and ROC/AUC curve metric. These evaluation metrics are selected based on literature review. There are many other research which has used these metrics to evaluate their models and they are cited in Related Work section.

7) Required Resources

The resources (hardware and software) required for this research are given below:

Hardware:

- Laptop- Mac/Windows
- Processor- 1.8 GHz Intel Core i5
- GPU – Intel Integrated Graphics Card or equivalent
- Memory – 6 GB or higher
- Operating System – macOS Mojave or Windows-10
- Kaggle Kernel and Google Colab
- Google Drive
- Github

Software:

- Anaconda
- Python and Jupyter Notebook
- Numpy, Pandas

- Matplotlib and Seaborn for visualization
- Scikit-learn, Scipy
- Keras and Tensorflow

8) Research Plan

The whole project plan with their tentative schedule is presented in Gantt chart below. We will first go for the problem analysis up to a week and then have detailed literature review for next 3 weeks. The preparation of research proposal will take around 2 weeks. Then, we will work with dataset starting with data understanding and cleaning and then exploratory data analysis for next 4 weeks. The mid-thesis report preparation will then take around 2 weeks. Furthermore, working with different models with hyperparameter optimization will then take around 4 weeks. We will do model evaluation for a week of our model and then finalize report submission within next 4 weeks.

USE OF NEURAL NETWORK ALONG WITH OTHER CLASSIC MACHINE LEARNING ALGORITHMS FOR ECG CLASSIFICATION

Period Highlights: Plan Duration % Complete

Activity	Plan start	Plan Duration	Actual Start	Actual Duration	Percent Complete	PERIODS (Wks) 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Problem Analysis-Review	1	1	1	1	100%	
Detailed Literature Review	2	3	2	2	100%	
Research Proposal Submission	5	2	5	5	100%	
Data Understanding and Cleaning	7	2	7	7	0%	
Exploratory Data Analysis	9	2	9	9	0%	
Data Preprocessing	11	2	11	11	0%	
Mid-Thesis Report Submission	13	2	14	13	0%	
Build and Tune Models	15	4	17	15	0%	
Model Evaluation	19	2	23	19	0%	
Review, Compare and Finalize Results	21	2	25	21	0%	
Create Final Presentation	23	2	27	23	0%	
Final Report Submission	25	1	29	25	0%	

REFERENCES

Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M., Gertych, A. and Tan, R.S., (2017) A deep convolutional neural network model to classify heartbeats. *Computers in Biology and Medicine*, [online] 89August, pp.389–396. Available at: https://doi.org/10.1016/j.compbiomed.2017.08.022.

Brien, J.O., (n.d.) Using Hidden Markov Models and Spark to Mine ECG Data. pp.4–5.
Celesti, F., Celesti, A., Carnevale, L., Galletta, A., Campo, S., Romano, A., Bramanti, P. and Villari, M., (2017) Big data analytics in genomics: The point on Deep Learning solutions. *Proceedings - IEEE Symposium on Computers and Communications*, Iscc, pp.306–309.

Daamouche, A., Hamami, L., Alajlan, N. and Melgani, F., (2012) A wavelet optimization approach for ECG signal classification. *Biomedical Signal Processing and Control*, [online] 74, pp.342–349. Available at: http://dx.doi.org/10.1016/j.bspc.2011.07.001.

Dallali, A., Kachouri, A., Samet, M., Dallali, A., Kachouri, A. and Samet, M., (2011) A Classification of Cardiac Arrhythmia Using WT, HRV, and Fuzzy C-Means Clustering Classification of Cardiac Arrhythmia Using WT, HRV, and Fuzzy C-Means Clustering. *Signal Processing: An International Journal (SPJI)*, [online] 5, p.101. Available at: https://www.researchgate.net/publication/237102471.

E., T., El-Sayed, A. and R., S., (2016) A Survey on Classification of ECG Signal Study. *Communications on Applied Electronics*, 65, pp.11–16.

Huang, H., Liu, J., Zhu, Q., Wang, R. and Hu, G., (2014) A new hierarchical method for inter-patient heartbeat classification using random projections and RR intervals. *BioMedical Engineering Online*, 131, pp.1–26.

Kachuee, M., Fazeli, S. and Sarrafzadeh, M., (2018) ECG heartbeat classification: A deep transferable representation. *Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018*, pp.443–444.

Korürek, M. and Doğan, B., (2010) *ECG beat classification using particle swarm optimization and radial basis function neural network*. *Expert Systems with Applications*, .

De Lannoy, G., François, D., Delbeke, J. and Verleysen, M., (2012) Weighted conditional random fields for supervised interpatient heartbeat classification. *IEEE Transactions on Biomedical Engineering*, 591, pp.241–247.

Li, H., Feng, X., Cao, L., Li, E., Liang, H. and Chen, X., (2016) A New ECG Signal Classification Based on WPD and ApEn Feature Extraction. *Circuits, Systems, and Signal Processing*, 351, pp.339–352.

Li, H., Yuan, D., Ma, X., Cui, D. and Cao, L., (2017) Genetic algorithm for the optimization of features and neural networks in ECG signals classification. *Scientific Reports*, [online] 7January, pp.1–12. Available at: http://dx.doi.org/10.1038/srep41011.

Llamedo, M. and Mart, J.P., (2011) Driven by Database Generalization Criteria. *IEEE Transactions on Biomedical Engineering*, 583, pp.616–625.

Llamedo Soria, M. and Martínez, J.P., (2009) Analysis of multidomain features for ECG classification. *Computers in Cardiology*, 36, pp.561–564.

Ma'Sum, M.A., Jatmiko, W. and Suhartanto, H., (2017) Enhanced tele ECG system using Hadoop framework to deal with big data processing. *2016 International Workshop on Big Data and Information Security, IWBIS 2016*, pp.121–126.

Moody, G.B. and Mark, R.G., (2001) The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 203, pp.45–50.

Park, K.S., Cho, B.H., Lee, D.H., Song, S.H., Lee, J.S., Chee, Y.J., Kim, I.Y. and Kim, S.I., (2008) Hierarchical support vector machine based heartbeat classification using higher order statistics and hermite basis function. *Computers in Cardiology*, 35, pp.229–232.

Pławiak, P., (2018a) Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals. *Swarm and Evolutionary Computation*, [online] 39October 2017, pp.192–208. Available at: http://dx.doi.org/10.1016/j.swevo.2017.10.002.

Pławiak, P., (2018b) Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system. *Expert Systems with Applications*, 92, pp.334–349.

Pławiak, P. and Acharya, U.R., (2020) Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals. *Neural Computing and Applications*, [online] 3215, pp.11137–11161. Available at: https://doi.org/10.1007/s00521-018-03980-2.

Singh, Y.N., Singh, S.K. and Ray, A.K., (2012) Bioelectrical Signals as Emerging Biometrics: Issues and Challenges. *ISRN Signal Processing*, 2012, pp.1–13.

Yıldırım, Ö., Pławiak, P., Tan, R.S. and Acharya, U.R., (2018) Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Computers in Biology and Medicine*, 102August, pp.411–420.

Zhang, D., Chen, Y., Chen, Y., Ye, S., Cai, W. and Chen, M., (2021) An ECG Heartbeat Classification Method Based on Deep Convolutional Neural Network. *Journal of Healthcare Engineering*, 2021.

Zhang, Z., Dong, J., Luo, X., Choi, K.S. and Wu, X., (2014) Heartbeat classification using disease-specific feature selection. *Computers in Biology and Medicine*, 461, pp.79–89.

Zubair, M., Kim, J. and Yoon, C., (2016) An automated ECG beat classification system using convolutional neural networks. *2016 6th International Conference on IT Convergence and Security, ICITCS 2016*.