Dr. D Y PATIL TECHNICAL CAMPUS
Dr. D Y PATIL SCHOOL OF ENGINEERING
Dr. D Y Patil Knowledge City, Charholi Bk, Via Lohegaon, PUNE 412 105

A
SEMINAR REPORT
ON
"Prediction of Cardiac Arrest using Machine Learning"

submitted to the Savitribai Phule Pune University
In partial fulfillment of the requirement for the award of the Degree of
BACHELOR OF ENGINEERING
IN
ELECTRONICS AND TELE COMMUNICATION ENGINEERING
Project Part-I

by

Mr. Prateek Prashant Awate (Exam. Seat No. 71723601D)
Mr. Vavilashetty Venkat Sai Kumar Shrinivas (Exam. Seat No. 71846121F)
Mr. Avinash Baburam Yadav (Exam. Seat No. 71846139J)

Under the guidance of

Prof. N B Surwade
And
Dr Rashmi Mahajan

Department of Electronics and Tele-Communication Engineering
Dr. D Y PATIL SCHOOL OF ENGINEERING
A.Y. 2020-21

Dr. D Y PATIL TECHNICAL CAMPUS
Dr. D Y PATIL SCHOOL OF ENGINEERING
Dr. D Y Patil Knowledge City, Charholi Bk, Via Lohegaon, PUNE 412 105

CERTIFICATE

This is to certify that

Mr. Prateek Prashant Awate -Seat No B150883007
Mr. Vavilashetty Venkat Sai Kumar Shrinivas -Seat No B150883058
Mr. Avinash Baburam Yadav -Seat No B150883060

have satisfactorily completed project part-I and presented report on topic titled "Prediction of Cardiac arrest using Machine Learning" at DR. D Y PATIL SCHOOL OF ENGINEERING in partial fulfillment of requirement of Savitribai Phule Pune University, Pune, for the degree of BE (Electronics and Telecommunication Engineering) in semester VII during academic year 2020-2021.

Guide	Coordinator	H.O.D	Principal
Dr Rashmi Mahajan	Dr Sania Ansari	Dr. S. M. Koli	Dr. Ashok Kasnale

Examiner 1
Name:

Examiner 2
Name:

ABSTRACT

In today's era cardiac arrest and other heart diseases are the most common problem in majority of people, and there are various factors that act as backbone of this problem like people are not paying attention towards health mainly because of work stress, laziness, substandard quality of food that results in increasing cholesterol and untimely diagnosis of heart disease due to lack of technology, methods used in diagnosing these diseases and consequently having a lot of tests. A lot of research and medical supporting systems are developing day by day, however, every system have its various features or advantages and limitations which are unknown to either side. This project aims to study different machine learning algorithms on dataset to be predict the possibility of a cardiac arrest based on various controlled and uncontrolled variables.

List of Tables

List of Figures

Index

CHAPTER 1 INTRODUCTION

The human heart is a finely-tuned instrument that serves the whole body. It is a muscular organ around the size of a closed fist, and it sits in the chest, slightly to the left of the centre.

The heart beats around 100,000 times a day, pumping approximately 8 pints of blood throughout the body 24/7. This delivers oxygen- and nutrient-rich blood to tissues and organs and carries away waste.

1.1 Anatomy of the heart

The heart consists of four chambers:

- The atria: These are the two upper chambers, which receive blood.
- The ventricles: These are the two lower chambers, which discharge blood.

A wall of tissue called the septum separates the left and right atria and the left and right ventricle. Valves separate the atria from the ventricles.
The heart's walls consist of three layers of tissue:

- Myocardium: This is the muscular tissue of the heart.
- Endocardium: This tissue lines the inside of the heart and protects the valves and chambers.
- Pericardium: This is a thin protective coating that surrounds the other parts.
- Epicardium: This protective layer consists mostly of connective tissue and forms the innermost layer of the pericardium.

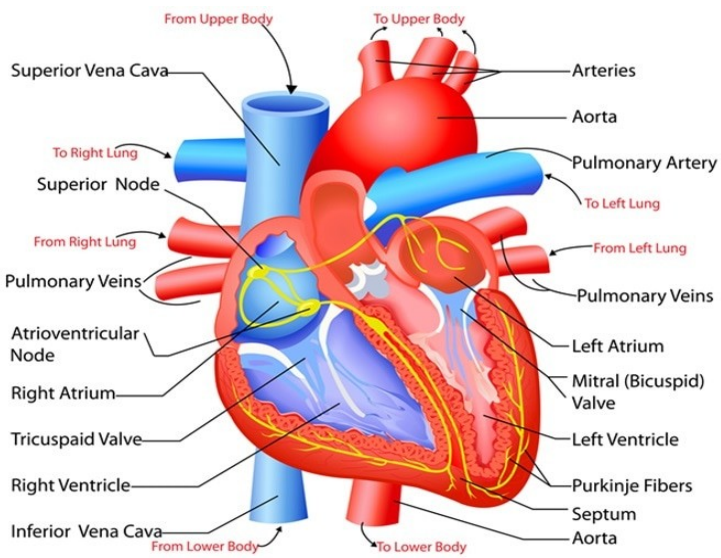


Fig 1.1 Structure of Human Heart

1.2 How the heart works

The rate at which the heart contracts depends on many factors, such as:

- activity and exercise
- emotional factors
- some medical conditions
- a fever
- some medications
- dehydration

At rest, the heart might beat around 60 times each minute. But this can increase to 100 beats per minute (bpm) or more.

1.2.1 Left and right sides

The left and right sides of the heart work in unison. The atria and ventricles contract and relax in turn, producing a rhythmic heartbeat.

1.2.1.1 Right side

The right side of the heart receives deoxygenated blood and sends it to the lungs.

- The right atrium receives deoxygenated blood from the body through veins called the superior and inferior vena cava. These are the largest veins in the body.
- The right atrium contracts and blood passes to the right ventricle.
- Once the right ventricle is full, it contracts and pumps the blood to the lungs via the pulmonary artery. In the lungs, the blood picks up oxygen and offloads carbon dioxide.

1.2.1.2 Left side

The left side of the heart receives blood from the lungs and pumps it to the rest of the body.

- Newly oxygenated blood returns to the left atrium via the pulmonary veins.
- The left atrium contracts, pushing the blood into the left ventricle.
- Once the left ventricle is full, it contracts and pushes the blood back out to the body via the aorta.

1.2.2 Diastole, systole, and blood pressure

Each heartbeat has two parts:

Diastole: The ventricles relax and fill with blood as the atria contract, emptying all blood into the ventricles.
Systole: The ventricles contract and pump blood out of the heart as the atria relax, filling with blood again.

When a person takes their blood pressure, the machine will give a high and a low number. The high number is the systolic blood pressure, and the lower number is the diastolic blood pressure.

Systolic pressure: This shows how much pressure the blood creates against the artery walls during systole.
Diastolic pressure: This shows how much pressure is in the arteries during diastole.

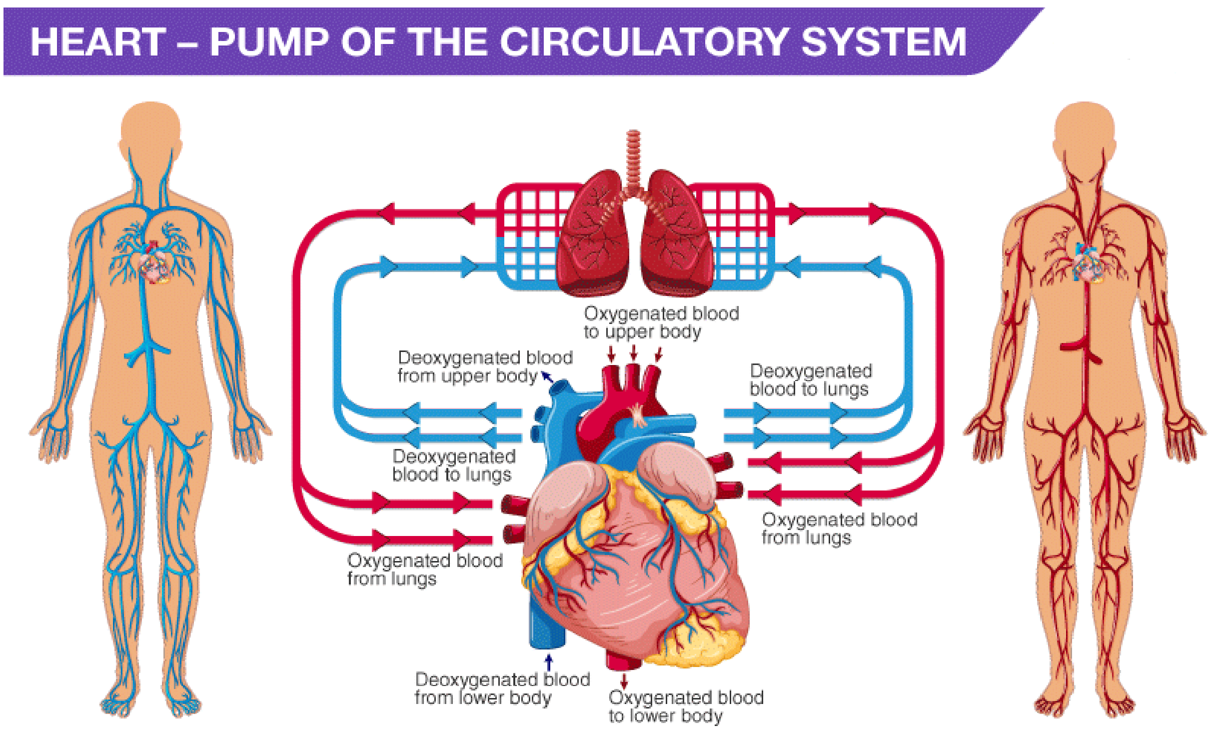


Fig 1.2 Process of pumping of human heart

1.2.3 Gas exchange

When blood travels through the pulmonary artery to the lungs, it passes through tiny capillaries that connect on the surface of the lung's air sacs, called the alveoli.

The body's cells need oxygen to function, and they produce carbon dioxide as a waste product. The heart enables the body to eliminate unwanted carbon dioxide.
Oxygen enters the blood and carbon dioxide leaves it through the capillaries of the alveoli.

The coronary arteries on the surface of the heart supply oxygenated blood to the heart muscle.

1.2.4 Pulse

A person can feel their pulse at points where arteries pass close to the skin's surface, such as on the wrist or neck. The pulse is the same as the heart rate. When you feel your pulse, you feel the rush of blood as the heart pumps it through the body.
A healthy pulse is usually 60–100 bpm, and what is normal can vary from person to person.

A very active person may have a pulse as low as 40 bpm. People with a larger body size tend to have a faster pulse, but it is not usually over 100 bpm.
Learn how to take the pulse here.

1.3 Valves

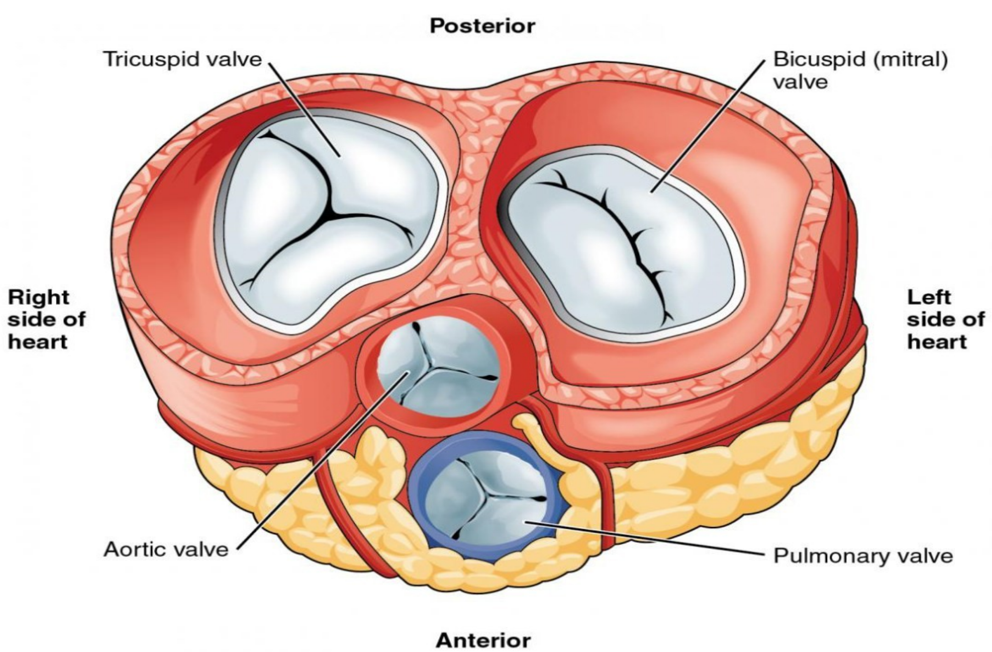


Fig 1.3 Diagram of the Heart's Valves

The heart has four valves to ensure that blood only flows in one direction:

- **Aortic valve:** This is between the left ventricle and the aorta.
- **Mitral valve:** This is between the left atrium and the left ventricle.
- **Pulmonary valve:** This is between the right ventricle and the pulmonary artery.
- **Tricuspid valve:** This is between the right atrium and right ventricle.

Most people are familiar with the sound of the heart. The heart makes many types of sound, and doctors can distinguish these to monitor the health of the heart.

The opening and closing of the valves are key contributors to the sound of the heartbeat. If there is leaking or a blockage of the heart valves, it can create sounds called "murmurs."

1.4 The heart's electrical system

To pump blood throughout the body, the muscles of the heart must work together to squeeze the blood in the right direction, at the right time, and with the right force. Electrical impulses coordinate this activity.

The electrical signal begins at the sino-atrial node, sometimes called the sinus, or SA, node. This is the heart's pacemaker, and it sits at the top of the right atrium. The signal causes the atria to contract, pushing blood down into the ventricles.
The electrical impulse then travels to an area of cells at the bottom of the right atrium, between the atria and ventricles, called the atrioventricular, or AV, node.

These cells act as a gatekeeper. They coordinate the signal so that the atria and ventricles do not contract at the same time. There needs to be a slight delay.
From here, the signal travels along fibres, called Purkinje fibres, within the ventricle walls. The fibres pass the impulse to the heart muscle, causing the ventricles to contract.

1.5 Heart Conditions

- **Stable angina pectoris:** Narrowed coronary arteries cause predictable chest pain or discomfort with exertion. The blockages prevent the heart from receiving the extra oxygen needed for strenuous activity. Symptoms typically get better with rest.

- **Unstable angina pectoris:** Chest pain or discomfort that is new, worsening or occurs at rest. This is an emergency as it can precede a heart attack, serious abnormal heart rhythm, or cardiac arrest.

- **Myocardial infarction (heart attack):** A coronary artery is suddenly blocked. Starved of oxygen, part of the heart muscle dies.

- **Arrhythmia (dysrhythmia):** An abnormal heart rhythm due to changes in the conduction of electrical impulses through the heart. Some arrhythmias are benign, but others are life-threatening.

- **Congestive heart failure:** The heart is either too weak or too stiff to effectively pump blood through the body. Shortness of breath and leg swelling are common symptoms.

- **Cardiomyopathy:** A disease of the heart muscle in which the heart is abnormally enlarged, thickened, and/or stiffened. As a result, the heart's ability to pump blood is weakened.

- **Myocarditis:** Inflammation of the heart muscle, most often due to a viral infection.

- **Pericarditis:** Inflammation of the lining of the heart (pericardium). Viral infections, kidney failure, and autoimmune conditions are common causes.

- **Pericardial effusion:** Fluid between the lining of the heart (pericardium) and the heart itself. Often, this is due to pericarditis.

- **Atrial fibrillation:** Abnormal electrical impulses in the atria causes an irregular heartbeat. Atrial fibrillation is one of the most common arrhythmias.

- **Pulmonary embolism:** Typically, a blood clot travels through the heart to the lungs.

- **Heart valve disease:** There are four heart valves, and each can develop problems. If severe, valve disease can cause congestive heart failure.

- **Heart murmur:** An abnormal sound heard when listening to the heart with a stethoscope. Some heart murmurs are benign; others suggest heart disease.

- **Endocarditis:** Inflammation of the inner lining or heart valves of the heart. Usually, endocarditis is due to a serious infection of the heart valves.

- **Mitral valve prolapses:** The mitral valve is forced backwards slightly after blood has passed through the valve.

- **Sudden cardiac death:** Death caused by a sudden loss of heart function (cardiac arrest).

- **Cardiac arrest:** Sudden loss of heart function.

- **STEMI: Classic or major Heart Attack:** occurs when a coronary artery becomes completely blocked and a large portion of the muscle stops receiving blood. It's a serious heart attack that can cause significant damage. STEMI has a classic symptom of pain in the centre of the chest. This chest discomfort may be described as pressure or tightness rather than sharp pain. Some people who experience STEMIs also describe feeling pain in one or both arms or their back, neck, or jaw.

- **CAS, silent heart attack, or heart attack without blockage:** Coronary Artery Spasm(CAS) is also known as a coronary spasm, unstable angina, or silent heart attack. The symptoms, which can be the same as a STEMI heart attack, may be mistaken for muscle pain, indigestion, and more. It occurs when one of the heart's arteries tightens so much that blood flow stops or becomes drastically reduced. Only imaging and blood test results can tell your doctor if you've had a silent heart attack. There is no permanent damage during a coronary artery spasm. While silent heart attacks aren't as serious, they do increase your risk of another heart attack or one that may be more serious.

- **NSTEMI:** Unlike in a STEMI, the affected coronary artery is only partially blocked in an NSTEMI. An NSTEMI won't show any change in the ST segment on the electrocardiogram. Coronary angiography will show the degree to which the artery is blocked. A blood test will also show elevated troponin protein levels.

1.6 The risk factors for different Heart conditions

- high levels of LDL ("bad") cholesterol
- high blood pressure
- obesity
- sedentary lifestyle
- smoking
- advanced age
- diabetes
- There are also risks associated with gender. For example, until age 55 or so, men are at a higher risk of a heart attack. After menopause, though, women tend to have similar risks as men. Also, men tend to have problems in the heart's larger arteries, while women often experience blockage in the smaller arteries of the heart.
- migraines
- excess thyroid hormone
- chronic allergy conditions
- smoking
- excessive alcohol consumption
- low magnesium levels
- taking drugs for chemotherapy

1.7 Symptoms of Heart Attack

- upper abdomen
- shoulder
- back
- neck/throat
- teeth or jaw
- chest pains or tightness or squeezing
- Sweating more than usual
- Night sweats

- Fatigue
- some women may even think their heart attack symptoms are flu-like symptoms.
- Shortness of breath
- Lightheadedness and dizziness
- Heart palpitations
- Indigestion
- nausea
- vomiting

1.8 Points to be considered for prediction of Heart Attack

- Age
- Diabetes
- Smoking
- Systolic Blood Pressure (SBP)
- Treatment for hypertension
- Total Cholesterol
- High-Density Lipoprotein (HDL) cholesterol
- The difference in BP while resting and other activities
- Fasting blood sugar
- Resting electrographic results
- Maximum heart rate achieved
- Obesity

1.9 Heart Tests

- **Electrocardiogram (ECG or EKG):** A tracing of the heart's electrical activity. Electrocardiograms can help diagnose many heart conditions.

- **Echocardiogram:** An ultrasound of the heart. An echocardiogram provides direct viewing of any problems with the heart muscle's pumping ability and heart valves.

- **Cardiac stress test:** By using a treadmill or medicines, the heart is stimulated to pump to near-maximum capacity. This may identify people with coronary artery disease.

- **Cardiac catheterization:** A catheter is inserted into the femoral artery in the groin and threaded into the coronary arteries. A doctor can then view X-ray images of the coronary arteries or any blockages and perform stenting or other procedures.

- **Holter monitor:** If a doctor suspects an arrhythmia, a portable heart monitor can be worn. Called a Holter monitor, it records the heart's rhythm continuously for 24 hours.

- **Event monitor:** If a doctor suspects an infrequent arrhythmia, a portable heart monitor called an event monitor can be worn. When you develop symptoms, you can push a button to record the heart's electrical rhythm.

1.10 Heart Treatments

- **Exercise:** Regular exercise is important for heart health and most heart conditions. Talk to your doctor before starting an exercise program if you have heart problems.

- **Angioplasty:** During cardiac catheterization, a doctor inflates a balloon inside a narrowed or blocked coronary artery to widen the artery. A stent is often then placed to keep the artery open.

- **Percutaneous coronary intervention (PCI):** Angioplasty is sometimes called a PCI or PTCA (percutaneous transluminal coronary angioplasty) by doctors.

- **Coronary artery stenting:** During cardiac catheterization, a doctor expands a wire metal stent inside a narrowed or blocked coronary artery to open up the area. This lets blood flow better and can abort a heart attack or relieve angina (chest pain).

- **Thrombolysis:** "Clot-busting" drugs injected into the veins can dissolve a blood clot causing a heart attack. Thrombolysis is generally only done if stenting is not possible.

- **Lipid-lowering agents:** Statins and other cholesterol (lipid) lowering drugs reduce the risk for heart attack in high-risk people.

- **Diuretics:** Commonly called water pills, diuretics increase urination and fluid loss. This reduces blood volume, improving symptoms of heart failure.

- **Beta-blockers:** These medicines reduce strain on the heart and lower heart rate. Beta-blockers are prescribed for many heart conditions, including heart failure and arrhythmias.

- **Angiotensin-converting enzyme inhibitors (ACE inhibitors):** These blood pressure medicines also help the heart after some heart attacks or in congestive heart failure.

- **Aspirin:** This powerful medicine helps prevent blood clots (the cause of heart attacks). Most people who have had heart attacks should take aspirin.

- **Clopidogrel (Plavix):** A clot-preventing medicine that prevents platelets from sticking together to form clots. Clopidogrel is especially important for many people who have had stents placed.

- **Antiarrhythmic medications:** Numerous medicines help control the heart's rate and electrical rhythm. This help prevents or controls arrhythmias.

- **AED (automated external defibrillator):** If someone has a sudden cardiac arrest, an AED can be used to assess the heart rhythm and send an electrical shock to the heart if necessary.

- **ICD (Implantable cardioverter-defibrillator):** If a doctor suspects you are at risk for a life-threatening arrhythmia, an implantable cardioverter-defibrillator may be surgically implanted to monitor your heart rhythm and send an electrical shock to the heart if necessary.

- **Pacemaker:** To maintain a stable heart rate, a pacemaker can be implanted. A pacemaker sends electrical signals to the heart when necessary to help it beat properly.

1.11 Facts about Human Heart

- The heart pumps around 5.7 litres of blood a day throughout the body.

- The heart is situated at the centre of the chest and points slightly towards the left.

- On average, the heart beats about 100,000 times a day, i.e., around 3 billion beats in a lifetime.

- The average male heart weighs around 280 to 340 grams (10 to 12 ounces). In females, it weighs around 230 to 280 grams (8 to 10 ounces).

- An adult heart beats about 60 to 100 times per minute, and newborn babies heart beats at a faster pace than an adult which is about 90 to 190 beats per minute.

1.12 Statistics for Cardio Vascular Diseases

- Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year. Four out of 5CVD deaths are due to heart attacks and strokes, and one-third of these deaths occur prematurely in people under 70 years of age.
- Heart diseases have plagued India as it is now known to be the leading cause of death in the country. According to a report by Global Burden of Diseases in 2016, 1.7 million Indians die due to heart disease out of the world 17.3 million deaths

- According to the World Health Organization, 17 million people on an average die due to heart-related diseases every year. Out of the 17 million, 3 million dies of cardiovascular diseases (CVDs) that include stroke and heart attacks.

- Ischemic heart disease and stroke are the predominant causes and are responsible for >80% of CVD deaths. The Global Burden of Disease study estimate of age-standardized CVD death rate of 272 per 100 000 population in India is higher than the global average of 235 per 100 000 population. Premature mortality in terms of years of life lost because of CVD in India increased by 59%, from 23.2 million (1990) to 37 million (2010)

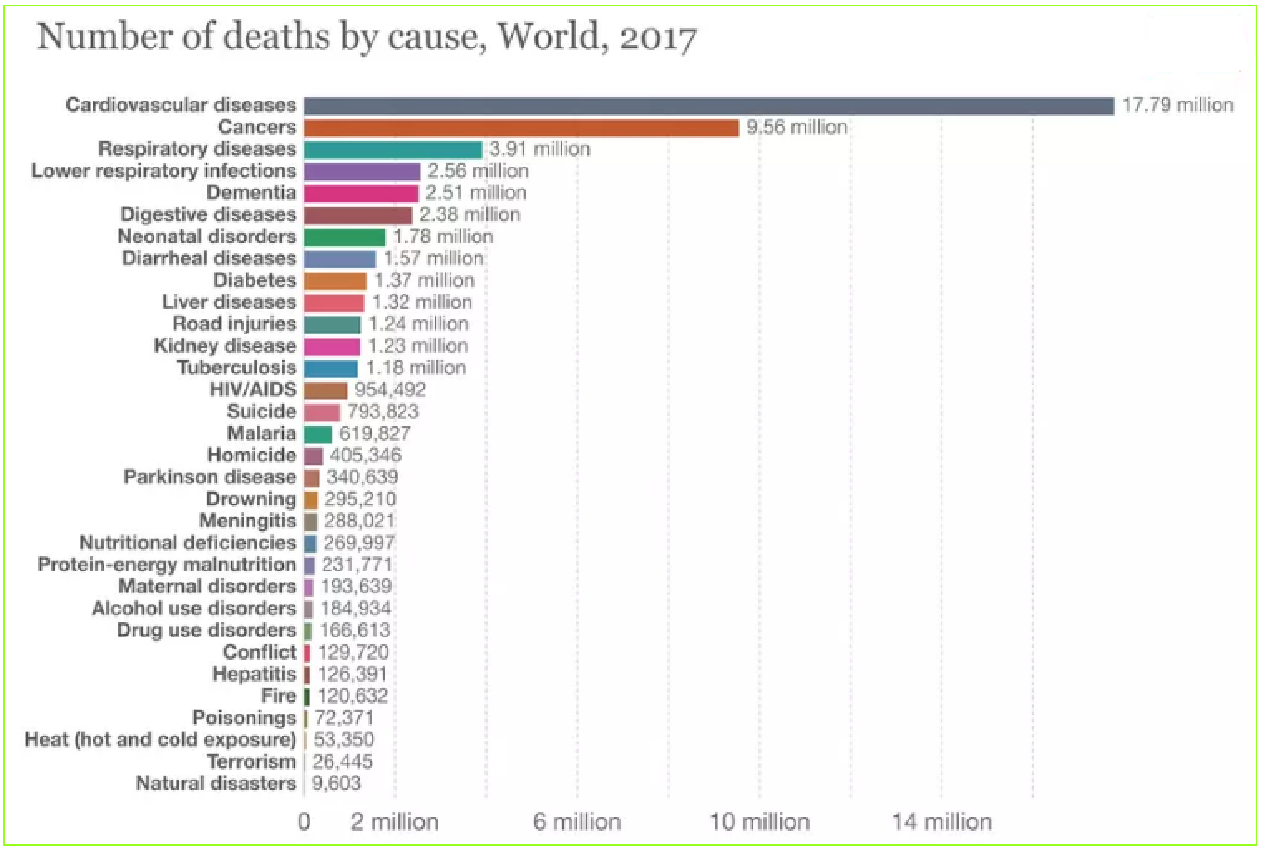


Fig 1.4 Graph showing deaths caused by various factors in year 2017

1.13 Aim

Our project aims to predict whether a person will have cardiac arrest or not using different ML classifiers and multiple datasets.

1.14 Objectives

- Extraction of patient information from the standard dataset

- Patient information includes different attributes such as...
    - Data cleaning and preprocessing on those datasets
    - Using machine learning algorithms to work on those datasets
    - Exploring data analysis
    - Predicting whether the patient will have cardiac arrest or not

1.15 Scope

We have taken multiple datasets and used 5 classifiers to increase their accuracy. We have compared each classifier with all datasets and choose to go on with Support Vector Machine (SVM). Currently, our scope is limited to finding accuracy and saving it but in future, we can expand it to visualizing real-time data.

1.16 Relevance

As our project is software-based so anyone from anywhere can feed their data and try to predict whether he or she is prone to cardiac arrest or not. This can be a game-changer if we can implement it with some fitness watch or after some time when multiple people have checked their score, based on their data we can increase the test accuracy of our machine

1.17 Applications

Our project can be used mainly in the following applications:

- In medical field
- In periodical health checks
- To check the risk of any other medical condition related to the heart.

CHAPTER 2 LITERATURE SURVEY

2.1 Literature survey

An increasing number of heart patients worldwide have motivated researchers to do comprehensive research to reveal hidden patterns in clinical datasets. An overview of reported computational studies on pattern recognition in heart disease is covered in this section. Not only are different techniques addressed, but also various heart disease datasets are provided. Finally, the gap in existing literature, which was the main motivation of this study is also mentioned. Some of the key studies are as follows:

Jyoti Soni, Ujma Ansari, Dipesh Sharma and Sunita Soni: In this journal the authors provides us with a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use in today's medical research particularly in Heart Disease Prediction using KNN, Neural Networks, Bayesian classification, classification based on clustering, Decision Tree. It gives us a total overview of heart attack predictions by using predictive data mining for medical diagnosis. [1]

Shantakumar B. Patil, Dr. Y. S. Kumaraswamy: In this paper the author is giving us a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use in today's medical research particularly in Heart Disease Prediction. It also to compare the performance of predictive data mining technique on the same dataset and the outcome reveals that Decision Tree outperforms and some time Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks. [2]

Anudeep Duba, Rajasekhar Reddy: This informational collection is utilized to anticipate the odds of an event of heart assault for a patient. Its main objective is to bring down the number of attributes by establishing and quantifying the relationships between attributes like Age, Gender, Chest, Pain, Blood, Pressure, Cholesterol etc. [3]

Cheryl Ann Alexander, **Lidong Wang:** The aim of this literature review was to identify usage of Big Data analytics in heart attack prediction and prevention, the use of technologies applicable to big data, privacy concerns for the patient, and challenges and future trends as well as suggestions for further use of these technologies. The national and international databases were examined to identify studies conducted about big data analytics in healthcare, heart attack prediction and prevention, technologies used in big data, and privacy concerns. [4]

Asha Rajkumar, Mrs. G. Sophia Reena: According to author the healthcare industry gathers enormous amounts of heart disease data that regrettably, are not "mined" to determine concealed information for effective decision making by healthcare practitioners. In this paper the data classification is based on supervised machine learning algorithms which result in accuracy, time taken to build the algorithm. Tanagra tool is used to classify the data and the data is evaluated using 10-fold cross validation and the results are compared. Naive Bayes, k-nn, Decision List are also used. [5]

Dr. Hidayet TAKCI: In this study, author used machine learning and feature selection algorithms together. Their aim is to determine the best machine learning method and the best feature selection algorithm to predict heart attacks. According to the author the best machine learning algorithm is the support vector machine algorithm with the linear kernel, while the best feature selection algorithm is the reliefF method. This pair gave the highest accuracy value of 84.81%. [6]

Anil Maybhate, Cheng Chen, Yama Akbari etc. all: In this paper the researchers used controlled laboratory experiment with animal models of CA, their primary focus here is on understanding the functional changes in the thalamus and the cortex, associated with the injury and acute recovery upon resuscitation. Specifically, they studied the changes in thalamocortical synchrony through these periods, they acquired local field potentials (LFPs) from the ventroposterior lateral (VPL) nucleus of the thalamus and the forelimb somatosensory cortex (S1FL) in rats after asphyxial CA. Band-specific relative Hilbert phases were used to analyze synchrony between the LFPs. observed that the CA induced global ischemia changes the local phase-relationships by introducing a phase-lag in both the thalamus and the cortex, while the synchrony between the two regions is nearly completely lost after CA. [7]

Junyun He, Hongyang Lu etc. all: In this paper the authors mentioned about their experiment, a full-field high-resolution optical imaging technique, was used for real-time monitoring of the fluctuations of CBF in a rat model of asphyxial-CA. The temporal changes of CBF were characterized and the relationship between CBF and mean arterial pressure (MAP) was evaluated. Their study provides a new technique to study the neurovascular coupling and metabolic regulation of CBF after CA. [8]

Hsiao-Ko Chang, Cheng-Tse Wu, etc. all: According to the author more than 54% IHCA patient had abnormal clinical manifestation before they suffered a cardiac arrest. If appropriate steps were taken, patients’ survival rate would be higher and medical expense would be decreased. They construct two types of shifting windows (corresponding to two tasks) that allow machine learning to be applied for their dataset which is severely imbalanced. The results show that their approach can effectively handle the imbalanced dataset for detecting cardiac arrest. [9]

Usman Rashed, Muhammad Javed Mirza: This paper introduces work that has been done to distinguish the Electrocardiogram (ECG) of a normal healthy human from that of a patient who may suffer from Sudden Cardiac Death (SCD), but this condition has not been detected. worked on normal portion of SCD ECG and compared its parameters with those of a healthy person’s ECG. The intention is to design an algorithm that may enable doctors to detect chances of myocardial infarction beforehand on the basis of spectral analysis of an ECG. [10]

Keisuke Kasahara, Masahito Shiobara, etc. all: This study examined the feasibility of using indices obtained from a long term Holter ECG record for sudden cardiac arrest (SCA) risk stratification. Patients were classified into high and low risk groups according to their clinical diagnosis, and the obtained indices were compared with those of 25 control subjects. The sensitivities and specificities of all three categories exceeded 0.8 except for the sensitivity to detect the high-risk patient group. Other short-term ECG parameters may need to be incorporated in order to improve the sensitivity. [11]

Archana Singh, Rakesh Kumar: The author uses the biological parameter as testing data such as cholesterol, Blood pressure, sex, age, etc. and on the basis of these calculate accuracy of machine learning algorithms for predicting heart disease, for this, algorithms are k-nearest neighbors, decision tree, linear regression and support vector machine(SVM). On the basis of calculation conclusion is done which one is best among them. [12]

Mohan etc. all: The author define how you can combine two different approaches to make a single approach called hybrid approach which have the accuracy 88.4% which is more than of all other. [13]

Kohali etc. all: The authors have worked on heart diseases prediction using logistic regression, diabetes prediction using support vector machine, breast cancer prediction using AdaBoost classifier and concluded that the logistic regression give the accuracy of 87.1%, support vector machine give the accuracy of 85.71%, AdaBoost classifier give the accuracy up to 98.57% which is good for predication point of view.[14]

Himanshu Sharma etc. all: The author utilizes feature vector and its various data types under various condition for predicating the heart disease, algorithms such as Naïve Bayes, Decision Tree, KNN, Neural Network, are used to predicate risk of heart diseases each algorithm has its specialty such as Naive Bayes used probability for predicating heart disease, whereas decision tree is used to provide classified report for the heart disease, whereas the Neural Network provides opportunities to minimize the error in predication of heart disease. All these techniques are using old patient record for getting predication about new patient. This predication system for heart disease helps doctors to predict heart disease in the early stage of disease resulting in saving millions of lives. [15]

Marjia Sultana etc. all: Authors have illustrated about how the datasets available for heart disease are generally a raw in nature which is highly redundant and inconsistent. There is a need of pre-processing of these data sets; in this phase high dimensional data set is reduced to low data set. They also show that extraction of crucial features from the data set because there is every kind of features. Selection of important features reduces work of training the algorithm and hence resulted in reduction in time complexity. [16]

Amin Ul Haq etc. all: A hybrid intelligent machine-learning based predictive system was proposed for the diagnosis of heart disease. The system was tested on Cleveland heart disease dataset. Seven well-known classifiers such as logistic regression, K-NN, ANN, SVM, NB, DT, and random forest were used with three feature selection algorithms Relief, mRMR, and LASSO used to select the important features. The K-fold cross-validation method was used in the system for validation. In order to check the performance of classifiers, different evaluation metrics were also adopted. The feature selection algorithms select important features that improve the performance of classifiers in terms of classification accuracy, specificity, and sensitivity, MCC and reduced the computation time of algorithms. The classifiers logistic regression with 10-fold cross-validation showed best accuracy 89% when selected by FS algorithm Relief. Due to the good performance of logistic regression with Relief, it is a better predictive system in terms of accuracy. [17]

Ansari etc. all: They performed a work, "Automated Diagnosis of Coronary Heart Disease Using Neuro-Fuzzy Integrated System". In this paper, the author offered a Neuro-fuzzy integrated system for the analysis of heart diseases. To show the effectiveness of the projected system, Simulation for computerized diagnosis is performed by means of the realistic causes of coronary heart disease. The author concluded that this kind of system is suitable for the identification of patients with high/low cardiac risk.[18]

Mrs .G .Subbalakshmi etc. all: They performed a work "Decision Support in Heart Disease Prediction System using Naive Bayes". The main objective of this research is to develop a Decision Support in Heart Disease Prediction System using Naive Bayes algorithm. The system extracts hidden useful information from the heart disease database. This model may possibly answer difficult queries, each one with its own potency with respect to ease of model analysis, access to complete information and accurateness. This model can be further enhanced and expanded by incorporating other data mining techniques.[19]

Mai Shouman, Tim Turner, and Rob Stocker etc. all: They performed a work "Applying k-Nearest neighbors in Diagnosing Heart Disease Patients". In this paper the author details work that applied KNN on a Cleveland Heart Disease dataset to investigate its efficiency in the prediction of heart disease. The author also investigated if the accuracy could be enhanced by integrating voting with KNN. The results show that applying KNN achieved an accuracy of 97.4%. The results also show that applying voting could not enhance the KNN accuracy in the diagnosis of heart disease.[20]

Ashish Kumar Sen, Shamsher Bahadur Patel, Dr. D. P. Shukla etc. all: They performed a work "A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level" .In this work, the author has designed a system which could identify the chances of a coronary heart disease. He has divided all parameters into two levels according to criticality of the parameter and assigned each level a separate weightage. Finally, both the levels are taken into consideration to arrive a final decision. The author has implemented neuro-fuzzy integrated approach at two levels. So, error rate is very low and work efficiency is high. The author concluded that this same approach could be used to perform the analysis on some other diseases also.[21]

Aditi Gavhane etc. all: They suggest to make use of the easily available sensors in watches and cell phones to measure the simple factors. Provides its users with a prediction result that gives the state of a user leading to CAD. Due to the recent advancements in technology, the machine learning algorithms are evolved a lot and hence we use Multi Layered Perceptron (MLP) in the proposed system because of its efficiency and accuracy. Also, the algorithm gives the nearby reliable output based on the input provided by the users. If the number of people using the system increases, then the awareness about their current heart status will be known and the rate of people dying due to heart diseases will reduce eventually. [22]

Jayshril S. Sonawane etc. all: They performed a work "Prediction of Heart Disease Using Multilayer Perceptron Neural Network" The system for prediction for heart disease using multilayer perceptron neural network is implemented in MATLAB R2012. In this system the database is divided in to two sets randomly that is training set and testing set. Out of total records 70% records are used for training and testing is done by using remaining 30% records. The evaluation of performance of the system is done by computing the percentage value of different parameters like Accuracy, Specificity and Sensitivity. [23]

Yubin Park, Joyce C. Ho, Joydeep Ghosh: They performed a work "Multivariate Temporal Symptomatic Characterization of Cardiac Arrest". In this work they modelled the temporal symptomatic characteristics of 171 cardiac arrest patients in Intensive Care Units. The temporal and feature dependencies in the data are illustrated using a mixture of matrix normal distributions. We found that the cardiac arrest temporal signature is best summarized with six hours data prior to cardiac arrest events, and its statistical descriptions are significantly different from the measurements taken in the past two days. This matrix normal model can classify these patterns better than logistic regressions with lagged features.[24]

Ali Bahrami Rad etc. all: They performed work "NEAREST-MANIFOLD CLASSIFICATION APPROACH FOR CARDIAC ARREST RHYTHM INTERPRETATION DURING RESUSCITATION" .In these paper in order to monitor the cardiac arrest patients response to therapy, there is a need for methods that can reliably interpret the different types of cardiac rhythms that can occur during a resuscitation episode. These rhythms can be categorized to five groups; ventricular tachycardia, ventricular fibrillation, pulseless electrical activity, asystole, and pulse generating rhythm. The objective of this study was to develop machine learning algorithms to automatically recognize these rhythms. They proposed a detection algorithm based on the nearest-manifold classification approach using a group of 8-time domain features as statistical measures on the signal itself, as well as the first and second differences. The overall accuracy of the cardiac arrest rhythm interpretation is 79%. The sensitivity/specificity of shockable/non-shockable rhythms is 92/95 %. [25]

N. Suresh etc. all: They performed a work "Improved Performance of FFT Based Cardiac Analyzer Using Advanced Booth Algorithm" .In these paper they have explained about Cardiac monitoring system which is used to analyze the abnormality of ECG signal using Fast Fourier Transform (FFT) and send comments to the receiver using short message service(SMS) through Global System Mobile(GSM), where the Fast Fourier Transform (FFT) used in this has more delay, which takes much time in computation due to the presence of multiplier in it. Thus to increase the speed of FFT the Advanced booth algorithm is proposed here, in which the number of slice, number of 4-input LUTs and number of bonded IOB gets reduced when compare to the other existing FFT speed increment algorithm, Hence it is proved that the proposed FFT analyzer improved the overall performance of existing system. [26]

Enrico G. Caiani, Giulia Marelli, etc. all: In the lambardy region Italy they used Geomatics which is a scientific term for gathering, storing, processing, and delivering geographic information. they hypothesized that approaches based on geomatics applied to CA and AED positions could provide useful information about the distribution of events and the definition of strategies for relocating emergency resources on a specific territory. Accordingly, there aim was to analyze by Health Geomatics methods the 2015 CA and AED database relevant to the Lombardy Region in Italy, and to focus on CA events distribution and AED potential coverage in the territory of the city of Milan, using spatial isochrones, i.e. connected points reached within the same time starting from the same origin. Isochrones are temporal lines that connect points that can be reached in the same temporal interval starting from the same origin. In this work we applied realistic isochrones considering the real path between two points based on the distance traveled along the street (therefore not the trivial and incorrect Euclidean distance), to evaluate the portion of territory effectively covered by the existing AED distribution. The analysis was restricted to the three healthcare districts that refer to the city of Milan, representing the portion of territory with higher population density. [27]

Dr. C. Nataraj, etc. all: The objective of the experiments was to determine if titrating CPR to blood pressure would improve 24-hour survival compared with traditional American Heart Association CPR in a porcine model of ventricular fibrillation and asphyxia-associated ventricular fibrillation. After induction of anesthesia and 7 minutes of untreated VF, 17 female 3-month-old swine received manual cardiopulmonary resuscitation. Similarly, 22 3-month- old female swine received manual cardiopulmonary resuscitation after 7 minutes of untreated asphyxia. Furthermore, it was observed that only the data obtained during the first 2 minutes of the resuscitation period in the interval between t ≈ 7 min and t ≈ 9 min is relevant for the purpose of this study, in order to eliminate potential influences on the data introduced through different CPR sequences after t = 9 min. Among other measurements, electrocardiography (ECG) and heart rate (HR)were recorded during the experiment and are the measurements used and referred to as data in this study. [28]

Jyh-Shing Roger Jang, etc. all: In this study, the raw dataset is collected from the electronic health records (EHRs) of the adult patients (age ʁ20 years) who visited emergency department (ED) and stayed in the emergency detention area for more than 6 hours during January 2014 to December 2015, and it is provided by National Taiwan University Hospital (NTUH).they perform the data preprocessing and cleaning for the dataset using a resampling technique to balance the data amount of CPR and Non-CPR patients of the dataset, and then they construct a sliding window and apply several classifiers for model training and reducing the possible overfitting problem. Additionally, they use the measures such as the Area Under the Receiver Operating Characteristic Curve (AUROC) and the Area Under the Precision-Recall Curve (AUPRC) to comparative evaluate the performance of their models built. [29]

Nan Liu, Zhiping Lin, etc. all: An intelligent prediction model is proposed to compute a risk score on a patient’s clinical outcome, utilizing both HRV parameters and vital signs. The scoring system is built based on the calculation of geometric distances among a set of feature vectors obtained from the records of multiple patients .The geometric distance-based scoring system was used to correlate HRV parameters and vital signs to cardiac arrest within 72 h, and ROC analysis was adopted to investigate prediction performance. In the evaluation results, sensitivity, specificity, PPV, and NPV were reported. [30]

M. Reza Pazhouhandeh, Omid Shoaei, etc. all: Two-electrode Impedance-sensing Cardiac Rhythm Monitor for Charge-Aware Shock Delivery in Cardiac Arrest, two-electrode amplifier with real-time body and electrodes impedance measuring capability. The amplifier employs two sinusoidal current sources to interrogate the electrodes/body impedance. Conventional two electrodes amplifiers are highly susceptible to CMV saturation. A CMFB keeps the CMV of the body constant and avoids saturating amplifier by any mismatch in the injected currents. The current sources are implemented with the popular Howland VCCS structure with one integrated circuit and a buffer in a positive feedback path to operate with only one external resistor. To minimize the sensitivity of the output impedance of the VCCS to intrinsic mismatch of components. [31]

Sombat Muengtaweepongsa, etc. all: In patients, HRV measurement is proposed to be a predictor of 24 hrs. mortality in successfully resuscitated 69 patients without-of hospital cardiac arrest. It has been reported that in non survivors have shown a sudden significant decrease of HRV in rewarming phase compared to those in survivors. The electrocardiogram signals were collected at intensive care unit, Thammasat University Hospital for all phases of study 6, 18 hours during sustainment in mild hypothermia (phase A and B), during rewarming period (phase C) and recover to normothermia (phase D). The 10 minute-lead II ECG with data sampling rate of 1000 Hz was recorded with Power Lab systems (AD Instruments). Then data were stored in personal computer, and RR intervals were extracted by using LabchartPro7 Software (AD Instruments). After extraction of RR intervals, time domain, frequency domain and nonlinear HRV parameters were analyzed by Heart Rate Variability Analysis Software (Kubios-HRV version 2.0). [32]

Utsav Chauhan, Vikas Kumar, etc. all: Artificial intelligence is used to solve cardiac arrest disease by predicting the risk percentage using Machine Learning. In Machine Learning, they used several predictive a classification algorithm such as Random Forest, decision tree, Linear Regression, Support Vector Machines (SVM) and Artificial Neural Networks (ANN). ANN is the most powerful tool in artificial intelligence due to its brain like functioning in which a neuron (also known as perceptron) is the fundamental unit of the network and a number of inputs are fed on one neuron which gives an output .The parameters they used while implementing the algorithm were age, sex, chest pain, rest bp, cholesterol, fbs, Rest ECG, Max HR, Ex Ang, Old peak and Decision. While tidying the data, the dataset used contained some missing values which had to be removed before the algorithms could be applied on it. The missing values in the dataset were removed by filling the spots with the mean values of the column. The dataset was checked for the outliers and none were found while mapping the data. After applying Support Vector Machine, Random Forest, Decision Tree, Logistic Regression and Artificial Neural Network algorithms on the dataset to predict the occurrence of cardiac arrest in patients it is found out that the accuracy of the Artificial Neural Network is the highest (~85 %). Also, since the dataset was limited, the accuracy of the algorithm is low. Had the dataset been bigger the accuracy of ANN would also have been much higher than the current average value of the outcome. Thus, Artificial Neural Network can be used as a basis to predict occurrence of cardiac arrest in the patients according to the results of the algorithms performed on the dataset. [33]

Joyce C. Ho and Yubin Park: they introduced TTL-Reg, a temporal transfer learning-based model to learn a robust cardiac arrest prediction model. The algorithm learns from different time perspectives by smoothing the estimated coefficients of logistic regression from adjacent time points. We show that the parameters of TTL-Reg can be solved iteratively using existing software packages by transforming the objective function into an `2-regularized logistic regression model. Their model not only yields a coefficient trajectory that can be easily interpreted and potentially uncover new trends but also results in improved early prediction of cardiac arrest patients. The results on 763 ICU patients illustrates the potential of their temporal transfer learning approach. [34]

Chih-Wei Sung, Jiann-Shng Shieh, etc. all: machine learning was used to predict Heart Rate Variability for the Detection of Seizures in Comatose Cardiac Arrest Survivors a real-time system predicting seizure events through long-term continuous data recording and extraction was developed. The system displays information on a physiological monitor and extracts real-time bio signals, including ECG and EEG signals. ECG-based HRV analysis can be used to immediately detect seizures in patients who have experienced sudden cardiac arrest and have then been resuscitated using CPR and admitted to the EICU. Four HRV parameters, namely SDNN intervals, HF, LF/HF, and SampEn, can be used to develop a reliable algorithm for seizure detection. Seizures has always been a critical issue in clinical care scenarios, especially for patients who have survived a cardiac arrest, because the long ischemic cerebral perfusion time may cause hypoxic encephalopathy, which deteriorates neurological outcomes. This study shows to overcomes the previous limitations on continuity and accessibility for real-time seizure detection. In this study, seizure events were detected in patients presenting with muscle convulsion through physician assessments and a real-time detection model. Seizure could be also found in anytime when the patients seemed no convulsion because abnormal electrical discharge in these patients only occurred in the cerebral cortex, but not in the limbs or face. [35]

Andoni Elola, Elisabete Aramendi, etc. all: a new multi-modal solution to classify circulation states during OHCA using concurrent information derived from the ECG, the TI and the capnogram. The solution allows the classification into two classes (PR/PEA) or three classes such as pseudo pulseless electrical activity, true pulseless electrical activity and pulse generating rhythms, with the final aim of monitoring the circulation state of the patient and the response to resuscitation treatment. The study is based on a unique dataset that includes IBP signals measured using arterial lines during OHCA to provide an accurate ground truth clinical annotation of the circulation state. in an organized rhythm (QRS complexes), and free of chest compression artefacts. The ECG, TI and capnogram were used to develop the algorithms. A clinician and two expert biomedical engineers used all other sources of information to annotate the circulation state pseudo pulseless electrical activity, true pulseless electrical activity and pulse generating rhythms for each interval, including: clinical patient charts with annotated restore spontaneous circulation intervals, the IBP waveform, and cerebral oxygen saturation when available. Systolic (Sys), diastolic (Dias) and pulse pressure were computed for each cardiac cycle and averaged to be displayed during annotation. The distinction between the three circulation states was possible using the objective values obtained from the IBP because systolic and pulse pressures are higher for pulse generating rhythms than for pulseless electrical activity. [36]

Donald R., AKM Jahangir, A. Majumder, etc. al: An embedded IoT system to monitor heart rates and body temperatures using smartphones. they use an ECG sensor system and the realtime detection of abnormality in users' ECG patterns. The results from sensors' data are also presented to show that this approach provides a high rate of classification correctness in distinguishing between normal and abnormal ECG patterns. The system may also find multiple applications in heart behavior detection for people with various disabilities who are at a high risk of cardiac arrest. The IoT system consists of a pulse sensor, a temperature sensor, an Arduino, and a Low Energy (LE) Bluetooth. A smartphone is used to collect sensors data. They collected data from ECG and temperature sensors using a smartphone. they used multiple subjects and collected data for different events of user daily activity. Data for each subject was collected from a smartphone placed in the subject's hand or a pocket. In the system, the ECG and temperature sensors are used to collect the raw ECG patterns and body temperature while the user is walking. Then, the resultant outputs are processed inside the mobile phone to identify the user's ECG pattern. Though the system continuously monitors for ECG patterns and body temperature, it only triggers a warning if the ECG pattern and body temperature of the user reaches a certain threshold where the user might face a potential heart attack. At that time, the system warns the user with a message and vibration to alert them about an imminent cardiac arrest. [37]

C. Jenefar Sheela and L. Vanitha: the cardiac prediction system is proposed with HRV parameters as input and support vector machines with RBF kernel as classifier. The prediction system has demonstrated its ability to generate human understandable factors, and has shown its effectiveness to being a powerful predictor of cardiac arrest before 30 minutes. The experimental validations on a time domain and frequency domain show that the proposed system is able to achieve satisfactory prediction results. Hybrid structure may be used to improve the classification efficiency. They predicted sudden cardiac arrest before 30 minutes of its occurrence on the basis of time domain and frequency domain features of Heart rate variability (HRV) obtained from ECG and using SVM classifier to classify SCD patient from Normal patient. The database of cardiac patients obtained from physionet is used to check the validity of the proposed work. Performance of SVM is better giving the classification efficiency of 88%. [38]

César O. Navarro, Nick A. Cromie, etc. all: Cardiac Arrest was detected using a Simplified Frequency Analysis of the Impedance Cardiogram recorded from Defibrillator Pads Estimation of the frequency spectrum of the first order derivative of the impedance cardiogram (dZ/dt) recorded through the 2 transthoracic defibrillator pads can be used as a marker of circulatory collapse. The use of less accurate integer filters for the estimation is a feasible solution to be applied in a less powerful CPU operating in a PAD defibrillator. The results provide initial tools for further development of applications for the use of ICG in defibrillators during emergency clinical practice. An in-house fully functional defibrillator was constructed (Samaritan AED; HeartSine Technologies, UK) which apart from the ECG, included the recording of ICG using a low amplitude sinusoidal current (30 kHz; 0.05 mA) between the 2 adhesive defibrillator pads (Samaritan. Its CPU is a Motorola 68336. The ECG and ICG signals were monitored, digitized and stored for retrospective analysis. [39]

2.2 Selected papers form literature (References)

[1] Jyoti Soni, Ujma Ansari, Dipesh Sharma and Sunita Soni, 'Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction', International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March 2011.

[2] Shantakumar B. Patil, Dr. Y. S. Kumaraswamy, 'Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction', International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March 2011.

[3] Anudeep Duba, Rajasekhar Reddy, 'Decision Making System using Machine Learning and Pearson for Heart Attack', International Conference on Electronics, Communication and Aerospace Technology ICECA 2017.

[4] Cheryl Ann Alexander, Lidong Wang, 'Big Data Analytics in Heart Attack Prediction', Article *in* Journal of Nursing & Care · January 2017.

[5] Asha Rajkumar, Mrs. G. Sophia Reena, 'Diagnosis of Heart Disease Using Datamining Algorithm', Vol. 10 Issue 10 Ver. 1.0 September 2010 Global Journal of Computer Science and Technology.

[6] Dr. Hidayet TAKCI, 'Improvement of heart attack prediction by the feature selection methods', Turkish Journal of Electrical Engineering & Computer Sciences, Turk J Elec Eng & Comp Sci (2018).

[7] Anil Maybhate, Cheng Chen, Yama Akbari, David L. Sherman, Kaiquan Shen, Xiaofeng Jia, Nitish V. Thakor, 'Band Specific Changes in Thalamocortical Synchrony in Field Potentials after Cardiac Arrest Induced Global Hypoxia', 35th Annual International Conference of the IEEE EMBS Osaka, Japan, 3 - 7 July, 2013.

[8] Junyun He, Hongyang Lu, Ruoxian Deng, Leanne Young, Shanbao Tong, Xiaofeng Jia, 'Real-time monitoring of cerebral blood flow by laser speckle contrast imaging after cardiac arrest in rat'.

[9] Hsiao-Ko Chang, Cheng-Tse Wu, Ji-Han Liu, Wee Shin Lim, Hui-Chih Wang, Shu-I Chiu, Jyh-Shing Roger Jang, 'Early Detecting In-Hospital Cardiac Arrest Based on Machine Learning on Imbalanced Data'.

[10] Usman Rashed, Muhammad Javed Mirza, ‘Identification of Sudden Cardiac Death Using Spectral Domain Analysis of Electrocardiogram (ECG)’, 2008 International conference on Emerging Technologies IEEE-ICET 2008.

[11] Keisuke Kasahara, Masahito Shiobara, Saya Nakamura, Koichiro Yamashiro, Kazuo Yana Takuya Ono, ‘Sudden Cardiac Arrest Risk Stratification based on 24-hour Holter ECG Statistics’.

[12] Archana Singh Rakesh Kumar "Heart Disease Prediction Using Machine Learning Algorithms" 2020 International Conference on Electrical and Electronics Engineering (ICE3-2020)

[13] Senthil Kumar Mohan, Chandrasekar Thirumalai and Gautam Srivastva, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques” IEEE Access 2019.

[14] Pahulpreet Singh Kohli and Shriya Arora, “Application of Machine Learning in Diseases Prediction”, 4th International Conference on Computing Communication and Automation (ICCCA), 2018.

[15] Himanshu Sharma and M. A. Rizvi, “Prediction of Heart Disease using Machine Learning Algorithms: A Survey” International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 8, IJRITCC August 2017.

[16] M. Sultana, A. Haider, and M. S. Uddin, “Analysis of data mining techniques for heart disease prediction,” 2016 3rd Int. Conf. Elect. Eng. Inf. Comm. Technol. iCEEiCT 2016, 2017.

[17] Amin Ul Haq A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms Hindawi Mobile Information Systems Volume 2018, Article ID 3860146, 21 pages

[18] A. Q. Ansari et. al.," Automated Diagnosis of Coronary Heart Disease Using Neuro-Fuzzy Integrated System", 2011 World Congress on Information and Communication Technologies 978-1-4673-0125-1@ 2011 IEEE (pp 1383-1388)

[19] Mrs. G. Subbalakshmi, Mr. M. Chinna Rao “Decision Support in heart disease prediction system using naïve bays”, IJCSE Indian journal of computer science and engineering, ISSN: 0976-5166 Vol. 2 No. 2 Apr-May 2011

[20] Mai Shouman, Tim Turner, and Rob Stocker, “Applying k-Nearest Neighbor in Diagnosing Heart Disease Patients”, International Journal of Information and Education Technology, Vol. 2, No. 3, June 2012

[21] Ashish Kumar Sen, Shamsher Bahadur Patel, Dr. D. P. Shukla, “A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach

Two Level", International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 2 Issue 9 Sept., 2013 Page No. 2663-2671

[22] Aditi Gavhane et al "Prediction of Heart Disease Using Machine Learning" Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1

[23] Jayshril S. Sonawane, D. R. Patil "Prediction of Heart Disease Using Multilayer Perceptron Neural Network" ICICES2014 - S. A. Engineering College, Chennai, Tamil Nadu, India ISBN No.978-1-4799-3834-6/14/$31.00©2014 IEEE

[24] Yubin Park and Joyce C. Ho and Joydeep Ghosh3 "Multivariate Temporal Symptomatic Characterization of Cardiac Arrest" 35th Annual International Conference of the IEEE EMBS Osaka, Japan, 3 - 7 July, 2013

[25] Ali Bahrami Rad et al "NEAREST-MANIFOLD CLASSIFICATION APPROACH FOR CARDIAC ARREST RHYTHM INTERPRETATION DURING RESUSCITATION" 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)

[26] N. Suresh "Improved Performance of FFT Based Cardiac Analyzer Using Advanced Booth Algorithm" International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB16)

[27] Enrico G. Caiani, Giulia Marelli, Maria A. Brovelli Politecnico di Milano Milan, Italy, Carolina Arias Muñoz Universidad de San Buenaventura, Pietro M Brambilla, Andrea Pagliosa, Giovanni Cesana Azienda Regionale Emergenza Urgenza – AREU Milan Italy: 'Development of a Health Geomatics analysis framework to evaluate cardiac arrests in Lombardy', January 2017

[28] Dieter Bender, Ph.D. Candidate- Villanova University, Dr. Ryan W. Morgan-Attending Physician of Critical Care Medicine, Children's Hospital of Philadelphia, Dr. Vinay M. Nadkarni-Endowed Chair of Critical Care Medicine, The Children's Hospital of Philadelphia and Dr. C. Nataraj, Moritz Endowed Professor, Villanova University: "Classification of Asphyxia & Ventricular Fibrillation Induced Cardiac Arrest For Cardiopulmonary Resuscitation "year 2017.

[29] Ji-Han Liu- Graduate Institute of Networking and Multimedia National Taiwan University Taipei, Taiwan, Hui-Chih Wang - Department of Emergency Medicine National Taiwan University Hospital Taipei, Taiwan, Hsiao-Ko Chang, Cheng-Tse Wu, Wee Shin Lim, Jyh-Shing Roger Jang -Department of Computer Science and Information Engineering National Taiwan University Taipei, Taiwan, 'Machine Learning Based Early Detection System of Cardiac Arrest', year 2019.

[30] Nan Liu, Zhiping Lin, Senior Member, IEEE, Jiuwen Cao, Zhixiong Koh, Tongtong Zhang, Guang-Bin Huang, Senior Member, IEEE, Wee Ser, Senior Member, IEEE, and

Marcus Eng Hock Ong, “An Intelligent Scoring System and Its Application to Cardiac Arrest Prediction “year 2012

[31] M. Reza Pazhouhandeh, Omid Shoaei, Roman Genov -Intelligent Sensory Microsystems Laboratory, ECE Department of University of Toronto, Toronto, Canada 2 Bio-Integrated Systems Laboratory, ECE Department of University of Tehran, Tehran, Iran- “Two-electrode Impedance-sensing Cardiac Rhythm Monitor for Charge-Aware Shock Delivery in Cardiac Arrest “year 2017.

[32] Sombat Muengtaweepongsa MD, Kesorn Suwanprasert Ph.D. Faculty of Medicine, Thammasat University (Rangsit Campus) Pathumthani, Thailand, Anon Jantanukul Medical engineering graduate program Thammasat University (Rangsit Campus) Pathumthani, Thailand “Suppression of Autonomic Drive Determined by Nonlinear HRV analysis in Therapeutic Hypothermia after Cardiac Arrest “year-2013 Biomedical Engineering International Conference.

[33] Vipul Chauhan, Amit Kumar, Sumit Tiwary, Vikas Kumar HM ,Utsav Chauhan HMR Institute of Technology and Management Delhi, India ,”Cardiac Arrest Prediction using Machine Learning Algorithms “ year-2019.

[34] Joyce C. Ho and Yubin Park,” Learning from Different Perspectives: Robust Cardiac Arrest Prediction via Temporal Transfer Learning “year -2017

[35]CHIH-WEI SUNG , JIANN-SHING SHIEH , WEI-TIEN CHANG, YI-WEI LEE , JI-HUAN LYU1 , HOOI-NEE ONG , WEI-TING CHEN, CHIEN-HUA HUANG , WEN-JONE CHEN , AND FU-SHAN JAW “Machine Learning Analysis of Heart Rate Variability for the Detection of Seizures in Comatose Cardiac Arrest Survivors “ year - July 30, 2020.

[36] Andoni Elola, Elisabete Aramendi, Member IEEE, Unai Irusta, Member IEEE, Per Olav Berve, Lars Wik “Multimodal algorithms for the classification of circulation states during out-of-hospital cardiac arrest “year-2020.

[37] Yosuf ElSaadany, AKM Jahangir A. Majumder, Donald R. Ucci Department of Electrical and Computer Engineering, Miami University, Oxford, OH, USA “A Wireless Early Prediction System of Cardiac Arrest through IoT “year -2017 IEEE 41st Annual Computer Software and Applications Conference “.

[38] C. Jenefar Sheela .PG Scholar- Loyola Institute of Technology Palanchur, Chennai, L. Vanitha. Assistant Professor -Loyola Institute of Technology, Palanchur, Chennai. “Prediction of Sudden Cardiac Death Using Support Vector Machine “year -2014 International Conference on Circuit, Power and Computing Technologies [ICCPCT]”.

[39] César O. Navarro, Nick A. Cromie, Colin Turner, Omar J. Escalona, Member IEEE, John McC Anderson “Detection of Cardiac Arrest using a Simplified Frequency Analysis of the Impedance Cardiogram recorded from Defibrillator Pads .presented in 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA “ year-August 30 - September 3, 2011.

2.3 Overcome the literature

Through this project we have presented the scientific community with a topic to debate on whether we can utilize the modern machineries to decrease the risks of deaths due to medical condition like cardiac arrest and increase life expectancy. During our preparation for this project we have read most of the research papers related to our topic or which can give insight related to our project.

We have seen that in one case one person was saved from cardiac arrest due to prediction by apple watch, we can conclude that from using tech with perfect parameters we can make valuable lifesaving prediction.

There have been many researches on this topic our project is much dependent on the research done before but no one tried to do this research on multiple datasets and bringing a cumulative answer to this problem. In our research we have applied multiple classifiers and used in total 4 datasets. Furthermore we also preserved the results as pickle file. Our idea of multiple datasets and pickling datasets is unique. But according to us the perfet model is far away to be made, we have made our contribution in the direction of the perfect model for prediction of cardiac arrest and continue to do our best in this direction.

2.4 Our Motivation for the project

In today's modern world cardiac arrest is one of the most dangerous disease .In earlier days it was expected to cause only to the old age people above 50 years ,but now as time changed due to the irregular eating habits ,poor diet, smoking, pollution and stress it is seeming to happen even for the person in his/her 20's.Most of the cardiac arrest are silent because of which many of them don't know if it is going to happen or whether it is cardiac arrest or not, because of which many lives are lost every year. It is impractical for a common man to frequently.

undergo costly tests like the ECG and thus there needs to be a system in place which is handy and at the same time reliable, in predicting the chances of a heart disease so we decided to make a project on prediction of cardiac arrest before it happens and by which many lives can be saved

2.5 Problem Definition

As there is an increase in problems related to heart, through this project we are going to observe different cardiac issues and habits of people which catalyzes the risk of CVD and predict whether someone is going to have a cardiac arrest and help people to correct their daily habits which pushes them towards the risk of CVD

CHAPTER 3 METHODLOGY

Introduction:

In this chapter, we are going to understand the need for a machine learning algorithm. How we can use machine learning in healthcare to process huge datasets beyond the scope of human capability is discussed in this chapter. For prediction of heart disease using machine learning algorithm, there are some steps such as identification of attributes which include age, sex, chest pain type, serum cholesterol, fasting blood sugar and many more which are discussed briefly below. For processing these attributes we need machine learning algorithms for example SVM, KNN, Decision Tree, Logistic Regression, Random Forest. The working of these algorithms and their implementation are discussed below. Algorithms have some advantages as well as some disadvantages. How the disadvantage of one algorithm is solved by another algorithm is also discussed below. Visualization of test results using histograms, graphs, images are also explained in this chapter.

3.1 Algorithms Used

3.1.1 Support Vector Machines (SVM)

- A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labelled training data for each category, they're able to categorize new text. So SVM is a fast and dependable classification algorithm that performs very well with a limited amount of data to analyse.

- SVM algorithm can be used for **Face detection, image classification, text categorization,** etc.

- SVC, NuSVC and LinearSVC are classes capable of performing binary and multi-class classification on a dataset. SVC and NuSVC are similar methods, but accept slightly different sets of parameters and have different mathematical formulations (see section Mathematical formulation). On the other hand, LinearSVC is another (faster) implementation of Support Vector Classification for the case of a linear kernel. Note that LinearSVC does not accept parameter kernel, as this is assumed to be linear. It also lacks some of the attributes of SVC and NuSVC, like support_.

- SVM can be of two types:

  a) Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
  b) Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and the classifier used is called a Non-linear SVM classifier.

- SVM can be understood with the example that we have used in the KNN classifier. Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, so such a model can be created by using the SVM algorithm. We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as the support vector creates a decision boundary between these two data (cat and dog) and choose extreme cases (support vectors), it will see the extreme case of cat and dog. Based on the support vectors, it will classify it as a cat. Consider the below diagram:

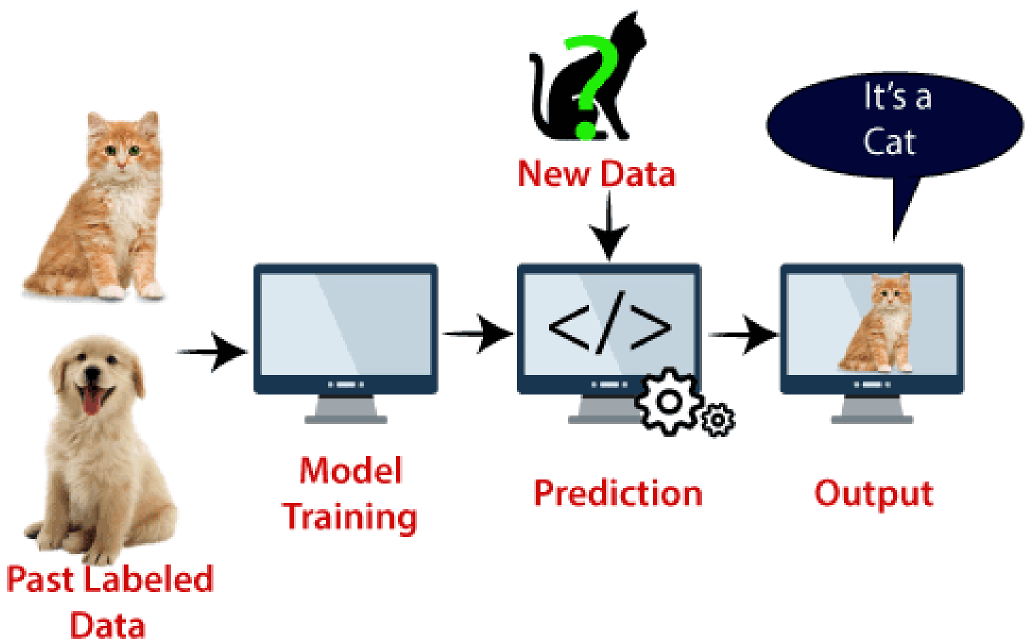


Fig 3.1 SVM Model

3.1.1.1 Advantages of support vector machines

- Effective in high dimensional spaces.

- Still effective in cases where the number of dimensions is greater than the number of samples.

- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.

- Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.

3.1.1.2 Disadvantages of support vector machines

- If the number of features is much greater than the number of samples, avoid over-fitting in choosing Kernel functions and the regularization term is crucial.

- SVMs do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation.

3.1.1.3 Working of SVM

An SVM model is a representation of different classes in a hyperplane in a multidimensional space. The hyperplane will be generated iteratively by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH).

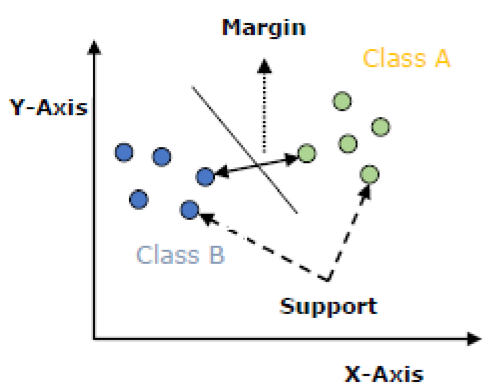


Fig 3.2 SVM Graph

The followings are important concepts in SVM −

- Support Vectors − Datapoints that are closest to the hyperplane is called support vectors. A separating line will be defined with the help of these data points.

- Hyperplane − As we can see in the above diagram, it is a decision plane or space which is divided between a set of objects having different classes.

- Margin − It may be defined as the gap between two lines on the closet data points of different classes. It can be calculated as the perpendicular distance from the line to the support vectors. A large margin is considered as a good margin and a small margin is considered as a bad margin.

The main goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH) and it can be done in the following two steps −

- First, SVM will generate hyperplanes iteratively that segregates the classes in the best way.

- Then, it will choose the hyperplane that separates the classes correctly.

3.1.1.4 Algorithm of SVC:

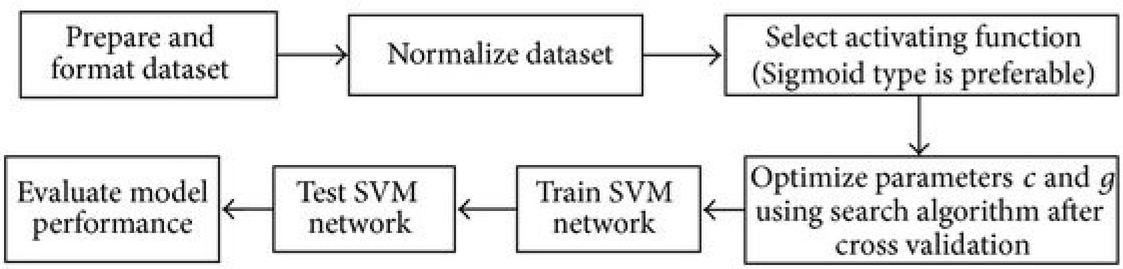


Fig No 3.3 SVM Block Diagram

- Step 1: In this step, we prepare the dataset and format it according to our needs.

- Step 2: Here we reduce data redundancy and eliminates undesirable characteristics like Insertion, Update and Deletion Anomalies.

- Step 3: Selection of an activating function for the algorithm.

- Step 4: Optimization of parameters using a search algorithm.

- Step 5: Training of SVM with the help of dataset collected in step 1.

- Step 6: Testing of the model performance in terms of accuracy/precision of the model.

- Step 7: Evaluation of model's performance on various aspects.

3.1.2 Decision Tree Algorithm

A decision tree is a flowchart tree-like structure that is made from training set tuples. The dataset is broken down into smaller subsets and is present in the form of nodes of a tree. The tree structure has a root node, internal nodes or decision nodes, leaf node, and branches.

The root node is the topmost node. It represents the best attribute selected for classification. Internal nodes of the decision nodes represent a test of an attribute of the dataset leaf node or terminal node which represents the classification or decision label. The branches show the outcome of the test performed.

- A decision tree is a supervised learning algorithm that works for both discrete and continuous variables. It splits the dataset into subsets based on the most significant attribute in the dataset. How the decision tree identifies this attribute and how this splitting is done is decided by the algorithms.

- The most significant predictor is designated as the root node, splitting is done to form sub-nodes called decision nodes, and the nodes which do not split further are terminal or leaf nodes.

- In the decision tree, the dataset is divided into homogeneous and non-overlapping regions. It follows a top-down approach as the top region presents all the observations at a single place which splits into two or more branches that further split. This approach is also called a *greedy approach* as it only considers the current node between the worked on without focusing on the future nodes.

- The decision tree algorithms will continue running until a stop criteria such as the minimum number of observations etc. is reached.

- Once a decision tree is built, many nodes may represent outliers or noisy data. The tree pruning method is applied to remove unwanted data. This, in turn, improves the accuracy of the classification model.

- To find the accuracy of the model, a test set consisting of test tuples and class labels is used. The percentages of the test set tuples are correctly classified by the model to identify the accuracy of the model. If the model is found to be accurate then it is used to classify the data tuples for which the class labels are not known.

- Some of the decision tree algorithms include Hunt's Algorithm, ID3, CD4.5, and CART.

- A decision tree algorithm is a process in which the data given is first split according to the given decision. The data splitting is based on which category thus the data belong (e.g.: age, height, material, etc). Then the split data is compared to the condition given. Then the compared data is checked to whether it satisfies the condition. If satisfy then the value is 1 or high and whether not satisfied the value is 0 or low then the high data is considered as a result and further process is taken. Many steps are involved in the working of a decision tree. It is the process of the partitioning of data into subsets. Splitting can be done on various factors as shown below i.e. on a gender basis, height basis or based on class.

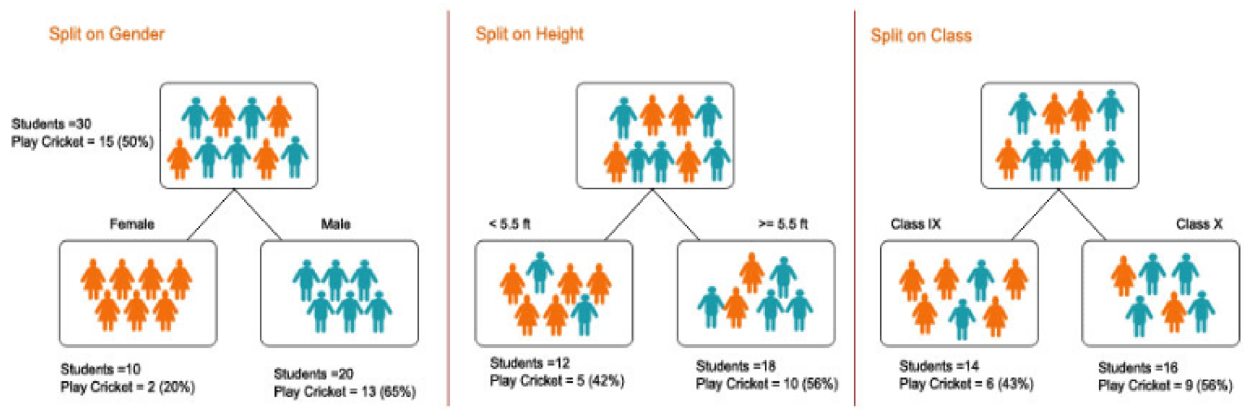


Fig 3.4 Decision Tree Model

3.1.2.1 Advantages of Decision tree

- It is very easy to understand and interpret.

- The data for decision trees require minimal preparation.

- They force you to find many possible outcomes of a decision.

- Can be easily used with many other decision tools.

- Helps you to make the best decisions and best guesses based on the information you have.

- Helps you to see the difference between controlled and uncontrolled events.
- Helps you estimate the likely results of one decision against another.

3.1.2.2 Disadvantages of Decision tree

- Sometimes decision trees can become too complex.
- The outcomes of decisions may be based mainly on your expectations. This can lead to unrealistic decision trees.
- The diagrams can narrow your focus to critical decisions and objectives.

3.1.2.3 Flow Chart

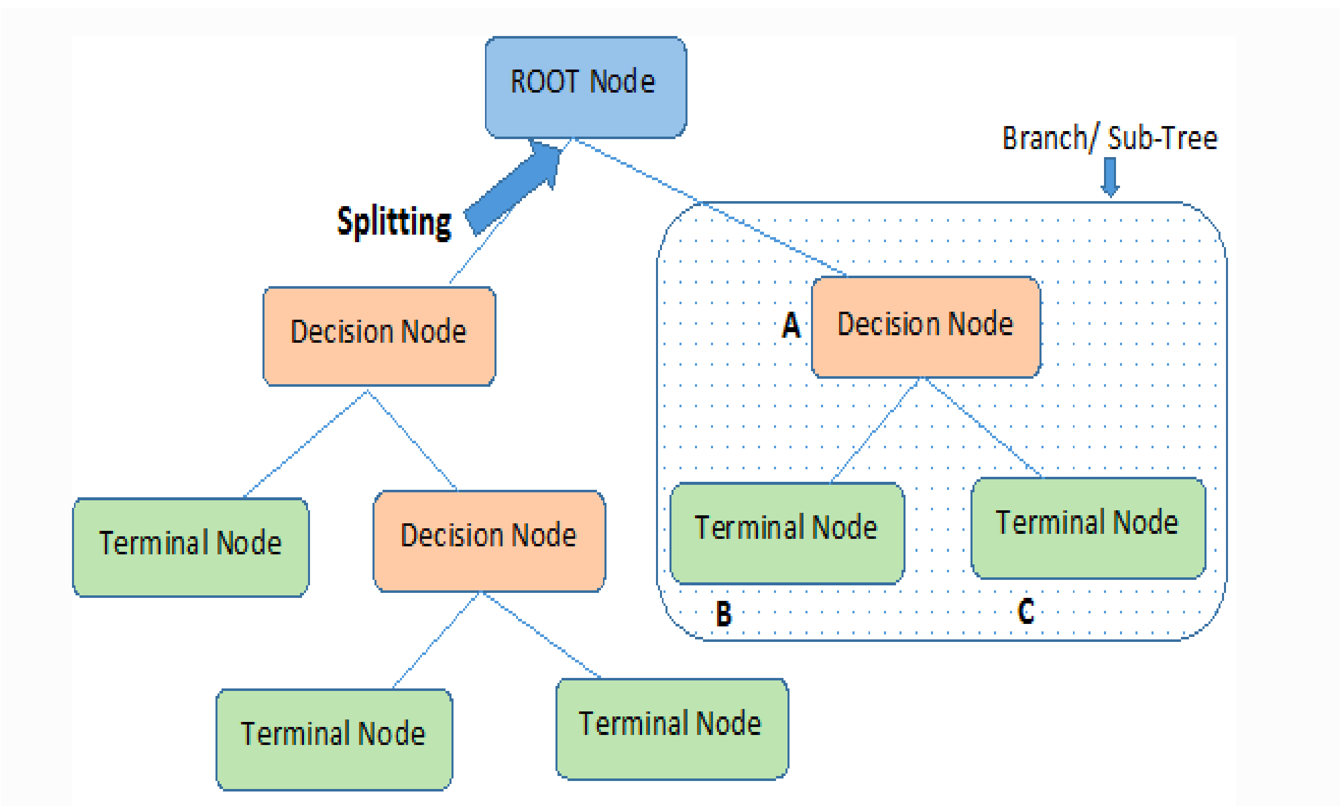


Fig No 3.5 Decision Tree Flowchart

- Step 1: **Root Node** It represents the entire population or sample and this further gets divided into two or more homogeneous sets.

- Step 2: **Splitting** is the process of dividing a node into two or more sub-nodes.
- Step 3: When a sub-node splits into further sub-nodes, then it is called the decision node.
- Step 4: **A node, which is divided into sub-nodes is called a parent node of sub-nodes whereas sub-nodes are the child of a parent node.**
- Step 5: When we remove sub-nodes of a decision node, this process is called pruning. You can say the opposite process of splitting.
- Step 6: A subsection of the entire tree is called a branch or sub-tree.
- Step 7: Nodes do not split is called Leaf or Terminal node.
- Step 8: Here A is parent node B and C.

3.1.3 K-Nearest Neighbour (KNN) Algorithm

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on the Supervised Learning technique.

- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a good suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for classification problems.
- K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.

- It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

- KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

- Example: Suppose, we have an image of a creature that looks similar to a cat and dog, but we want to know either it is a cat or dog. So, for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cat's and dog's images and based on the most similar features it will put it in either cat or dog category.

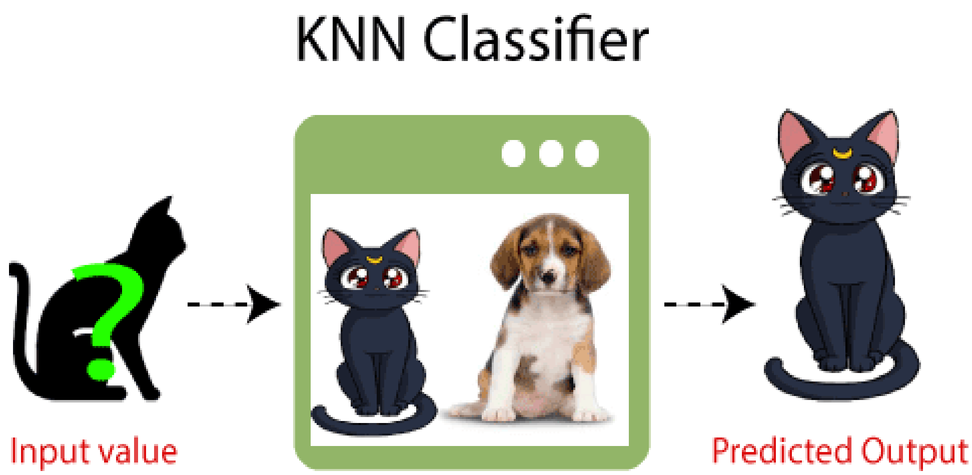


Fig 3.6 K-nn Model

3.1.3.1 Need of K-NN Algorithm:

Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x1, so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. Consider the below diagram:

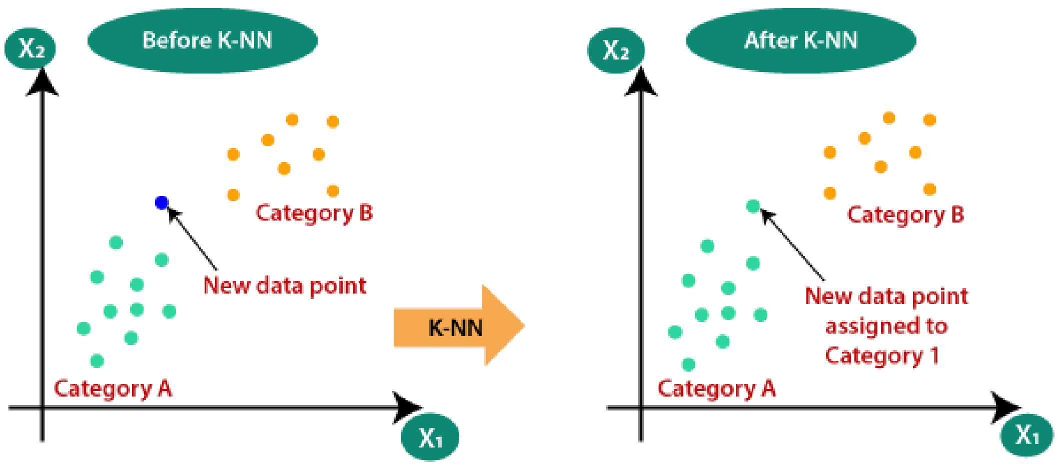


Fig 3.7 K-nn Graphs

3.1.3.2 Working Of K-NN:

The K-NN working can be explained based on the below algorithm:

- **Step-1:** Select the number K of the neighbours
- **Step-2:** Calculate the Euclidean distance of **K number of neighbours**
- **Step-3:** Take the K nearest neighbours as per the calculated Euclidean distance.
- **Step-4:** Among these k neighbours, count the number of the data points in each category.
- **Step-5:** Assign the new data points to that category for which the number of neighbours is maximum.
- **Step-6:** Our model is ready.

Suppose we have a new data point and we need to put it in the required category. Consider the below image:

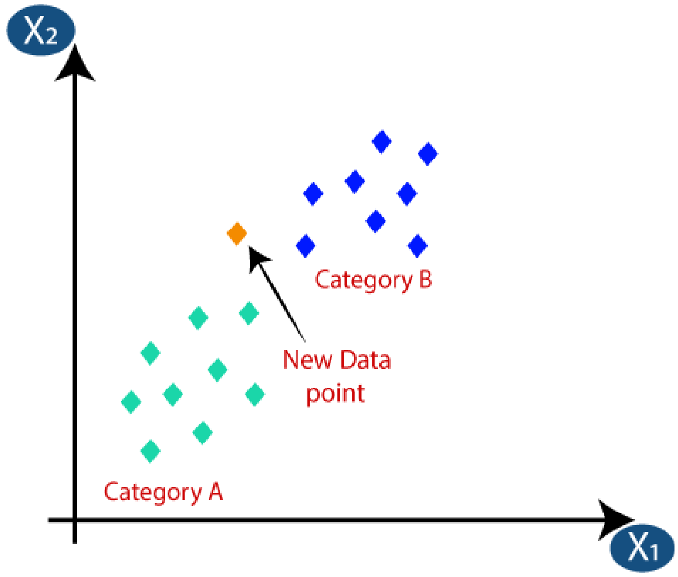


Fig 3.8 K-nn Graph

- Firstly, we will choose the number of neighbours, so we will choose k=5.
- Next, we will calculate the Euclidean distance between the data points. The Euclidean distance is the distance between two points, which we have already studied in geometry. It can be calculated as:

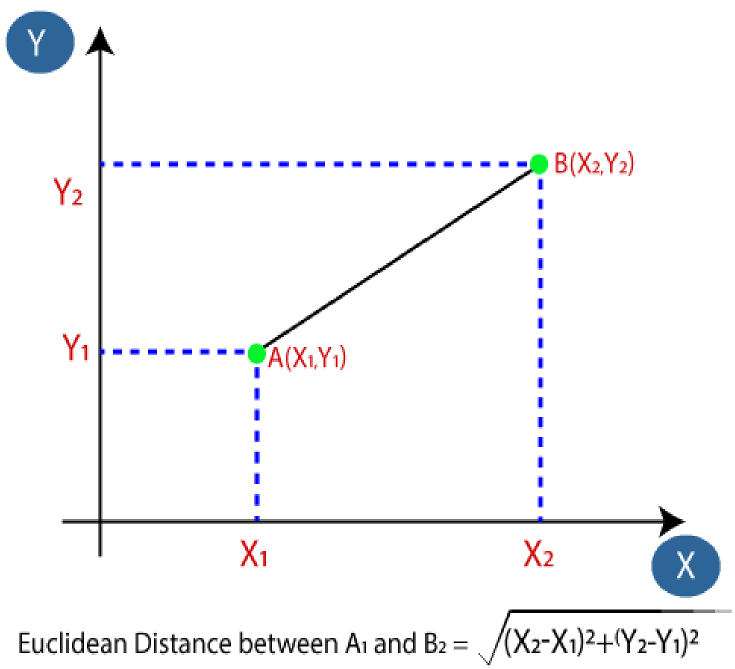


Fig 3.9 Euclidean Distance (K-nn)

- As we can see the 3 nearest neighbours are from category A, hence this new data point must belong to category A.

3.1.3.3 Selection of K value in the K-NN Algorithm

Below are some points to remember while selecting the value of K in the K-NN algorithm:

- There is no particular way to determine the best value for "K", so we need to try some values to find the best out of them. The most preferred value for K is 5.
- A very low value for K such as K=1 or K=2, can be noisy and lead to the effects of outliers in the model.
- Large values for K are good, but it may find some difficulties.

3.1.3.4 Advantages of KNN

- It is simple to implement.
- It is robust to the noisy training data
- It can be more effective if the training data is large.

3.1.3.5 Disadvantages of KNN

- Always needs to determine the value of K which may be complex sometimes.
- The computation cost is high because of calculating the distance between the data points for all the training samples.

3.1.3.6 Steps to implement the K-NN algorithm:

- Data Pre-processing step
- Fitting the K-NN algorithm to the Training set
- Predicting the test result

- Test accuracy of the result(Creation of Confusion matrix)
- Visualizing the test set result.

3.1.4 Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model.*

As the name suggests, *"Random Forest is a classifier that contains several decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."* Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

The below diagram explains the working of the Random Forest algorithm:

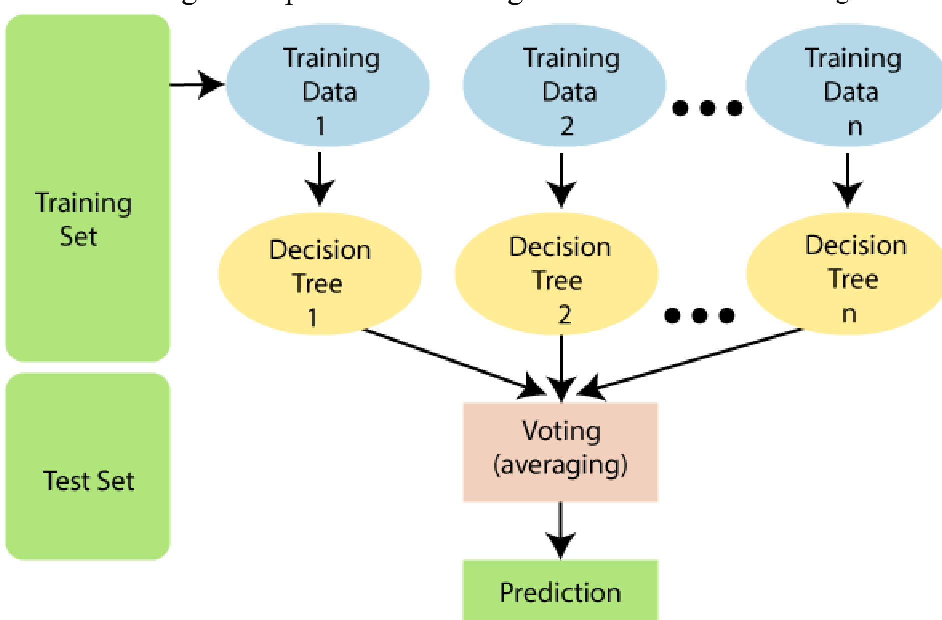


Fig 3.10 Random Forest Model

3.1.4.1 Assumptions for Random Forest

Since the random forest combines multiple trees to predict the class of the dataset, some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random forest classifier:

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.

3.1.4.2 Use of Random Forest:

Below are some points that explain why we should use the Random Forest algorithm:

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.
- It can also maintain accuracy when a large proportion of data is missing.

3.1.4.3 Working of Random Forest algorithm:

- Random Forest works in two-phase first is to create the random forest by combining the N decision tree, and the second is to make predictions for each tree created in the first phase.
- The Working process can be explained in the below steps and diagram:
    - Step-1: Select random K data points from the training set.
    - Step-2: Build the decision trees associated with the selected data points (Subsets).
    - Step-3: Choose the number N for decision trees that you want to build.
    - Step-4: Repeat Step 1 & 2.
    - Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.
- The working of the algorithm can be better understood by the below example:
  Example: Suppose there is a dataset that contains multiple fruit images. So, this dataset is given to the Random forest classifier. The dataset is divided into subsets and

given to each decision tree. During the training phase, each decision tree produces a prediction result, and when a new data point occurs, then based on the majority of results, the Random Forest classifier predicts the final decision. Consider the below image:

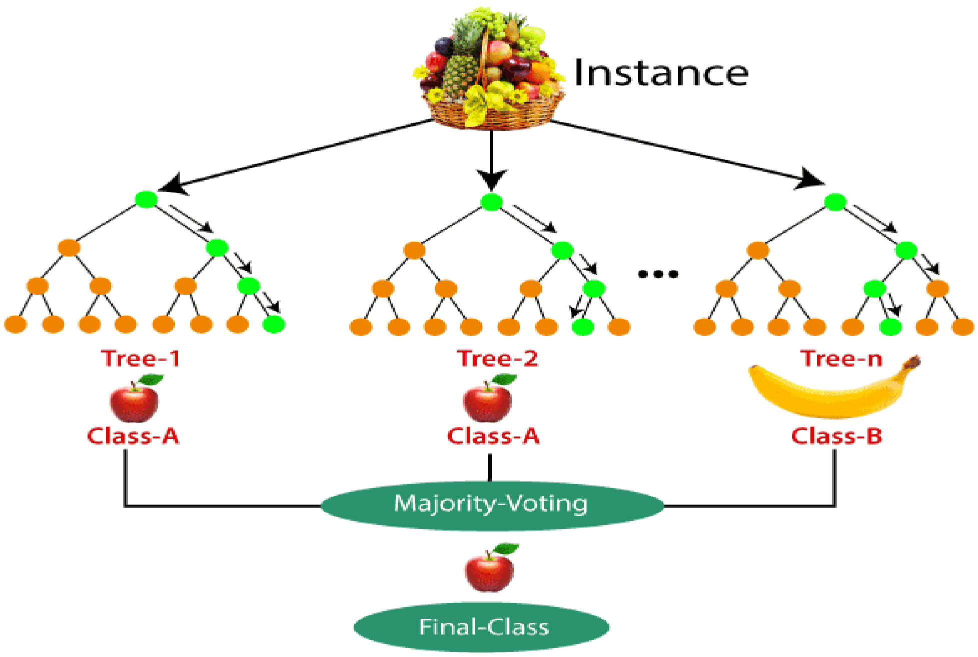


Fig 3.11 Random Forest (Fruit) Example

3.1.4.4 Applications of Random Forest

There are mainly four sectors where Random forest mostly used:

- Banking: The banking sector mostly uses this algorithm for the identification of loan risk.
- Medicine: With the help of this algorithm, disease trends and risks of the disease can be identified.
- Land Use: We can identify the areas of similar land use by this algorithm.
- Marketing: Marketing trends can be identified using this algorithm.

3.1.4.5 Advantages of Random Forest

- Random Forest is capable of performing both Classification and Regression tasks.

- It is capable of handling large datasets with high dimensionality.
- It enhances the accuracy of the model and prevents the overfitting issue.

3.1.4.6 Disadvantages of Random Forest

Although random forest can be used for both classification and regression tasks, it is not more suitable for Regression tasks.

3.1.4.7 Python Implementation of Random Forest Algorithm

- Now we will implement the Random Forest Algorithm tree using Python. For this, we will use the same dataset "user_data.csv", which we have used in previous classification models. By using the same dataset, we can compare the Random Forest classifier with other classification models such as Decision tree Classifier, KNN, SVM, Logistic Regression, etc.
- Implementation Steps are given below:
  - Data Pre-processing step
  - Fitting the Random forest algorithm to the Training set
  - Predicting the test result
  - Test accuracy of the result (Creation of Confusion matrix)
  - Visualizing the test set result.

3.1.5 Logistic Regression

- Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.
- Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

- Logistic Regression is much similar to Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.

- In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).

- The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.

- Logistic Regression is a significant machine learning algorithm because it can provide probabilities and classify new data using continuous and discrete datasets.

- Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification. The below image is showing the logistic function:

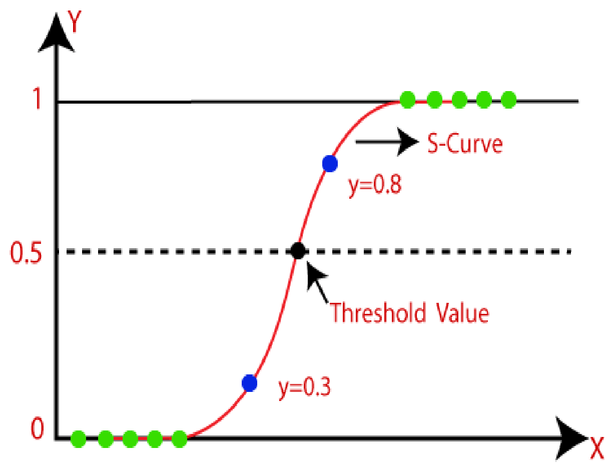


Fig 3.12 Logistic Regression Graph

3.1.5.1 Type of Logistic Regression:

Based on the categories, Logistic Regression can be classified into three types:

- **Binomial:** In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.

- **Multinomial:** In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as "cat", "dogs", or "sheep"

- **Ordinal:** In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High".

3.1.5.2 Steps in Logistic Regression

To implement the Logistic Regression using Python, we will use the same steps as we have done in previous topics of Regression. Below are the steps:

- Data Pre-processing step
- Fitting Logistic Regression to the Training set
- Predicting the test result
- Test accuracy of the result (Creation of Confusion matrix)
- Visualizing the test set result.

3.1.5.3 Working of Logistic Regression:

- Logistic regression measures the relationship between the dependent variable (our label, what we want to predict) and the one or more independent variables (or features), by estimating probabilities using its underlying logistic function.

- These probabilities must then be transformed into binary values to make a prediction. This is the task of the logistic function, also called the sigmoid function. The Sigmoid-Function is an S-shaped curve that can take any real-valued number and map it into a value between the range of 0 and 1, but never exactly at those limits. These values between 0 and 1 will then be transformed into either 0 or 1 using a threshold classifier.

- The picture below illustrates the steps that logistic regression goes through to give you your desired output.

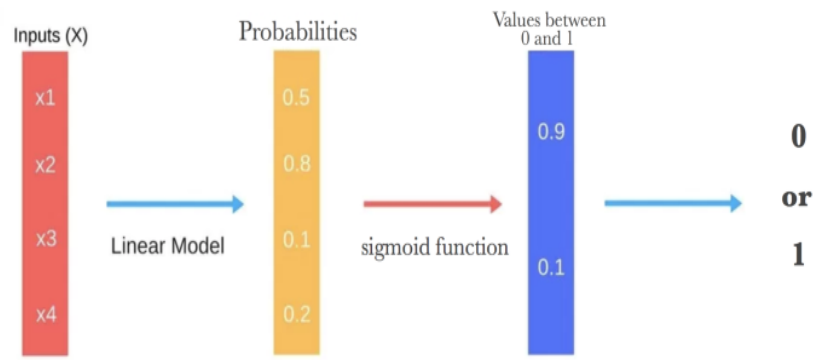


Fig 3.13 Steps of Logistic Regression

Below you can see how the logistic function (sigmoid function) looks like:

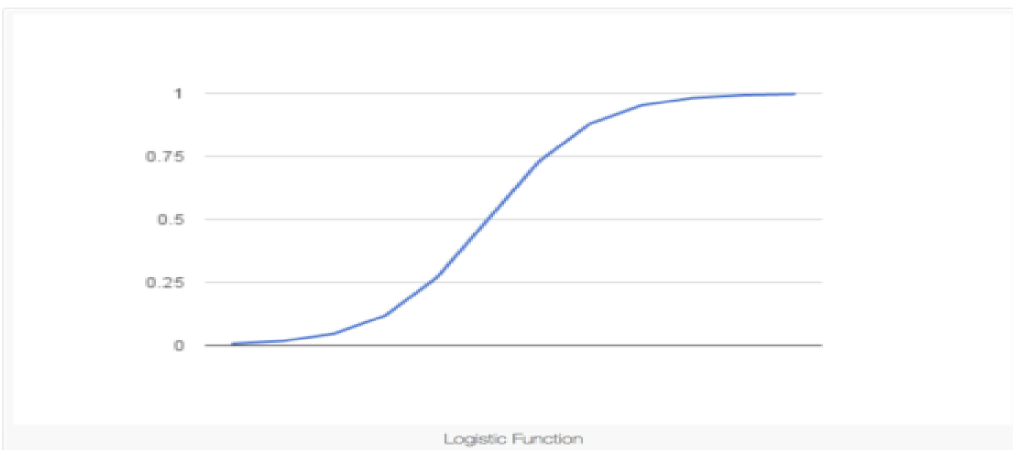


Fig 3.14 Logistic function (sigmoid function)

- We want to maximize the likelihood that a random data point gets classified correctly, which is called Maximum Likelihood Estimation. Maximum Likelihood Estimation is a general approach to estimating parameters in statistical models.

- We can maximize the likelihood using different methods like an optimization algorithm. Newton's Method is such an algorithm and can be used to find the maximum (or minimum) of many different functions, including the likelihood function. Instead of Newton's Method, you could also use Gradient Descent.

3.1.5.4 Advantages

- Logistic Regression is **one of the simplest machine learning algorithms** and is easy to implement yet provides great training efficiency in some cases. Also due to these reasons, training a model with this algorithm doesn't require high computation power.

- The predicted parameters (trained weights) give **inference about the importance of each feature**. The direction of association i.e. positive or negative is also given. So we can use logistic regression to find out the relationship between the features.

- This algorithm allows models to be **updated easily to reflect new data**, unlike decision trees or support vector machines. The update can be done using stochastic gradient descent.

- Logistic Regression **outputs well-calibrated probabilities** along with classification results. This is an advantage over models that only give the final classification as results. If a training example has a 95% probability for a class, and another has a 55% probability for the same class, we get an inference about which training examples are more accurate for the formulated problem.

3.1.5.5 Disadvantages

- Logistic Regression is a statistical analysis model that attempts to predict precise probabilistic outcomes based on independent features. On **high dimensional datasets**, this may lead to the model being **over-fit on the training set**, which means overstating the accuracy of predictions on the training set and thus the model **may not be able to predict accurate results on the test set**. This usually happens in the case when the model is trained on little training data with lots of features. So on high dimensional datasets, Regularization techniques should be considered to avoid over-fitting (but this makes the model complex). Very high regularization factors may even lead to the model being under-fit on the training data.

- **Non-linear problems can't be solved** with logistic regression **since it has a linear decision surface**. Linearly separable data is rarely found in real-world scenarios. So the transformation of non-linear features is required which can be done by increasing the number of features such that the data becomes linearly separable in higher dimensions.

- It is **difficult to capture complex relationships** using logistic regression. More powerful and complex algorithms such as Neural Networks can easily outperform this algorithm.

- The training features are known as independent variables. Logistic Regression **requires moderate or no multicollinearity between independent variables**. This means if two independent variables have a high correlation, only one of them should be used. **Repetition of information could lead to wrong training of parameters** (weights) during minimizing the cost function. Multicollinearity can be removed using dimensionality reduction techniques.

3.2 Block Diagram of Project Flow

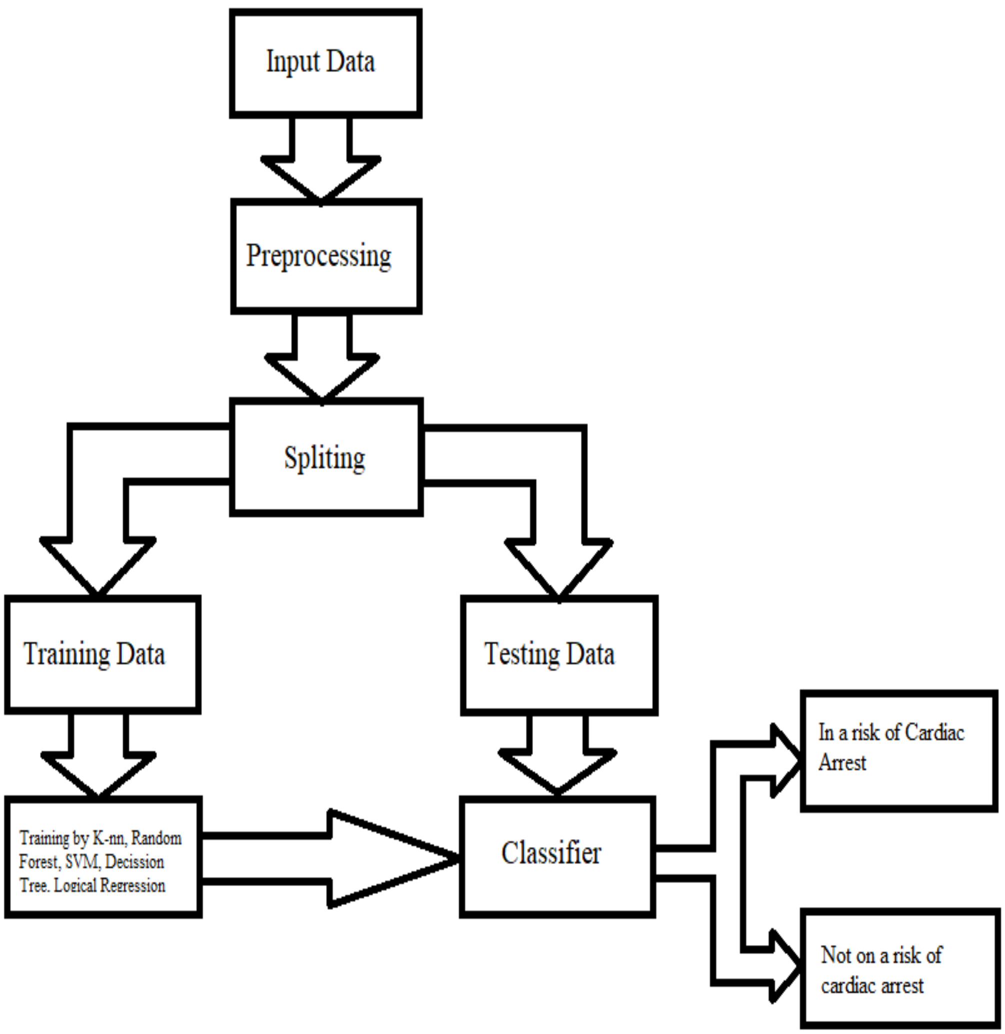


Fig 3.15 Block diagram of ML model for Cardiac Arrest Prediction

3.2.1 Description

3.2.2.1 Input Data

- Cleveland database from UCI's repository was used for the heart disease prediction system. Because the Cleveland database is the most commonly used database by ML researchers. The dataset contains 303 instances and 76 attributes, but only 14 of them are referred to by all published studies.

- The dataset was split into three parts: one for training (%70), the second one for testing (%15) and the third one for validation (%15). There are 213 instances and 13 attributes in training data. Test data and validation data contain 45 instances and 13 attributes.

- A Cleveland dataset is a collection of information that is organized so that it can be easily accessed, managed and updated. Dataset process workloads to create and update themselves, querying the data they contain and running applications against it.

- In this project of Prediction of the Risk of Heart Attack after finding different parameters that could help analyze the person's risk of having CVD. A huge database is required to store the information.

- The mentioned dataset is taken from different patients considering their age, gender, blood pressure, family history, cholesterol etc.

3.2.2.2 Pre-Processing

- Data preprocessing is an essential step use to clean the data and make it useful for any experiment associated with machine learning or data mining. In this study, multiple preprocessing steps applied to the selected dataset.

- The dataset contains 14 columns and 303 rows. We see that there are only 6 cells with null values with 4 belonging to attribute *ca* and 2 to *that.* As the null values are very less, we can either drop them or impute them. I have imputed the mean in place of the null values however one can also delete these rows entirely.

3.2.2.3 Splitting

- Data is at the heart of every ML problem. Without proper data, ML models are just like bodies without soul. But in today's world of 'big data' collecting data is not a major problem anymore. We are knowingly (or unknowingly) generating huge datasets every day. However, having surplus data at hand still does not solve the problem. For ML models to give reasonable results, we not only need to feed in large quantities of data but also have to ensure the quality of data.

- Data is at the heart of every ML problem. Without proper data, ML models are just like bodies without soul. But in today's world of 'big data' collecting data is not a major problem anymore. We are knowingly (or unknowingly) generating huge datasets every day. However, having surplus data at hand still does not solve the problem. For ML models to give reasonable results, we not only need to feed in large quantities of data but also have to ensure the quality of data.

- The data should ideally be divided into 2 sets – namely, train and test set. 80% of data is taken for training while the remaining 20% of data is used for testing.

3.2.2.3 Training Data

- The train set would contain the data which will be fed into the model. In simple terms, our model would learn from this data. For instance, a Regression model would use the examples in this data to find gradients to reduce the cost function.

- Then these gradients will be used to reduce the cost and predict data effectively. The evaluation metric used is the confusion matrix. The confusion matrix displays the correctly predicted as well as incorrectly predicted values by a classifier.

- The sum of TP and TN, from the confusion matrix, is the number of correctly classified entries by the classifier. The algorithms are implemented with the default parameters only.

- In our case training of the machine is done with the help of K-nn, Decision tree, SCM, logical regression, Random forest as base classifier.

3.2.2.4 Testing Data

- The test set contains the data on which we test the trained and validated model. It tells us how efficient our overall model is and how likely is it going to predict something which does not make sense.

- There is a plethora of evaluation metrics (like precision, recall, accuracy, etc.) that can be used to measure the performance of our model. In simple terms, it is used to evaluate the fit machine learning model.

3.2.2.5 Classifier

- Classification is the process of predicting the class of given data points. Classes are sometimes called targets/ labels or categories. Classification predictive modelling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y).

- A classifier utilizes some training data to understand how given input variables relate to the class. Classification belongs to the category of supervised learning where the targets also provided with the input data.

- In our case final classification is done by hyperparameter tuning of logistic regression and N-nn and prediction is done whether the subject is going to have a cardiac arrest or not (valued 0 or 1).

3.3 Flow Chart

Fig 3.16 Block Diagram showing Proposed Methodology

3.3.1 Understanding the need for a Machine Learning algorithm

Machine Learning is the core subarea of artificial intelligence. It makes computers get into a self-learning mode without explicit programming. When fed new data, these computers learn, grow, change, and develop by themselves. The value of machine learning in healthcare is its ability to process huge datasets beyond the scope of human capability, and then reliably convert analysis of that data into clinical insights that aid physicians in planning and providing care, ultimately leading to better outcomes, lower costs of care, and increased patient satisfaction.

3.3.2 Collecting the information for the dataset

This step in the checklist is akin to what is often referred to as Exploratory Data Analysis (EDA). The goal is to try and gain insights from the data before modelling. Recall that in the first step assumptions about the data were to be identified and explored; this is a good time to more deeply investigate these assumptions. Human experts can be of particular use in this step, answering questions about correlations that may not be obvious to the machine learning practitioner. Studying features and their characteristics are done here, as is general visualization of features and their values (think of how much easier it is, for example, to quickly identify outliers by box plot than by numerical interrogation). Documenting the findings of your exploration for later use is good practice.

3.3.3 Identification of attributes causing Cardiac Arrest and pre-processing on it

The dataset consists of 303 individual data. There are 14 columns in the dataset, which are described below.

1. ***Age*****:** displays the age of the individual.

2. ***Sex*****:** displays the gender of the individual using the following format :
   1 = male
   0 = female

3. ***Chest-pain type*****:** displays the type of chest pain experienced by the individual using the following format:
   1 = typical angina
   2 = atypical angina
   3 = non — anginal pain
   4 = asymptotic

4. ***Resting Blood Pressure*****:** displays the resting blood pressure value of an individual in mmHg (unit)

5. ***Serum Cholesterol*****:** displays the serum cholesterol in mg/dl (unit)

6. ***Fasting Blood Sugar*****:** compares the fasting blood sugar value of an individual with 120mg/dl.
If fasting blood sugar > 120mg/dl then: 1 (true)
else: 0 (false)

7. ***Resting ECG*****:** displays resting electrocardiographic results
0 = normal
1 = having ST-T wave abnormality
2 = left ventricular hyperthrophy

8. ***Max heart rate achieved*****:** displays the max heart rate achieved by an individual.

9. *Exercise-induced angina*:
1 = yes
0 = no

10. ***ST depression induced by exercise relative to rest*****:** displays the value which is an integer or float.

11. ***Peak exercise ST segment*****:**
1 = upsloping
2 = flat
3 = down sloping

12. ***The number of major vessels (0–3) coloured by fluoroscopy*****:** displays the value as integer or float.

13. ***Thal*****:** displays the thalassemia:
3 = normal
6 = fixed defect
7 = reversible defect

14. ***Diagnosis of heart disease*****:** Displays whether the individual is suffering from heart disease or not:

0 = absence
1, 2, 3, 4 = present.

3.3.4 Fitting the classifier algorithm to the Training set

- Model fitting is a measure of how well a machine learning model generalizes to similar data to that on which it was trained. A well-fitted model produces more accurate outcomes. A model that is overfitted matches the data too closely. A model that is under fitted doesn't match closely enough. Each machine learning algorithm has a basic set of parameters that can be changed to improve its accuracy.

- During the fitting process, you run an algorithm on data for which you know the target variable, known as "labelled" data, and produce a machine learning model. Then, you compare the outcomes to real, observed values of the target variable to determine their accuracy.

- Next, we use that information to adjust the algorithm's standard parameters to reduce the level of error, making it more accurate in uncovering patterns and relationships between the rest of its features and the target.

- We repeat this process until the algorithm finds the optimal parameters that produce valid, practical, applicable insights for our practical medical application problem.

3.3.5 Predicting the test result

- Our primary focus for reducing waste in healthcare expenditure is identifying and discouraging unnecessary repeat lab tests. A machine learning model which could reliably predict low information lab tests could provide personalized, real-time predictions to discourage over-testing. to predict when the next measurement of a lab test is likely to be the "same" as the previous one.

- This points to potential areas where machine learning approaches may identify and prevent unneeded testing before it occurs, and a methodological framework for how these tasks can be accomplished.

3.3.6 Test accuracy of the result (Creation of Confusion matrix)

- A confusion matrix is a technique for summarizing the performance of a classification algorithm.

- Classification accuracy is the ratio of correct predictions to total predictions made.
- The confusion matrix shows how your classification model is confused when it makes predictions.
- It gives you insight not only into the errors being made by your classifier but more importantly the types of errors that are being made.
- It is this breakdown that overcomes the limitation of using classification accuracy alone.

3.3.6 Visualizing the test set result

- During the training of models, it is necessary to visualize the training process and helps to understand debug models and track errors and progress of the models. Visualize the test and train result values at each step.

- Scalars (Loss and Accuracy) – Scalar can be used to show the trends of error during training. Besides logging the loss and accuracy to the stdout regularly, we record and plot them to analyze its long-term trend.

- Histograms – Visualize how the distribution of tensors in the model graph has changed over time. Showing many histograms visualizations of tensor at different points in time.

- Weight and bias –Monitor the weights and the biases during training time by visualizing them on histograms.

- Activation – For gradient descent to perform the best, the node outputs before the activation functions usually distributed.

- Gradients – Gradients can be visualized for each layer to identify deep learning problems like gradient diminishing or exploding problems.

- Graphs –Graphs visualize the internal structure or architecture of the model.

- Image –Images at each step of training means intermediate images generated can be visualized and visualize tensors.

- Projector – Visualize the principal component analysis and t-sne algorithm model results. This technique mainly uses for dimensionality reduction.

CHAPTER 4 RESULTS AND DISCUSSION

4.1 Results

4.1.1 Dataset 1

	Train Accuracy	Test Accuracy
Logistic Regression	82.82	71.05
Random Forest Classifier	100	69.74
K-Nearest Neighbour	83.26	76.32
Support Vector Machine	87.22	76.32
Decision tree	100	69.74

Table 4.1 Accuracy Results of Dataset 1

4.1.2 Dataset 2

	Train Accuracy	Test Accuracy
Logistic Regression	94.25	86.21
Random Forest Classifier	100	93.1
K-Nearest Neighbour	91.95	89.66
Support Vector Machine	93.1	93.1
Decision tree	100	89.66

Table 4.2 Accuracy Results of Dataset 2

4.1.3 Dataset 3

	Train Accuracy	Test Accuracy
Logistic Regression	81.59	83.82
Random Forest Classifier	100	82.35
K-Nearest Neighbour	85.57	82.35
Support Vector Machine	86.57	85.29
Decision tree	100	77.94

Table 4.3 Accuracy Results of Dataset 3

4.1.4 Dataset 4

	Train Accuracy	Test Accuracy
Logistic Regression	83.78	72
Random Forest Classifier	100	68
K-Nearest Neighbour	83.78	64
Support Vector Machine	83.78	64
Decision tree	100	80

Table 4.4 Accuracy Results of Dataset 4

4.1.5 Data Balancing of Dataset 1

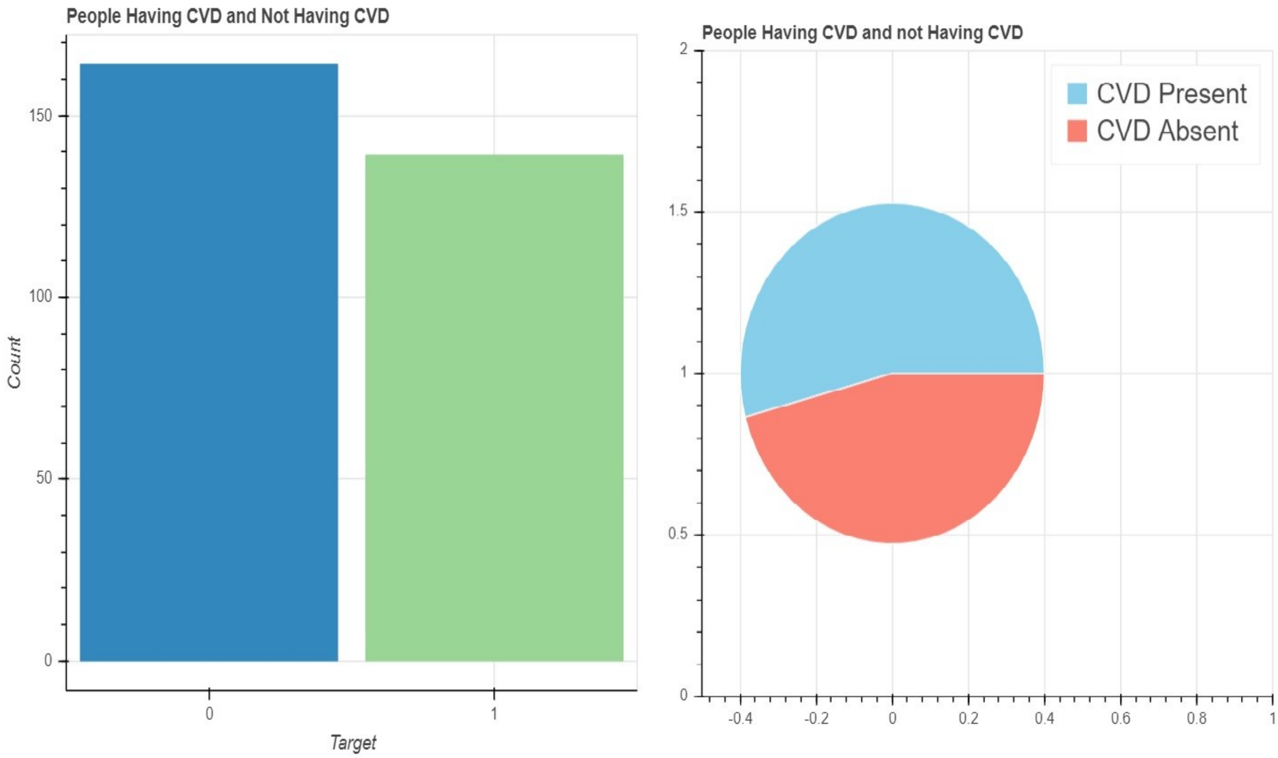


Fig 4.1 Proportion of People Having CVD and not Having CVD(Dataset 1)

4.1.6 Data Balancing of Dataset 3

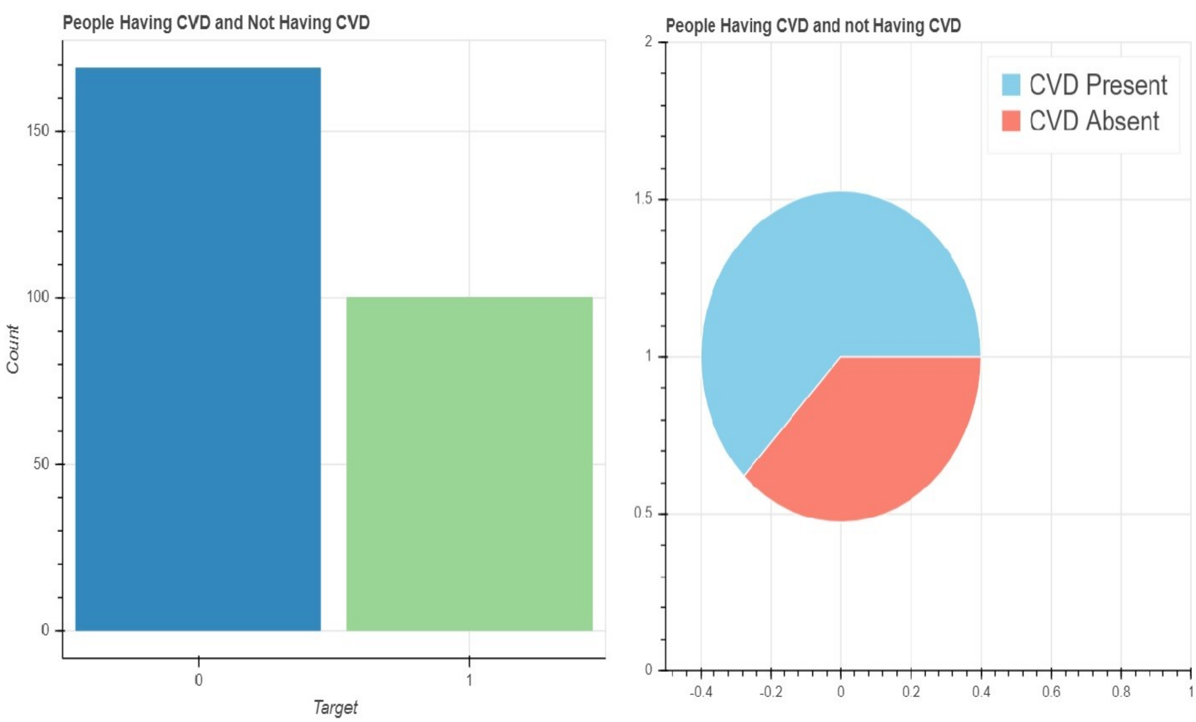


Fig 4.2 Proportion of People Having CVD and not Having CVD(Dataset 3)

4.1.7 Data Balancing of Dataset 4

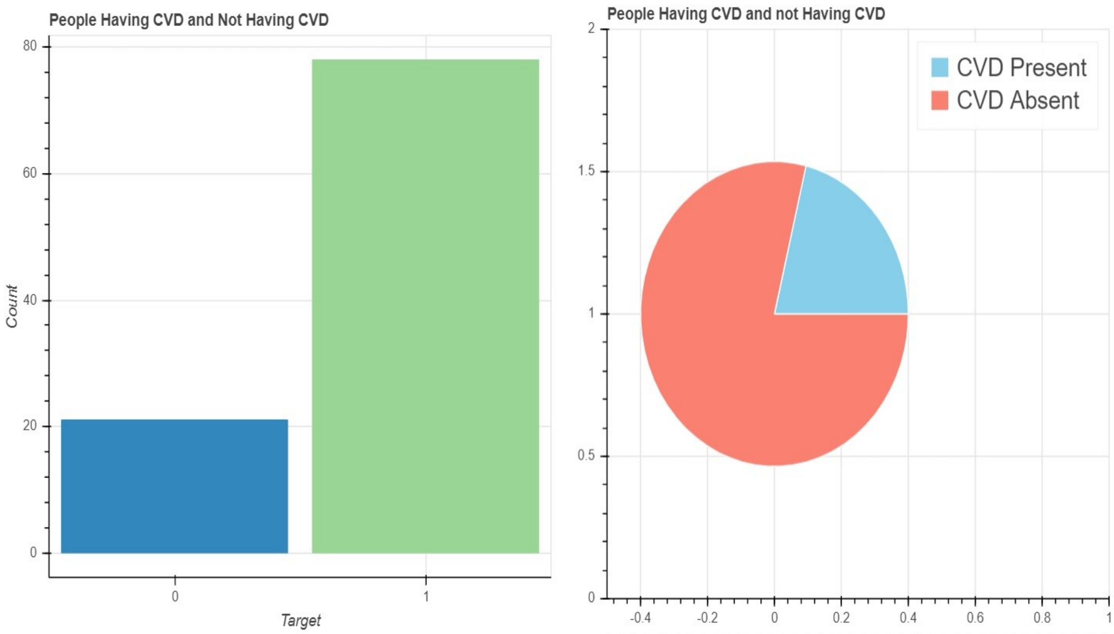


Fig 4.3 Proportion of People Having CVD and not Having CVD(Dataset 4)

4.1.8 Correlation Matrix of Dataset 1

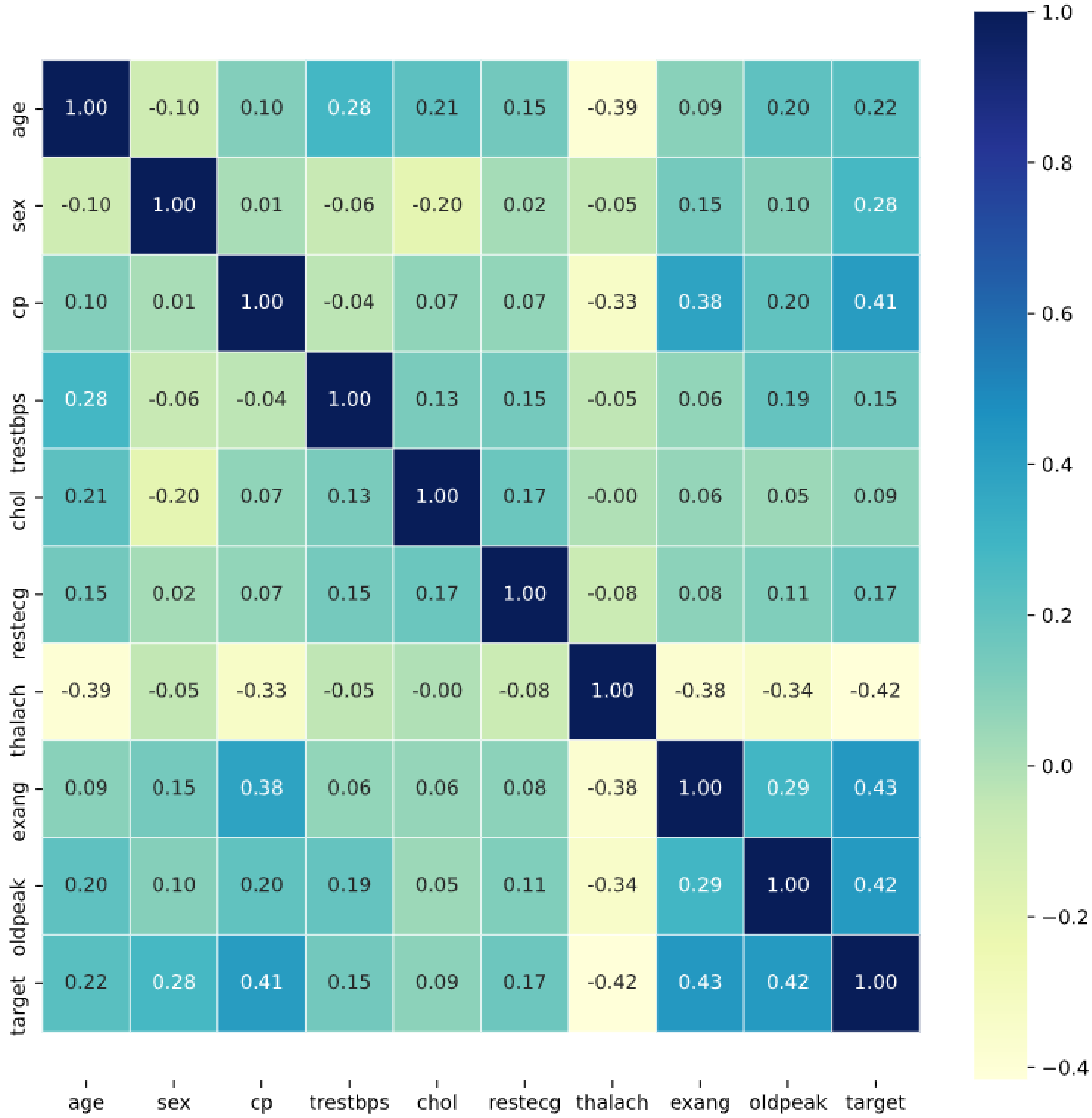
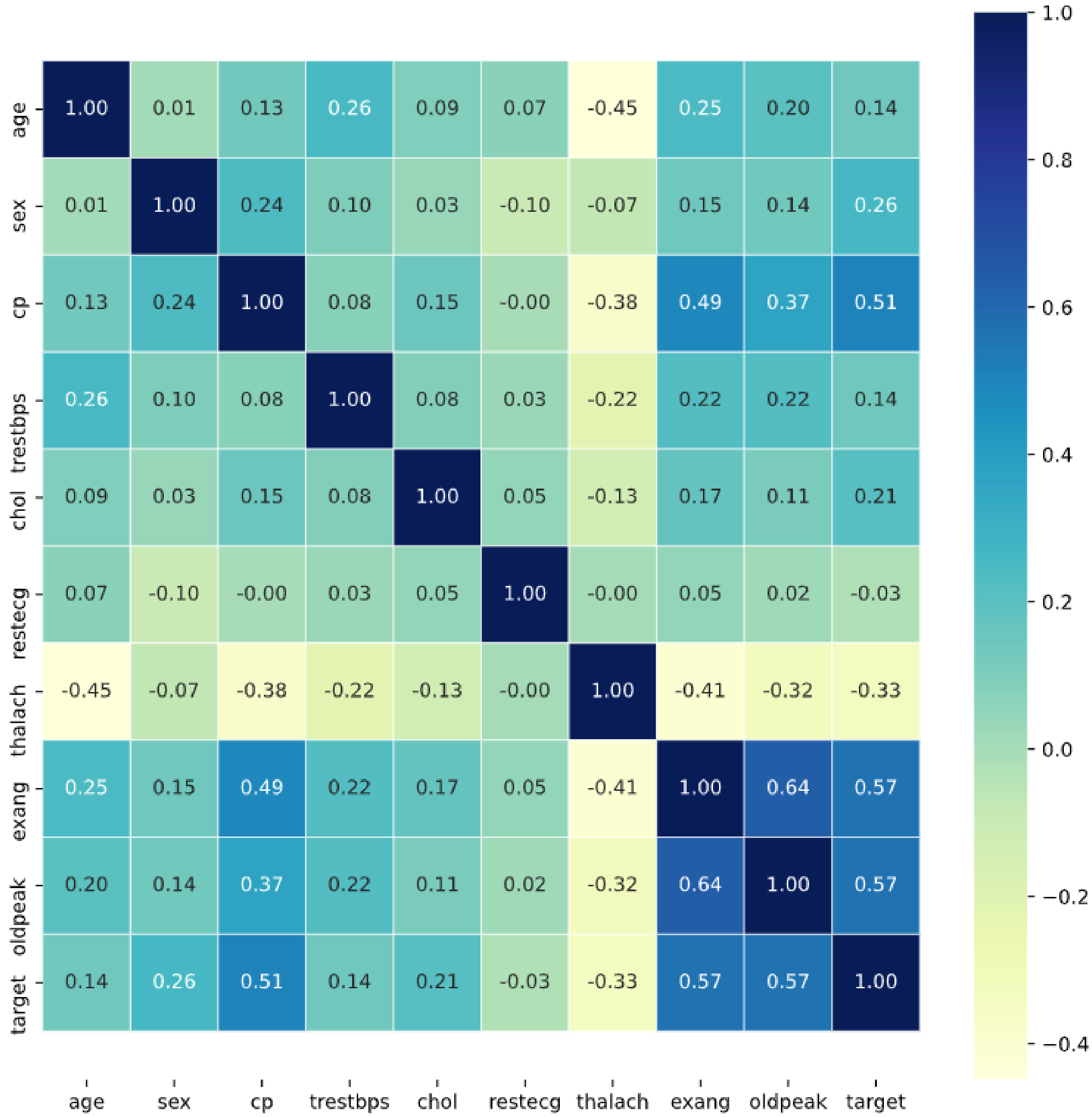


Fig 4.4 Correlation Matrix of Dataset 1

4.1.9 Correlation Matrix of Dataset 3



<u>Fig 4.5 Correlation Matrix of Dataset 3</u>

4.1.10 Correlation Matrix of Dataset 4

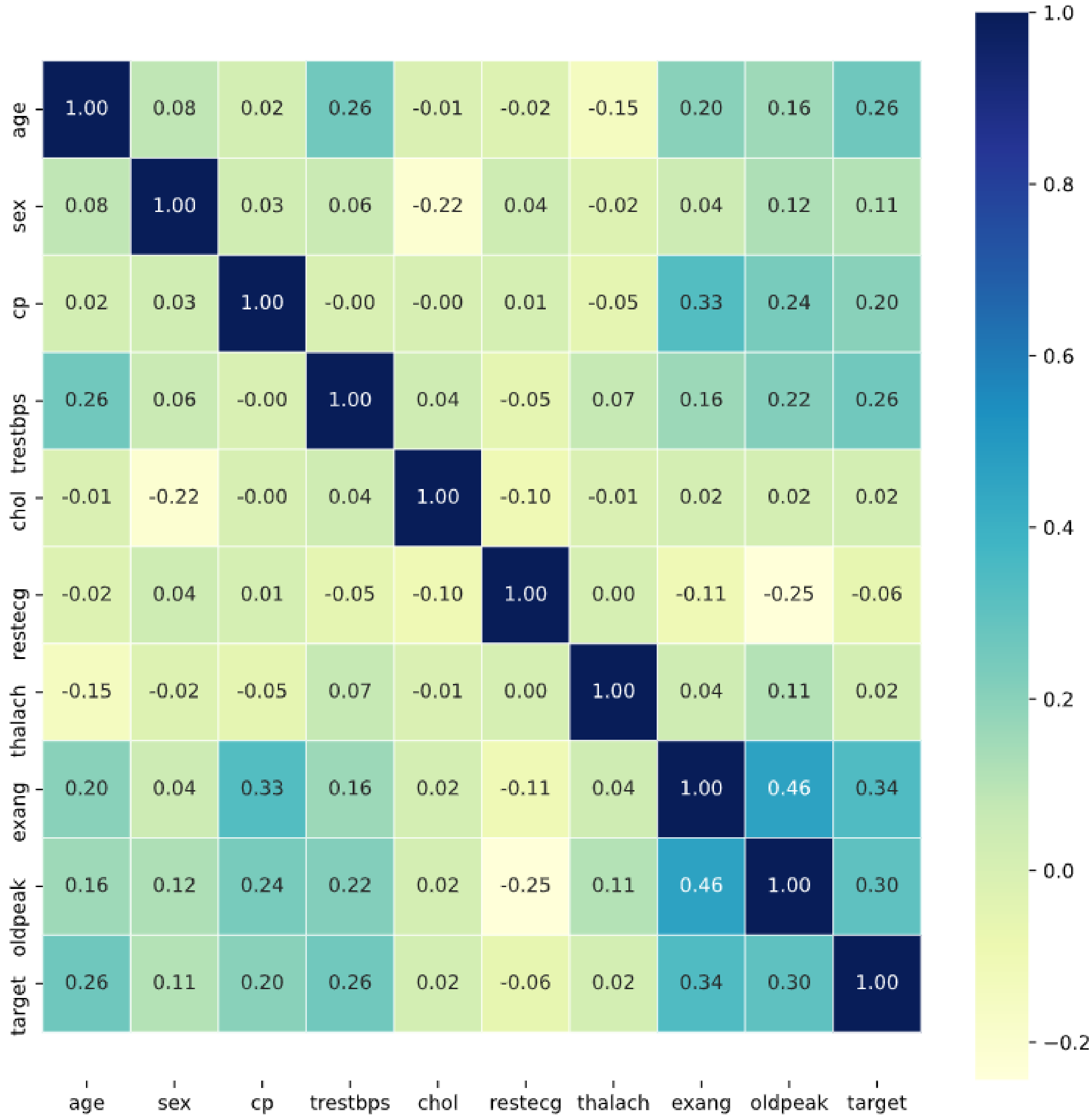


Fig 4.6 Correlation Matrix of Dataset 4

4.1.11 Correlation with target of Dataset 1

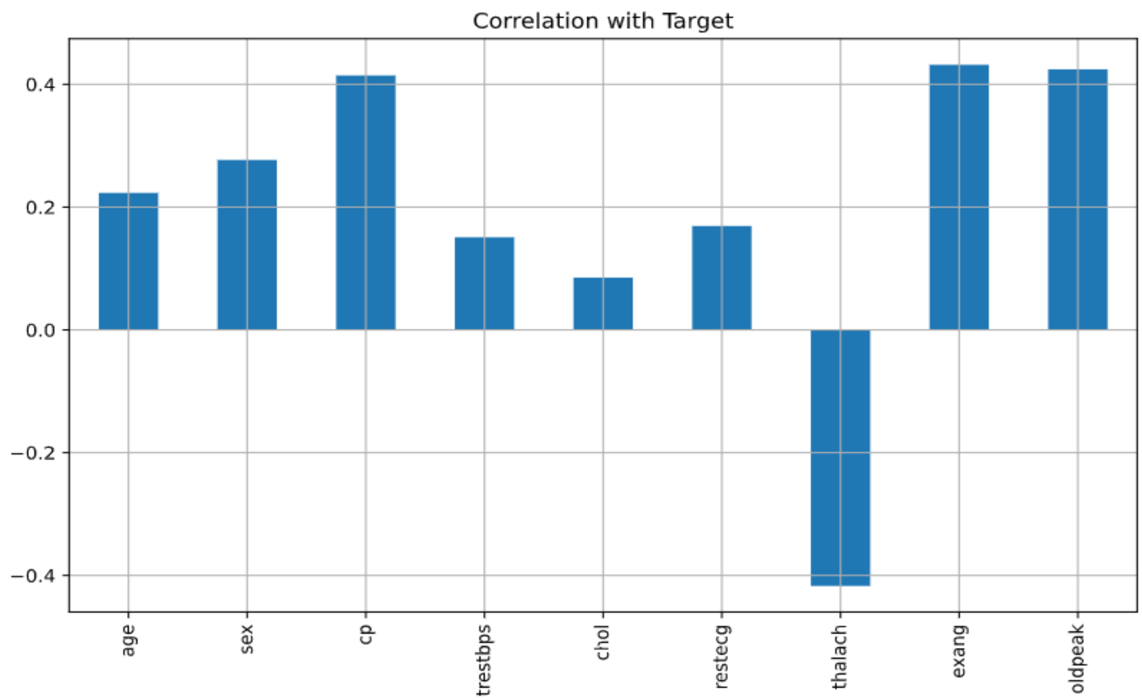


Fig 4.7 Correlation with target of Dataset 1

4.1.12 Correlation with target of Dataset 4

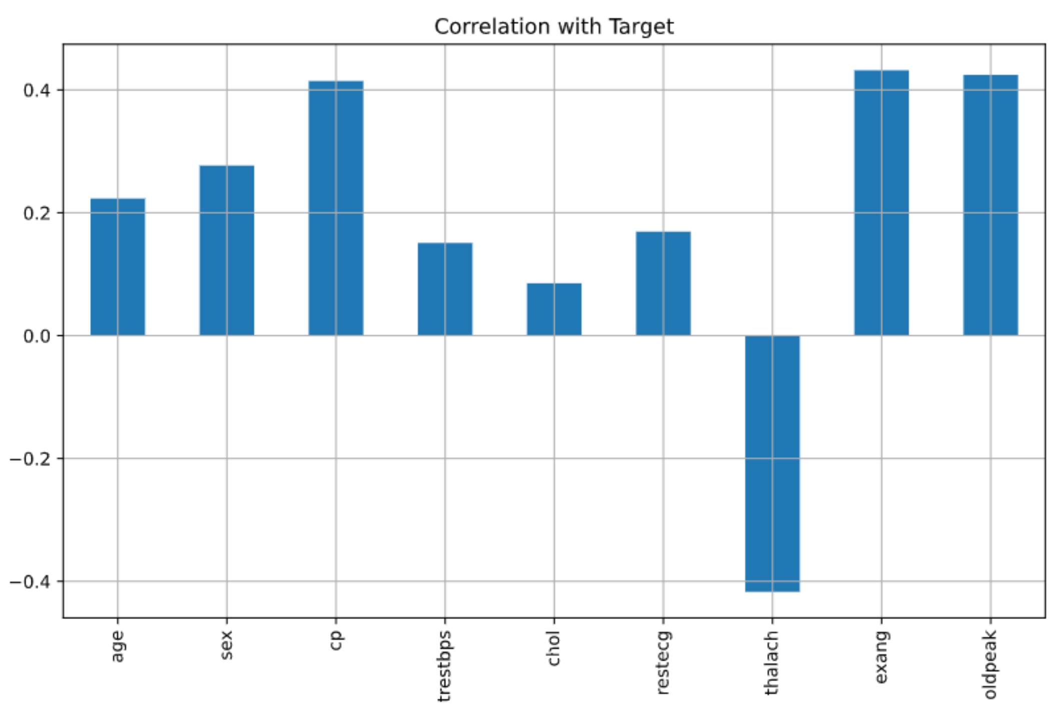


Fig 4.8 Correlation with target of Dataset 3

4.1.13 Correlation with target of Dataset 4

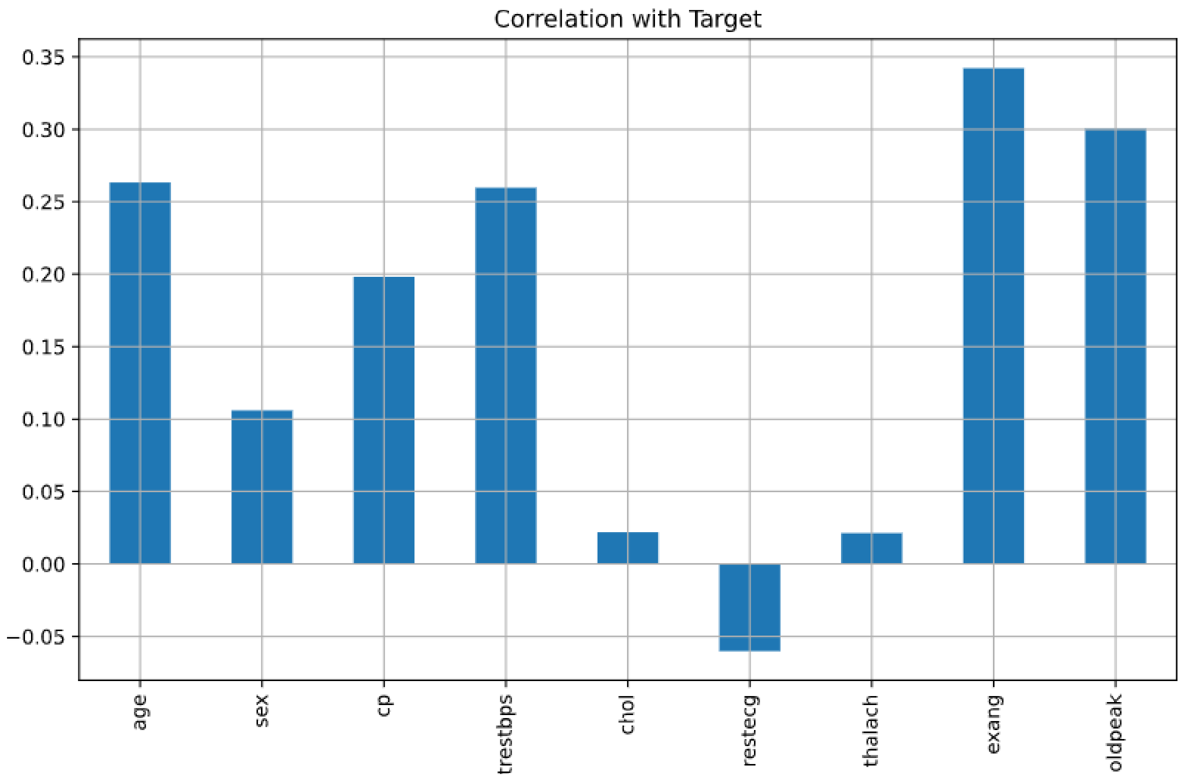


Fig 4.9 Correlation with target of Dataset 4

4.2 Discussion

- In stage 1 we have discussed about datasets ,previously we took dataset which was already preprocessed from kaggle. We also discussed about how preprocessing can be done.Later in our code we have also shown about the preprocessing steps . We have checked whether the dataset is balanced or not.

- We discussed about various attribute such as age,sex,cp etc. We plotted graphs of attributes with respect to target .From that graphs we understood what factors are responsible for heart disease

- Different machine learning algorithms such as knn,svm,logistic regression were discussed and we tried to implement this algorithms to our datasets.

- Before implementing such algorithms we have seen how training ,splitting and testing of datasets take place . After implementing various algorithms we did comparisons which algorithm is giving best accuracy for our dataset.

- In stage 2 we performed above steps with multiple datasets. We have also discussed about pickle in stage 2. We have shown in our code how we use pickle to store accuracy .

Chapter 5 CONCLUSION AND FUTURE SCOPES

5.1 Conclusion

After having availability of such a huge research facilities, targeting development as well as deployment of decision support systems for better understanding of medical services and clinical research using various data mining models and systems, still we are somewhere stuck to identify that icebreaking point where we will define our field of data mining and machine learning in the list of our expertise.

Data mining contributed in pattern finding by using analytical tool set that helps in finding patterns, while predictive analytics used in this project will provide the answers to the next big decision based on the results. In this project, we were trying to build a framework that'll help to generate an automated decision model for better result delivery and continue efforts of researchers of this field will definitely build such a system with maximum efficiency.

Hungarian dataset clinical markers like resting ECG, cholesterol level and heart rate played an important role in depicting heart disease while echocardiogram data set on the basis of clinical markers predicted about the mortality of heart patients.

In Medical Science, various factors like age, sex, lifestyle, mental status, eating habits etc. are root cause that interrupts in our biological functioning of body and slowly leads to heart disease that increases the possibility of having sudden cardiac death.

The methods of machine learning has been discussed and implemented in this project are such as KNN(K-nearest neighbors) , SVM (support vector machine),decision tree, logistic regression and random forest and an alytical comparison has been done for finding out best available algorithm for medical dataset.

So ,we deveeloped a cardiac arrest prediction model in the Emergency department using machine learning and sequential characteristics .T he model was validated for clinical usefulness using chronological visualization focused on clinical usability.

In this project we have summarized state of art techniques and available methods for prediction of the cardiac arrest. we have taken multiple data set and processed it with multiple machine learning algorithms and concluded the best way to predict the cardiac arrest in patient in his/her early stage.

Future Scope

The future work for this project holds the great potential for researchers to analyze various data mining algorithms with their variants in large amount of data. Hence, this gap further widens the opportunities for researchers to work in the proper treatment .. In a nutshell, we can conclude that SCD(Sudden Cardiac Death) is most dangerous situation for both heart disease and no heart disease patients that depends on the medical history as well as life style factors of a human being and the process of decision making and predictions will be made more concise and precise by using predictive analytics to generate automated decision based model.

At some point in future, the machine learning model will make use of a larger training dataset, possibly more than a million different data points maintainedin electronic health record system. Although it would be a huge leap in terms of computational power and software sophistication but a system that will workon artificial intelligence might allow the medical practitioner to decide the best suited treatment for the concerned patient as soon as possible . A software API can be developed to enable health websites and apps to provide access to the patients free of cost.The probability prediction would be performed with zero or virtually no delay in processing .

REFRENCES

1) Jyoti Soni, Ujma Ansari, Dipesh Sharma and Sunita Soni, 'Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction', International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March 2011.

2) Shantakumar B. Patil, Dr. Y. S. Kumaraswamy, 'Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction', International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March 2011.

3) Anudeep Duba, Rajasekhar Reddy, 'Decision Making System using Machine Learning and Pearson for Heart Attack', International Conference on Electronics, Communication and Aerospace Technology ICECA 2017.

4) Cheryl Ann Alexander, Lidong Wang, 'Big Data Analytics in Heart Attack Prediction', Article in Journal of Nursing & Care · January 2017.

5) Asha Rajkumar, Mrs. G. Sophia Reena, 'Diagnosis of Heart Disease Using Datamining Algorithm', Vol. 10 Issue 10 Ver. 1.0 September 2010 Global Journal of Computer Science and Technology.

6) Dr. Hidayet TAKCI, 'Improvement of heart attack prediction by the feature selection methods', Turkish Journal of Electrical Engineering & Computer Sciences, Turk J Elec Eng & Comp Sci (2018).

7) Anil Maybhate, Cheng Chen, Yama Akbari, David L. Sherman, Kaiquan Shen, Xiaofeng Jia, Nitish V. Thakor, 'Band Specific Changes in Thalamocortical Synchrony in Field Potentials after Cardiac Arrest Induced Global Hypoxia', 35th Annual International Conference of the IEEE EMBS Osaka, Japan, 3 - 7 July, 2013.

8) Junyun He, Hongyang Lu, Ruoxian Deng, Leanne Young, Shanbao Tong, Xiaofeng Jia, 'Real-time monitoring of cerebral blood flow by laser speckle contrast imaging after cardiac arrest in rat'.

9) Hsiao-Ko Chang, Cheng-Tse Wu, Ji-Han Liu, Wee Shin Lim, Hui-Chih Wang, Shu-I Chiu, Jyh-Shing Roger Jang, 'Early Detecting In-Hospital Cardiac Arrest Based on Machine Learning on Imbalanced Data'.

10) Usman Rashed, Muhammad Javed Mirza, ‘Identification of Sudden Cardiac Death Using Spectral Domain Analysis of Electrocardiogram (ECG)’, 2008 International conference on Emerging Technologies IEEE-ICET 2008.

11) Keisuke Kasahara, Masahito Shiobara, Saya Nakamura, Koichiro Yamashiro, Kazuo Yana Takuya Ono, ‘Sudden Cardiac Arrest Risk Stratification based on 24-hour Holter ECG Statistics’.

12) Archana Singh Rakesh Kumar "Heart Disease Prediction Using Machine Learning Algorithms" 2020 International Conference on Electrical and Electronics Engineering (ICE3-2020)

13) Senthil Kumar Mohan, Chandrasekar Thirumalai and Gautam Srivastva, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques” IEEE Access 2019.

14) Pahulpreet Singh Kohli and Shriya Arora, “Application of Machine Learning in Diseases Prediction”, 4th International Conference on Computing Communication and Automation (ICCCA), 2018.

15) Himanshu Sharma and M. A. Rizvi, “Prediction of Heart Disease using Machine Learning Algorithms: A Survey” International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 8, IJRITCC August 2017.

16) M. Sultana, A. Haider, and M. S. Uddin, “Analysis of data mining techniques for heart disease prediction,” 2016 3rd Int. Conf. Elect. Eng. Inf. Comm. Technol. iCEEiCT 2016, 2017.

17) Amin Ul Haq A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms Hindawi Mobile Information Systems Volume 2018, Article ID 3860146, 21 pages

18) A. Q. Ansari et. al.," Automated Diagnosis of Coronary Heart Disease Using Neuro-Fuzzy Integrated System", 2011 World Congress on Information and Communication Technologies 978-1-4673-0125-1@ 2011 IEEE (pp 1383-1388)

19) Mrs. G. Subbalakshmi, Mr. M. Chinna Rao “Decision Support in heart disease prediction system using naïve bays”, IJCSE Indian journal of computer science and engineering, ISSN: 0976-5166 Vol. 2 No. 2 Apr-May 2011

20) Mai Shouman, Tim Turner, and Rob Stocker, “Applying k-Nearest Neighbor in Diagnosing Heart Disease Patients”, International Journal of Information and Education Technology, Vol. 2, No. 3, June 2012

21) Ashish Kumar Sen, Shamsher Bahadur Patel, Dr. D. P. Shukla, “A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level”, International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 2 Issue 9 Sept., 2013 Page No. 2663-2671

22) Aditi Gavhane et al "Prediction of Heart Disease Using Machine Learning" Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1

23) Jayshril S. Sonawane, D. R. Patil "Prediction of Heart Disease Using Multilayer Perceptron Neural Network" ICICES2014 - S. A. Engineering College, Chennai, Tamil Nadu, India ISBN No.978-1-4799-3834-6/14/$31.00©2014 IEEE

24) Yubin Park and Joyce C. Ho and Joydeep Ghosh3 "Multivariate Temporal Symptomatic Characterization of Cardiac Arrest" 35th Annual International Conference of the IEEE EMBS Osaka, Japan, 3 - 7 July, 2013

25) Ali Bahrami Rad et al "NEAREST-MANIFOLD CLASSIFICATION APPROACH FOR CARDIAC ARREST RHYTHM INTERPRETATION DURING RESUSCITATION" 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)

26) N. Suresh "Improved Performance of FFT Based Cardiac Analyzer Using Advanced Booth Algorithm" International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB16)

27) Enrico G. Caiani, Giulia Marelli, Maria A. Brovelli Politecnico di Milano Milan, Italy, Carolina Arias Muñoz Universidad de San Buenaventura, Pietro M Brambilla, Andrea Pagliosa, Giovanni Cesana Azienda Regionale Emergenza

Urgenza – AREU Milan Italy: 'Development of a Health Geomatics analysis framework to evaluate cardiac arrests in Lombardy', January 2017

28) Dieter Bender, Ph.D. Candidate- Villanova University, Dr. Ryan W. Morgan- Attending Physician of Critical Care Medicine, Children's Hospital of Philadelphia, Dr. Vinay M. Nadkarni-Endowed Chair of Critical Care Medicine, The Children's Hospital of Philadelphia and Dr. C. Nataraj, Moritz Endowed Professor, Villanova University: "Classification of Asphyxia & Ventricular Fibrillation Induced Cardiac Arrest For Cardiopulmonary Resuscitation "year 2017.

29) Ji-Han Liu- Graduate Institute of Networking and Multimedia National Taiwan University Taipei, Taiwan, Hui-Chih Wang - Department of Emergency Medicine National Taiwan University Hospital Taipei, Taiwan, Hsiao-Ko Chang, Cheng-Tse Wu, Wee Shin Lim, Jyh-Shing Roger Jang -Department of Computer Science and Information Engineering National Taiwan University Taipei, Taiwan, 'Machine Learning Based Early Detection System of Cardiac Arrest', year 2019.

30) Nan Liu, Zhiping Lin, Senior Member, IEEE, Jiuwen Cao, Zhixiong Koh, Tongtong Zhang, Guang-Bin Huang, Senior Member, IEEE, Wee Ser, Senior Member, IEEE, and Marcus Eng Hock Ong, "An Intelligent Scoring System and Its Application to Cardiac Arrest Prediction "year 2012

31) M. Reza Pazhouhandeh, Omid Shoaei, Roman Genov -Intelligent Sensory Microsystems Laboratory, ECE Department of University of Toronto, Toronto, Canada 2 Bio-Integrated Systems Laboratory, ECE Department of University of Tehran, Tehran, Iran- "Two-electrode Impedance-sensing Cardiac Rhythm Monitor for Charge-Aware Shock Delivery in Cardiac Arrest "year 2017.

32) Sombat Muengtaweepongsa MD, Kesorn Suwanprasert Ph.D. Faculty of Medicine, Thammasat University (Rangsit Campus) Pathumthani, Thailand, Anon Jantanukul Medical engineering graduate program Thammasat University (Rangsit Campus) Pathumthani, Thailand "Suppression of Autonomic Drive Determined by Nonlinear HRV analysis in Therapeutic Hypothermia after Cardiac Arrest "year-2013 Biomedical Engineering International Conference.

33) Vipul Chauhan, Amit Kumar, Sumit Tiwary, Vikas Kumar HM ,Utsav Chauhan HMR Institute of Technology and Management Delhi, India ,"Cardiac Arrest Prediction using Machine Learning Algorithms " year-2019.

34) Joyce C. Ho and Yubin Park,” Learning from Different Perspectives: Robust Cardiac Arrest Prediction via Temporal Transfer Learning “year -2017

35) CHIH-WEI SUNG, JIANN-SHING SHIEH, WEI-TIEN CHANG, YI-WEI LEE , JI-HUAN LYU1 , HOOI-NEE ONG , WEI-TING CHEN, CHIEN-HUA HUANG , WEN-JONE CHEN , AND FU-SHAN JAW “Machine Learning Analysis of Heart Rate Variability for the Detection of Seizures in Comatose Cardiac Arrest Survivors “ year - July 30, 2020.

36) Andoni Elola, Elisabete Aramendi, Member IEEE, Unai Irusta, Member IEEE, Per Olav Berve, Lars Wik “Multimodal algorithms for the classification of circulation states during out-of-hospital cardiac arrest “year-2020.

37) Yosuf ElSaadany, AKM Jahangir A. Majumder, Donald R. Ucci Department of Electrical and Computer Engineering, Miami University, Oxford, OH, USA “A Wireless Early Prediction System of Cardiac Arrest through IoT “year -2017 IEEE 41st Annual Computer Software and Applications Conference “.

38) C. Jenefar Sheela .PG Scholar- Loyola Institute of Technology Palanchur, Chennai, L. Vanitha. Assistant Professor -Loyola Institute of Technology, Palanchur, Chennai. “Prediction of Sudden Cardiac Death Using Support Vector Machine “year -2014 International Conference on Circuit, Power and Computing Technologies [ICCPCT]”.

39) César O. Navarro, Nick A. Cromie, Colin Turner, Omar J. Escalona, Member IEEE, John McC Anderson “Detection of Cardiac Arrest using a Simplified Frequency Analysis of the Impedance Cardiogram recorded from Defibrillator Pads .presented in 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA “ year-August 30 - September 3, 2011.

40) British Heart Foundation - https://www.bhf.org.uk/informationsupport/how-a-healthy-heart-works

41) Medical News Today - https://www.medicalnewstoday.com/articles/168266

42) NHS inform - https://www.nhsinform.scot/illnesses-and-conditions/heart-and-blood-vessels/about-the-heart/understanding-how-your-heart-functions

43) American Heart Association - https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings

NCBI - https://www.ncbi.nlm.nih.gov/books/NBK541010/

44) Healthline.com - https://www.healthline.com/health/coronary-artery-spasm

45) https://www.javatpoint.com/k-nearest-neighbor-algorithm-for- machine-learning

46) https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

47) https://scholar.google.co.in/scholar?q=prediction+of+heart+disease+using+machine+learning+algorithms&hl=en&as_sdt=0&as_vis=1&oi=scholart

48) https://www.kaggle.com/faressayah/predicting-heart-disease-using-machine-learning

49) https://www.researchgate.net/publication/326733163_Prediction_of_Heart_Disease_Using_Machine_Learning_Algorithms

50) https://www.javatpoint.com/machine-learning

51)