

UNIVERSITY OF HERTFORDSHIRE

School of Physics, Engineering and Computer Science

MSc Data Science and Analytics with Advanced Research

7COM1075 –0501-2020: Data Science and Analytics Masters Project

**Final Project Report**

03/12/2021

**A Symptom-based Predictive Model to Analyze and Detect Heart Attack**

Name: Tharun Divvela

Student ID:19023017

Supervisor: **Eamonn Ramsay**

**Abstract**

Heart attack is one of the most widely seen health issues in the world. Many people die because of this heart attack. Different symptoms and underlying causes are there for heart attack, which people do not generally know. Because of the fact, they ignore those symptoms like diabetes, chest pain, etc., that cause heart attack to those people. So, the understanding and identification of important symptoms are necessary to determine the heart attack. In this project, the symptom-based predictive model has been proposed through which the heart attack can be determined by emphasizing the symptoms. In this project, the classifiers from machine learning and deep learning have been chosen and applied to the heart attack dataset that has been collected from Kaggle. By applying all six classifiers, it has been observed that Convolutional Neural Network has produced 91.8% accuracy for predicting a heart attack, which is higher than the previous models. Therefore Convolutional Neural Network will be selected as the best classifier regarding the higher accuracy, f1-score, precision, and recall.

**MSc Declarations**

This report is submitted in partial fulfilment of the requirement for the qualification Master of Science in Data Science and Analytics at the University of Hertfordshire (U.H.).

It is my true self work except where indicated in the report. I did not allow any sort of human participants in my MSc Project. I hereby give my full permission for the report to be made available on the university website on a condition that the source is acknowledged genuinely.

**Acknowledgement**

I express my utmost gratefulness to my supervisor Mr. Eamonn Ramsay for his unceasing support and guidance throughout my project journey with valuable feedback which assisted me for the completion of all my tasks on time thereby completing this dissertation successfully.

I am extremely thankful to my parents who always strive for my prosperity, their exquisite support, motivation and encouragement in times of deprivation throughout my masters period. Without them, I won’t be able to see myself accomplishing a Data Science project and writing this thesis. This thesis is truly dedicated to them.

**List of Abbreviations**

CNN: Convolutional Neural Network

SVM: Support Vector Machine

KNN: K-Nearest Neighbours

H.F: Heart Failure

CAD: Coronary Heart Disease

MLP: Multi-Layer Perceptron Classifier

IHD: Ischemic Heart Disease

SGD: Stochastic Gradient Descent

A.I: Artificial Intelligence

**List of Figures**

[**Figure 1:Heart Disease Dataset** 22](#_Toc89436161)

[**Figure 2:Methodology for the Detection of Heart Attack** 29](#_Toc89436162)

[**Figure 3:The tendency of Heart Attack by Age Factor** 30](#_Toc89436163)

[**Figure 4:The tendency of Heart Attack by Resting Blood Pressure** 31](#_Toc89436164)

[**Figure 5:The tendency of Heart Attack by Blood Cholesterol** 31](#_Toc89436165)

[**Figure 6:The tendency of Heart Attack by Maximum Heart Rate** 31](#_Toc89436166)

[**Figure 7:The tendency of Heart Attack by Gender** 32](#_Toc89436167)

[**Figure 8:Correlation of Features with Outcome** 32](#_Toc89436168)

[**Figure 9:Selection of Features from Correlation** 33](#_Toc89436169)

[**Figure 10:Data Information before Resampling** 33](#_Toc89436170)

[**Figure 11:Data Information after Resampling** 34](#_Toc89436171)

[**Figure 12:Data Segregation** 34](#_Toc89436172)

[**Figure 13:Outcome of Classification by Logistic Regression** 35](#_Toc89436173)

[**Figure 14:Outcome of Classification by Naïve Bayes** 36](#_Toc89436174)

[**Figure 15:Outcome of Classification by Passive Aggressive Classifier** 37](#_Toc89436175)

[**Figure 16:Outcome of Classification by SGD Classifier** 38](#_Toc89436176)

[**Figure 17:Outcome of Classification by MLP Classifier** 39](#_Toc89436177)

[**Figure 18:Model for Convolutional Neural Network** 40](#_Toc89436178)

[**Figure 19:Accuracy Comparison for CNN** 40](#_Toc89436179)

[**Figure 20:Model Loss Comparison for CNN** 40](#_Toc89436180)

[**Figure 21:Model Loss Comparison for CNN** 41](#_Toc89436181)

[**Figure 22:Comparison of Precisions and Recalls** 41](#_Toc89436182)

[**Figure 23:Comparison of Accuracies and F1-Scores** 42](#_Toc89436183)

**List of Tables**

[**Table 1:Structure of Confusion Matrix** 27](#_Toc89379667)

Table of Contents

[1 Chapter-1 Introduction 8](#_Toc89379629)

[1.1 Heart Attack 8](#_Toc89379630)

[1.1.1 Symptoms 9](#_Toc89379631)

[1.1.2 Heart Attack Detection 10](#_Toc89379632)

[1.2 Aim of Project 11](#_Toc89379633)

[1.3 Research Questions 11](#_Toc89379634)

[1.4 Objectives of Project 11](#_Toc89379635)

[1.5 Issues Regarding the Porject 12](#_Toc89379636)

[1.5.1 Issues for Ethical and Legal Aspects 12](#_Toc89379637)

[1.5.2 Issues for Professional Aspects 12](#_Toc89379638)

[1.5.3 Social Issue 12](#_Toc89379639)

[2 Chapter-2 Literature Review 14](#_Toc89379640)

[2.1 Previous Researches on Disease Detection 14](#_Toc89379641)

[2.2 Previous Researches on Heart Attack Detection using Machine Learning 17](#_Toc89379642)

[2.3 Comparison of Researches 20](#_Toc89379643)

[3 Chapter-3 Methodology 22](#_Toc89379644)

[3.1 Data Selection for Project 22](#_Toc89379645)

[3.1.1 Data Creators 22](#_Toc89379646)

[3.1.2 Data Features and Discussions 23](#_Toc89379647)

[3.2 Tool Selection for Project 24](#_Toc89379648)

[3.3 Algorithms Selection of Project 25](#_Toc89379649)

[3.4 Metrics Selection for Project 27](#_Toc89379650)

[3.5 Proposed Methodology 29](#_Toc89379651)

[4 Chapter-4 Analysis and Result 30](#_Toc89379652)

[4.1 Analysis of Heart Attack Data 30](#_Toc89379653)

[4.2 Heart Attack Prediction 32](#_Toc89379654)

[4.2.1 Feature Selection 32](#_Toc89379655)

[4.2.2 Data Resampling 33](#_Toc89379656)

[4.2.3 Creating Data Segments 34](#_Toc89379657)

[4.2.4 Prediction Using Machine Learning 34](#_Toc89379658)

[4.2.5 Prediction using Deep Learning 39](#_Toc89379659)

[4.3 Comparison of Performances 41](#_Toc89379660)

[5 Chapter-5 Discussion and Conclusion 43](#_Toc89379661)

[5.1 Addressing the Research Questions 44](#_Toc89379662)

[5.2 Limitation of the Project 44](#_Toc89379663)

[5.3 Future Scope 44](#_Toc89379664)

[6 References 45](#_Toc89379665)

[7 Appendix 48](#_Toc89379666)

# Chapter-1 Introduction

Time plays a crucial role in the diagnosis of a disease in all medical procedures. To acquire accurate predictions, a large set of data is required. Heart attack has a significant role in all medical procedures sit it helps in making final decisions about life's last phase. A good prognostication helps determine the correct cause of disease and facilitates medical procedures to provide better healthcare facilities. The Advanced Care Planning (ACP) system helps patients in making decisions about their future treatments. A well-designed ACP process helps enhance the quality of life of a patient and reduces the death rate by timely providing healthcare services (Chaves & Ramírez, 2010). However, to design the Advanced Care Planning process, accurate prediction of heart attack is essential for a physician. Therefore, this research project aims to study data science techniques to predict medical procedures and diseases.

Different types of diseases are there for which people badly suffer. One such disease is heart disease that leads a person to a heart attack if the reason behind the disease cannot be detected. Predicting the disease is more critical than treating certain diseases, as several may have similar symptoms (Geetha & Arunachalam, 2021). For example, if a person has a disease like pneumonia, he may face chest pain and headaches. However, it does not mean that the person has heart disease caused by blood clotting in arteries and veins. So, the type of chest pain can tell the main reason for the disease. Thus, the detection of a specific disease is more important by considering the symptoms. The predictive model will be designed to detect heart attack using the signs (A. S. Abdullah, 2012).

## Heart Attack

Heart attack is one of the most significant diseases for which thousands of people die yearly (BHF, 2021). The leading cause of a heart attack is the blockage of the blood in arteries. This will prohibit oxygen flow to the entire body and the flow of carbon-di-oxide to be extracted from the human body’s cells. In most cases, blood clotting is caused by the building up of fats and cholesterols in the arteries. Due to the storage of those materials, the blood cannot pass, and the essential operations in the human body will stop immediately. This causes a sudden heart attack and leads to stroke. In a survey of the British Council, the following facts have been observed (BHF, 2021):

1. Most of the deaths in the U.K. are caused by heart and circulatory diseases. A survey has been detected that more than 160 thousand deaths are generated each year, with an average of 460 people each day. Thus, heart attack and cardiovascular disease are concerned to be the most effective and dangerous diseases.
2. Approximately 7.6 million people in the U.K. live with heart disease, out of which 4 million are men, and the rest 3.6 million are women.
3. Coronary heart disease is the most common category of heart disease that can be seen in the U.K. This disease is one of the most common causes of a heart attack. Coronary heart disease was the biggest killer of people worldwide by affecting people with a heart attack in 2019.
4. More than 100,000 admissions in the hospitals can be seen in the U.K. every five minutes with cases of heart attack and other relevant diseases closely related to a heart attack.
5. Approximately 1.4 million people in the U.K. have survived a heart attack in the U.K. 2021 to date.
6. Approximately more than 900 thousand people are living with heart failure in the U.K.
7. Strokes cause around 34,000 deaths in the U.K. each year and are the most significant cause of severe disability.
8. In the U.K., vascular dementia is developed by the advancement of coronary heart disease, for which most people in the U.K. are affected by cardiac arrests. With the increase in the number of cardiac arrests, many people die without treatment, as for the shortage in the hospital. The survival rate from this disease is 1 to 10.
9. More than 4.1 million adults in the U.K. have been diagnosed with diabetes, with thousands more undiagnosed.

### Symptoms

Several symptoms can be observed for heart disease, and some of the most prominent symptoms for heart disease are as follows:

1. Discomfort in breathing
2. Feeling pressure and heaviness in the chest while taking a breath
3. Chest pain
4. Back pain and discomforts in the jaw, throat, and arm
5. Heartburn
6. Heart Choaking
7. Significant Indigestion in everyday life
8. Weakness and Anxiety
9. Observable fatigue and the unwillingness to work
10. Fast or uneven heartbeat

### Heart Attack Detection

**Role of Machine Learning**

Machine learning supports different supervised and unsupervised learning algorithms that can help in the handling of a large amount of data set to generate meaningful results as the medical sector keeps the record in the form of medical images, videos, reports, or signals which can be analyzed through ML techniques to help the decision-making process. Data produced by different medical sectors can be structured or unstructured, which can be harnessed using ML algorithms to extract useful information to aid advance medical procedures (Akshai & Anitha, 2021). A.I. has been assisting many medical procedures in detecting abnormalities. Modern ML techniques can help refine the information processing capabilities of the healthcare sector, such as analyzing and detecting the diseases like heart attacks and strokes. Natural Language Processing techniques have enabled the identification of synaptic structures, which help develop speech or text recognition systems for impaired persons. Although this is a separate issue, the application of Natural Language Processing will help analyse the comments made by the patients expressing their views on the diseases with which they have been affected (Chaves & Ramírez, 2010). ML has also been used in medical image analytics in Computer Tomography, Magnetic Resonance Imaging, X- rays, etc. These systems capture high-resolution medical images and, using image analytics, extract valuable patterns from these images that can help in decision making.

**Role of Deep Learning**

Deep learning is a subfield of artificial intelligence that consists of artificial neural networks that can act and work like the human brain. Deep learning models can detect complex structures by adding a hidden layer between the input and output layer to represent complex intermediatory solutions (Gogi & M.N., 2018). The research was conducted to find the one-year mortality rate of patients during their hospital admission through EMR data. A binary classification model was used to extract only structured data from clinical records. The data extracted was later used as an input for the deep learning model to make predictions. However, the adaptation of deep learning and other analytic tools in medical science is still limited due to the implementation of the EHR system. Many healthcare sectors are stock to traditional methods to keep the patient record. Therefore, in such scenarios predicting a patient’s disease is a trivial task (Geetha & Arunachalam, 2021).

## Aim of Project

The project aims to design a predictive model for the analysis and detection of a heart attack. In this project, the heart attack will be detected using data analysis and machine and deep learning. In this context, the data will be collected, and the analytics will be applied there. The choice of the features that are the symptoms will be detected using correlation. Those symptoms will be selected as most important, which have the higher correlation coefficient. Then by applying machine and deep learning classifiers, the heart attack will be predicted.

## Research Questions

The research questions for fulfilling the aim of the project and to achieving the goal of the project are as follows:

1. Can we precisely predict the heart attack using symptoms?
2. Which symptoms are more important to detect heart attack?
3. Which algorithm is best for the detection of heart attack based on symptoms?

## Objectives of Project

The objectives of the projects have been taken into consideration for making the project satisfactory and addressing the research question successfully. Hence, ten objectives for this project are as follows:

1. To review the existing model for the heart attack prediction and collect resources such as the method that the previous researchers have used along with the data on which they have worked.
2. To select the machine learning and deep learning models for the review
3. To select the data containing the records of the previous patients that have been under the test for heart attack
4. To analyze the selected data to identify the symptoms and the effect of those symptoms on heart attack
5. To detect heart attacks using the selected models and record the performances concerning the classification metrics such as accuracy, precision, recall, and the f1-scores.
6. To compare the metrics and determine the best model for detecting a heart attack.

## Issues Regarding the Porject

As the project uses the data dataset provided by the Kaggle repository to find the elements that affect the heart attack rate, therefore, there is no legal issue related to the project. However, the research project will cite the work of the other researchers to avoid copyright issues. Moreover, the project will not have any personal data and will only analyze the dataset obtained from Kaggle.

### Issues for Ethical and Legal Aspects

The issues are as follows:

1. The source of the data will be the open-source website from where it can be collected and used without any payment issues.
2. The free-source data will be considered as it will not incur any kind of legal isues like permissions.
3. The research will be conducted by own effort and no replication from other’s idea
4. If the approach or method is used in the research by taking from previous research, those will be referenced to authenticate.
5. The data will be collected so that it will be relevant to the present research.
6. The whole Research proposal and implementation were with my own knowledge.

### Issues for Professional Aspects

The issues are as follows:

1. The description id the project will be written in a well-structured report format.
2. The research will be done professionally so that valuable insights can be obtained and documented.
3. The approaches, if it will be taken from the existing papers, will be cited.
4. The source of data and the previous researches will be cited in the research document to avoid any copyright conflict.

### Social Issue

The issues are as follows:

1. The research will not impose any manual effort to collect the healthcare data.
2. The research will facilities the healthcare system by providing the symptom-based detection procedure of heart attack.

# Chapter-2 Literature Review

A literature review is one of the most important parts of every research work. It helps to develop the understanding of a symptom-based predictive model to analyze and detect a heart attack in an according way. It helps to evaluate important and significant research-related information to improve the initial research work. In this specific research work, the entire literature review section helps shed light on the Machine learning support processes. The primary machine learning techniques offer several supervised and unsupervised learning algorithms to help large data sets generate meaningful results. The following section involves a large number of related articles to increase the relevancy of the research work effectively.

## Previous Researches on Disease Detection

It is pretty apparent to highlight that several researchers already exist on disease detection. That helps to improve the entire work process of the investigation and examination. All of the major works are helping to increase the strengths of the research paper positively by involving the key and interrelated information. The following articles are trying to offer an overview of the previous researches in a relevant way.

According to Tarik et al. (2021), Potato is the most important crop in a country like Bangladesh. So, the entire process of potato farming has become very common in Bangladesh in recent days. But the process of potato farming is being vulnerable because of the diseases that are cumulative the cost of ranchers in potato construction. Nevertheless, some potato diseases are obstructing potato construction that is growing the worth of agriculturalists. That puts a direct and complete negative impact on the life of a farmer. So, it is essential to note that an automated and quick illness discovery procedure is needed to upsurge potato construction and digitise the prearrangement. The essential purpose of this article is to identify potato illness by applying leaf images that analyze via advanced machine learning knowledge. This specific article reflects the image is dealing out as well as machine learning depends upon an automated system that includes the potato leaf illnesses. That will be recognized as well. Image dealing out is the best explanation for sensing and evaluating these illnesses. The primary result of this article has revealed that machine learning surpasses all current responsibilities in potato illness discovery.

As stated by Korkut et al. (2018), The key purpose of this specific article is associated with the detection of plant diseases by machine learning in the present time. The article mainly reflects an automatic discovery of plant illnesses performed by applying the image handing out and machine learning approaches. In addition, the accurate and on-time discovery of plant illnesses is significant for crop superiority as well as yield relevancy. Moreover, the primary diagnosis, as well as involvement, can reduce the cost of plant illnesses as well as decrease needless drug application. On the other hand, the leaf imageries of diverse plant classes were gathered, along with feature withdrawal was done from the descriptions by applying the transfer learning technique. With numerous machine learning approaches, the planned model attained a 94% of accuracy rate.

As mentioned by Chitnis et al. (2020), It is essential to note that initial dental illnesses have been prevalent in this modern world. It is necessary to develop the entire dental treatment method. At present, maximum medical experts depend upon a manual examination of a patient's verbalized void for preliminary analysis. Apart from that, they also depend on the manual investigation of radiographs or x-rays for progressive analysis. To decrease this determination, arrangements are planned for illness finding methods working with x-rays or radiograms. Those are only available to dental experts. Some supports are also needed in the current times to repurpose the wide-ranging application of machine learning algorithms like CNNs for the specific action of illness discovery as well as sorting in medical imagination. Most importantly, the ground of dentistry can advantage significantly by focusing more research on visual light images, letting experts unburden the early evaluation of a patient to machines. The main aim of this review is to offer an all-inclusive investigation of currently planned machine learning which depends upon the dental illness detection arrangements along with proposals to what can be better-quality in the upcoming to offer a better understanding to investigators working in this area.

According to Bose et al. (2020), At first, it is essential to highlight that Hemp disease detection is a versatile process that has medicinal along with industrial worth. The plant is growing very effortlessly. It is equally easy to maintain below any environment. Though, like additional plants, Hemp illnesses influence plant developments as well as the reason for crucial financial damage in hemp construction. Along with the speedy progression of artificial intelligence as well as machine learning expertise, investigators have in progress applying data-driven machine learning methods in innovative cultivation as well as rural methods. Plant disease detection, as well as arrangement, is a process of the intelligent agriculture method. The essential purpose of this article is to highlight hemp disease detection by proposing one SVM-based machine learning model and three deep learning ensemble models. For that, the article implements pre-trained deep learning supportive models along with transmission learning. It reports qualified evaluation consequences of the three deep learning collaborative models with an SVM-based model with manual feature removal. The assessment consequences from diverse models reflect 98% of the total accuracy rate as well as a robust application.

As mentioned by Kohli & Arora (2018), In modern times, the implementation of machine learning in the ground of medical analysis is cumulative in a slow process. It can be donated chiefly to the development in the organization as well as acknowledgement arrangements that are applied in the present illness diagnosis that is capable of offering several pieces of information that support the medical specialists in the initial discovery of deadly illnesses. Therefore, upsurging of the existing rate of patients meaningfully is also affected by this. The main aim of this article is to reflect the different organizational processes. Each benefit from three distinct disease factors is obtainable in UCI source for disease detection and prediction. The upcoming choices for each dataset were skilled by regressive modelling that applies the p-value examination. The consequences of the study reinforce the knowledge of the experimentation of machine learning in the initial discovery of illnesses.

As per Khan et al. (2021), The entire discovery process of brain illnesses at an initial phase can generate an enormous alteration in endeavouring treatment to them. The application of artificial intelligence (A.I.) is rising and falling through all compasses of science. It is reforming the ground of clinical neurology. The application of artificial intelligence in medical science has completed brain illness forecasts and detection more specifically. The major findings reflect that the reviewed articles are brief as well as include several most important problems associated with the machine learning or deep learning approaches that depend upon the brain illness diagnostic methods, which are deliberated. This specific article tries to objectify the significant finding with the utmost accurate method for detecting diverse brain illnesses that can be active for forthcoming improvement.

As stated by Abdulhadi & Al-Mousa (2021), In the present situation, the entire process of diabetes detection by applying machine learning classification methods has become very much essential and impactful. The key purpose of this investigation is to forecast the probable existence of diabetes definitely in females at an initial period by applying diverse machine learning methods. Primary discovery of diabetes can meaningfully avoid the development of the illness along with decrease the danger of thoughtful difficulties like a heart as well as kidney diseases and generating the appropriate lifestyle variations at the right time can support evade diabetes and all the diseases related with it. The outcome of this article reflects an 82% of accuracy rate, which depends upon the random forest classifier model.

## Previous Researches on Heart Attack Detection using Machine Learning

Following the brain, the heart is among the most important components of the human body. The heart's principal duty is to pump blood to all regions of the system. Heart disease refers to any issue that might impair the heart's ability to operate. There are numerous forms of heart illness around the globe; the most frequent heart disorders are coronary artery disease (CAD) and heart failure (H.F.). The obstruction or shrinking of the coronary arteries is the cause of coronary heart disease (CAD). It is also in charge of delivering blood to the cardiac system. Coronary heart disease (CAD) has been the biggest cause of mortality worldwide, affecting approximately 26 million individuals (Katarya & Srinivas, 2020). Identifying and detecting cardiac disease has generally been a time-consuming and challenging undertaking for doctors. To treat cardiac disorders, hospitals and other clinics are giving costly treatments and surgeries. As a result, anticipating cardiac disease in its initial phases will be beneficial to individuals all around the globe, allowing them to take required treatment before it becomes severe. Heart disease has been a significant issue in recent years, with the primary causes being excessive alcohol use, cigarette use, and a shortage of physical activity. Machine learning has shown to be helpful in making choices and predictions from a large amount of data created by the healthcare business over time. This study by Katarya & Srinivas (2020) detailed several of the expert automated processes. Feature selection and prediction, according to the findings, are fundamental for any automated system. Researchers can improve their outcomes in predicting heart disease by selecting characteristics wisely. This research has compiled a list of methods that may be used to pick features, such as the hybrid grid search algorithm as well as the random search algorithm.

The purpose of this research of Tao et al. (2018) was to create a system for detecting and localizing ischemic heart disease that was both quick and accurate. Methods: T waves were separated from averaged Magnetocardiography (MCG) recordings, and 164 characteristics were retrieved after that. Time-domain characteristics, frequency domain elements, and information theory characteristics were divided into three categories. Furthermore, researchers examined k-nearest neighbour, decision tree, support vector machine (SVM), and XGBoost as machine learning classifiers. They chose three classifiers with the greatest efficiency and used a model ensemble to aggregate the results to detect ischemic heart disease (IHD) cases. This stage made use of all 164 characteristics. The researchers categorized IHD patients into three groups based on the site of stenosis: left anterior descending (LAD), left circumflex artery (LCX), and right coronary artery (RCA). The findings show that 1. T wave repolarization synchronization is a critical part of distinguishing IHD from healthy people. 2. The location of the stenosis is linked to the magnetic field pattern. The suggested machine learning technique offers doctors a quick and reliable diagnostic tool for interpreting MCG data, increasing its acceptability in clinics. Additionally, the magnetic pole features presented by the approach are linked to the ischemia site, allowing for noninvasive ischemia detection.

As stated by Atallah & Al-Mousa (2021), it can be stated that with the help of the voting ensemble method, understanding of the threat related to heart disease can be predicted properly. Furthermore, based on the understanding of the research report, it can be stated that heart disease in recent years has become one of the huge health threats which have the credibility of creating a long term impact on the human health condition. Moreover, to work on the predictability approach of heart disease identification taking help from the simple and affordable medical tests are being conducted in the local clinics. With the help of the selected research presentation, understanding how doctors can manage the entire diagnosis process based on the usage of the real-life data of both the healthy and ill patient, both of the data and information has the credibility of predicting the disease with utmost importance effectivity and efficiency. Depending on the research presentation, it can be stated that with the help of the machine learning method, the prediction method can be driven towards a proper direction. Therefore, with the help of the entire prediction model, around 89% accuracy rate can be provided, which takes help from the complex voting ensemble model.

As defined by Ed-Daoudy & Maalmi (2021), it can be mentioned the entire heart disease hugely influences the global death rate as a threat; therefore, based on the understanding of the research presentation, it can be stated that by predicting heart disease at an early stage the monitoring of the health issue can be mitigated on the other hand the mortality rate of the prediction technique can better the global health management concept in a positive direction. Moreover, the usage of wearable sensor devices in the Internet of Things health monitoring is increasing the usage of the growth of data management. Moreover, based on the understanding of the report, it can be stated that the streaming system and similar aspects have the credibility of generating a vast number of data regularly. In addition to the research report, it can be stated that in the case of the early detection of heart disease, the importance of streaming big data analytics and machine learning cannot be ignored as based on the attainment of a long term health value can be attained. In this report, a proposal related to a real-time heart disease prediction system based on Apache Spark has been given, which is standing on the large solid scale distributed computing platform. Therefore, depending on the entire system, the attainment of a better and accurate result (87%) is being achieved in the case of the prediction model.

As described by Keya et al. (2021), the heart is one of the most complicated organs of the human body, and due to this complicacy, if any part of the heart is damaged, the rest of the parts will keep operating. However, due to the damage, the heart will not be able to pump as much as blood it was supposed to, and it also weakens the organ's activities too. On the other hand, according to the selected research report, it can be stated that after heart attacks, actions like timely detection of multiple possible hamstring issues, proper care, and dietary changes can reduce the risk of being affected for the long term. At the same time, various algorithms such as logistic regression, random forest, bagging, MLP, and decision tree is used in the machine learning method to manage the entire prediction system with an accuracy of around 80 to 87%.

As mentioned by Raja et al. (2021), it is quite evident that heart diseases are one of the most threatening facts of the world health rate, and regarding this health issue, one person is losing their life per minute. Therefore, based on the understanding of the report, it can be stated that predicting the issue regarding heart disease is one of the most challenging approaches; therefore to deal with the issue taking help from machine learning techniques play a very important role, as with the help of this aspect early prediction of the disease becomes relatively easy and accurate. Random Forest Algorithm is being used in the machine learning system to enhance the reliability of the heart disease prediction system. They have got 85% accuracy in their experiment. Moreover, with the help of this aspect, enhancement of the entire accuracy margin of the system is gained at a positive rate as well.

As stated by Haq et al. (2021), the reliability of invoice based techniques as well as ordinary medical-based methods in the case of heart disease prediction cannot be trusted; moreover, based on the report result, it is quite clear that non-invoice based techniques are most effective in the entire prediction system. Therefore, with the help of the research presentation, the capability and value of different Machine Learning (ML) classifiers and deep learning classifiers are being understood from here. Moreover, six machine learning classifiers and BPNN have been selected here to diagnose heart disease accurately. On the other hand, BPNN has achieved around 88% accuracy rate depending on the entire system, while SVM has achieved around 87.45% accuracy margin.

As defined by Khan & Mondal (2021), by detecting heart disease at the early stage, the prevention of the progression of the illness can be managed. Moreover, with the help of the selected research report, learning about the importance of machine learning algorithms that can accurately predict heart disease. On the other hand, the selected research paper is trying to manage a hybrid dataset of the machine learning system to understand more about the techniques. Moreover, with the help of the result, it can be stated that two in method A randomly separated hybrid dataset and in method B one dataset is used as training the model and the other dataset for testing the model. and depending on the presentation, it is quite evident that method A has provided around 73.87% accuracy while method B has provided 58.41 per cent accuracy margin.

As mentioned by Ahmed et al. (2021), heart disease prediction and detection as a concept are one critical aspect of the current health management and predicting heart disease in the management of the early stages of the health care services (HCS) can be done in an accurate direction. On the other hand, in the case of the United States, heart disease is one of the prominent public chronic illnesses. It is influenced by different factors like consumption, bad lifestyle, lack of physical activity, and alcohol intake. Therefore, with the help of the selected project, an understanding of a cloud-based architecture has been used here that can better the entire health information prediction and tracking aspect. Furthermore, with the help of machine learning techniques and cloud-based 4- tier architecture, an Artificial Neural Network (ANN) management and accuracy of the disease prediction are being carried out properly. Finally, they have got 89.56% accuracy in the detection of a heart attack.

## Comparison of Researches

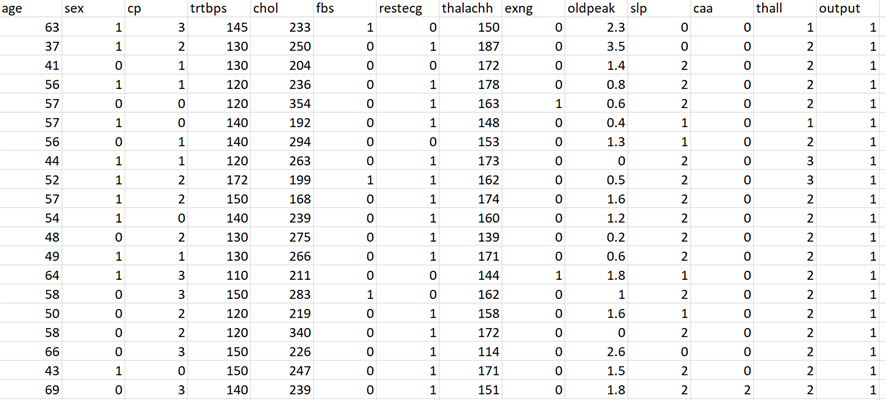
The review of the previous research reflects the idea of the application of machine learning and deep learning to detect disease. The study has been done from different angles and views that will reflect the aspects of detecting different types of diseases. It has been seen that Tarik et al. (2021), Bose et al. (2020) and Korkut et al. (2018) have applied deep learning classifiers to detect the disease of plants. On the other hand, the authors of the research, Kohli & Arora (2018), Katarya & Jain (2020), Khan et al. (2021) and Abdulhadi & Al-Mousa (2021), have focused on the detection of disease for humans. Katarya & Srinivas (2020), Tao et al. (2018) and Atallah & Al-Mousa (2021) have applied machine learni8ng classifiers to detect a heart attack. Keya et al. (2021), Raja et al. (2021), Haq et al. (2021) and Khan & Mondal (2021) have used machine learning classifiers to detect heart disease, but they have obtained moderate accuracy below 90%. However, with the help of an Artificial Neural Network and by the combination of machine learning, the accuracy has been improved by Ahmed et al. (2021). But, the overall accuracy for the detection of heart disease has been seen to be less than 90% which is moderate. So, in this research, the symptoms will be emphasized to detect heart disease at an early stage.

# Chapter-3 Methodology

In this chapter, the method of executing the project will be discussed, along with the focusing of the algorithms to be used here. The chapter will be discussing the application process of classifiers of machine learning, and deep learning and the description of each of the selected classifiers will be given here (Khan, et al., 2021). Now, for the application of the classifiers, the data will be chosen and described in this chapter regarding the features and the symptoms in it.

## Data Selection for Project

The data has been collected for the detection of heart disease. The data mainly contains the symptoms for which the heart disease has been observed. However, the data shows both the people with and without heart disease corresponding to the symptoms. The data has been collected from the Kaggle data repository (Ronit, 2020). The data was originally uploaded to the UCI data repository, and four scientists created it. The data is shown below:



**Figure 1:Heart Disease Dataset**

### Data Creators

The data has been created by four scientists in different locations and with collaboration. The data had been gathered and framed by Andras Janosi from Hungarian Institute of Cardiology, William Steinbrunn from University Hospital, Matthias Pfisterer from University Hospital and Robert Detrano from Long Beach and Cleveland Clinic Foundation and donated by E. David.

### Data Features and Discussions

The original dataset contains 76 attributes that are the symptoms of a heart attack. Later for the feasibility of research, only 13 important attributes, along with the attribute with the name “output”, have been taken into consideration. So, the stored 13 attributes of features of the data are the selected symptoms of the heart attack based upon which the heart attack can be determined (Kamata, et al., 2015). The present data contains the observations of 303 patients. The detailed attributes of the data are discussed below:

* age: This is one of the predictor features of the data and signifies the age of the patient who was under the test for the determination of heart attack.
* sex: This is one of the predictor features of the data and signifies the gender of the patient who was under the test for the determination of heart attack.
* cp: This is one of the predictor features and symptoms of the data and signifies the type of chest pain of the patients under the test. Four different types of chest pain have been enlisted in the dataset based upon which the possibility of heart attack can be determined.
* trtbps: This is one of the predictor features and symptoms of the data and signifies the blood pressure at the resting condition of the patients who were under the test for the determination of heart attack.
* chol: This is one of the predictor features and symptoms of the data and signifies the value of cholesterol in the blood of the patient who was under the test for the determination of heart attack.
* fbs: This is one of the predictor features and symptoms of the data and signifies the blood sugar level in the body of the patients.
* restecg: This is one of the predictor features and symptoms of the data and signifies the electrocardiographic results at the resting condition of the patients who were under the test for the determination of heart attack.
* thalachh: This is one of the predictor features and symptoms of the data and signifies the maximum heart rate that has been achieved from the patients while at the time of testing.
* exng: This is one of the predictor features and symptoms of the data and signifies the value of angina while at the time exercise by the patients.
* oldpeak: This is one of the predictor features and symptoms of the data and signifies the S.T. depression level that has been induced by the exercise of the patients relative to resting condition.
* slp: This is one of the predictor features and symptoms of the data and signifies the slope of the peak exercise for the S.T. segment of the patient who was under the test for the determination of heart attack.
* caa: This is one of the predictor features and symptoms of the data and signifies the number of major vessels (has been labelled from 0-3) of the patient, and those have been coloured by fluoroscopy for ease of understanding.
* thall: This is one of the predictor features and symptoms of the data and signifies the defects of the heart of the patients, and it has been labelled by 1, 2 & 3 meaning normal heart, heart condition with fixed defect and heart condition with the reversible defect.
* output: This is the target feature of the dataset and labelled by 1 and 0 which means patients with heart attack and patients without a heart attack.

## Tool Selection for Project

The project will be aiming to detect heart disease based upon the symptoms. In this context, the method of classification will be used as discussed in the aim of the project and also found from the review of the literature. So, to operate, the application of machine learning is necessary (Khan & Mondal, 2021). In this viewpoint, Python 3 has been chosen as the tool because of the robustness and availability of libraries. Using Python 3, the execution for the detection of heart attack will be done. The reasons for choosing the tool ars as follows:

1. Python 3 is a programming tool that does not depend upon the operating system, and the code can be executed in any operating system without changing the code structures.
2. It is a dynamically types programming tool that helps the codes to attain the feasibility of writing the code.
3. Python 3 has good support of the online documentation of coding that helps the coder to get help if any issue is faced during the coding.
4. Python 3 has the provision of accessing the Jupyter Notebook, which is an interactive IDE for Python. It provides the output of the code based on the code snippets, which helps the coder to get the code response from each of the code segments.
5. Python 3 is free programming tool.

## Algorithms Selection of Project

The algorithms have been chosen to detect and predict heart attacks. As seen from the review of previous researches, the previous researchers have applied both the classifiers from machine learning and deep learning (Ismaeel, et al., 2015). By following this, the classifiers from machine learning and deep learning have been chosen here. The algorithms that will be chosen for the classification and detection of heart attacks are discussed below.

**Logistic Regression**

This is a binary classifier that is used to classify the labelled data. The meaning of a binary classifier is that this classifier is good for the prediction of the classes where the number of class labels is two (Geetha & Arunachalam, 2021). Logistic regression employs the sigmoid rule for classifying the data. This belongs to the supervised machine learning that is used to classify the labelled data. This is a supervised learning algorithm because it takes the data from the users that contain the target features and is based on which the classification will be done. Based on the result of classification, the final prediction will be made to detect whether a patient has a heart attack or not. This algorithm will be applied in this project for the detection of a heart attack (Atallah & Al-Mousa, 2021).

**Naive Bayes**

This is a classifier that works of the probability to find the predictive labels of then class using likelihood function and posterior probability. This is a multiclass classifier as well as a binary classifier that can classify the data with a binary class or multiclass. It will find the probability of each of the classes in the target feature and classify the data with that class having a higher probability (Elstad & Walløe, 2015). This is another supervised algorithm from machine learning that is applied for classification. The classification is done based on the predictor features, and the detection is done based on the target feature of the data. In this context, the classifier will be applied for the classification and detection of heart attacks of the patients. So, the symptoms will be taken as the predictor features, and the target attribute (output) will be taken as the target feature. The final classification and the prediction will be made based on the labels of the target feature (Chitnis, et al., 2020).

**Passive Aggressive Classifier**

This machine learning classifier is widely used to classify the data with binary class and multiclass. The main essence of this classifier is that it is used to hold the highest accuracy while checking and predicting the data. In some cases, this classifier produced very good precision over the classes (Bose, et al., 2020). This classifier belongs to the supervised machine learning that is used to classify the labelled data. This is a supervised learning algorithm because it takes the data from the users that contain the target features and is based on which the classification will be done. Based on the result of classification, the final prediction will be made to detect whether a patient has a heart attack or not. This algorithm will be applied in this project for the detection of a heart attack (Faiz, et al., 2019).

**SGD Classifier**

SGD stands for Scholastic Gradient Descent that is a member of linear models in machine learning. It is used to classify binary and multiclass data. It applied the method called gradient descent for classification. This is another supervised algorithm that is applied for the classification and detection of a heart attack (Katarya & Jain, 2020). The classification is done based on the predictor features, and the detection is done based on the target feature of the data. In this context, the classifier will be applied for the classification and detection of heart attacks of the patients. So, the symptoms will be taken as the predictor features to classify the target feature. The final classification and prediction will be made based on the labels of the target feature (Geetha & Arunachalam, 2021).

**MLP Classifier**

This is a neural network classifier that is used to classify binary data and multiclass data. It applied neural network architecture for classification. It has three significant sections, namely the input section (from where the data is inputted), the hidden layer (where the feature engineering and decision making is done) and the output layer (where the final output of the classification is obtained). Perceptron refers to the neural network, and the meaning of multilayered perceptron is the multi-layer neural network. This belongs to the supervised machine learning that is used to classify the labelled data (Islam, et al., 2020). This is a supervised learning algorithm because it takes the data from the users that contain the target features and is based on which the classification will be done. Based on the result of the classification, the final prediction will be made to detect whether a patient has a heart attack or not. This algorithm will be applied in this project for the detection of a heart attack (Bose, et al., 2020).

**Convolutional Neural Network**

This is a deep learning classifier that employes the neural network in two-dimensional architecture (Tao, et al., 2018). The main difference between MLP with Convolutional Neural Networks is that the Convolutional Neural Network has two-dimensional, whereas MLP has a one-dimensional architecture and is eventually known as an Artificial Neural network. The classification is done based on the predictor features, and the detection is done based on the target feature of the data. In this context, the classifier will be applied for the classification and detection of heart attacks of the patients (Akhtar, et al., 2021). So, the symptoms will be taken as the predictor features, and the target feature will be taken as the target feature. The final classification and the prediction will be made based on the labels of the target feature.

## Metrics Selection for Project

The discussion on the chosen classifiers has been done in the last section. Those classifiers will be applied to the heart attack dataset to predict the possibility of heart attack based on the symptoms (Chitnis, et al., 2020). The effectiveness of those algorithms will be determined using the following classification measures and metrics.

**Confusion Matrix**

A confusion matrix is one of the important measures of the classification operations through which the strength of the classifier can be judged. The confusion matrices are generated based on the actual target features in the test data with the predicted target features. The demonstrative structure for the confusion matrix is shown below:

**Table 1:Structure of Confusion Matrix**

|  |  |  |  |
| --- | --- | --- | --- |
|  | | **Prediction** | |
| **A** | **B** |
| **Actual** | **A** | AA | AB |
| **B** | BA | BB |

The A.A. and B.B. are referred to as the True Positive and Negative of the classification, respectively, and those are the correct prediction of the classifier. On the other hand, the A.B. and B.A. are the False-negative and positive values of the classification, and those are considered to be the wrong classification (Ed-Daoudy & Maalmi, 2021).

**Accuracy**

The accuracy of the classifier is defined by the count of total correct predictions over the total observational data. In this context, the accuracy can be determined by the following formula:

**Precision**

Precision is the measure of the classification to determine how the classifier is precise in identifying the classes (Chaves & Ramírez, 2010). Concerning the confusion matrix, the precision of the classifier can be determined as follows:

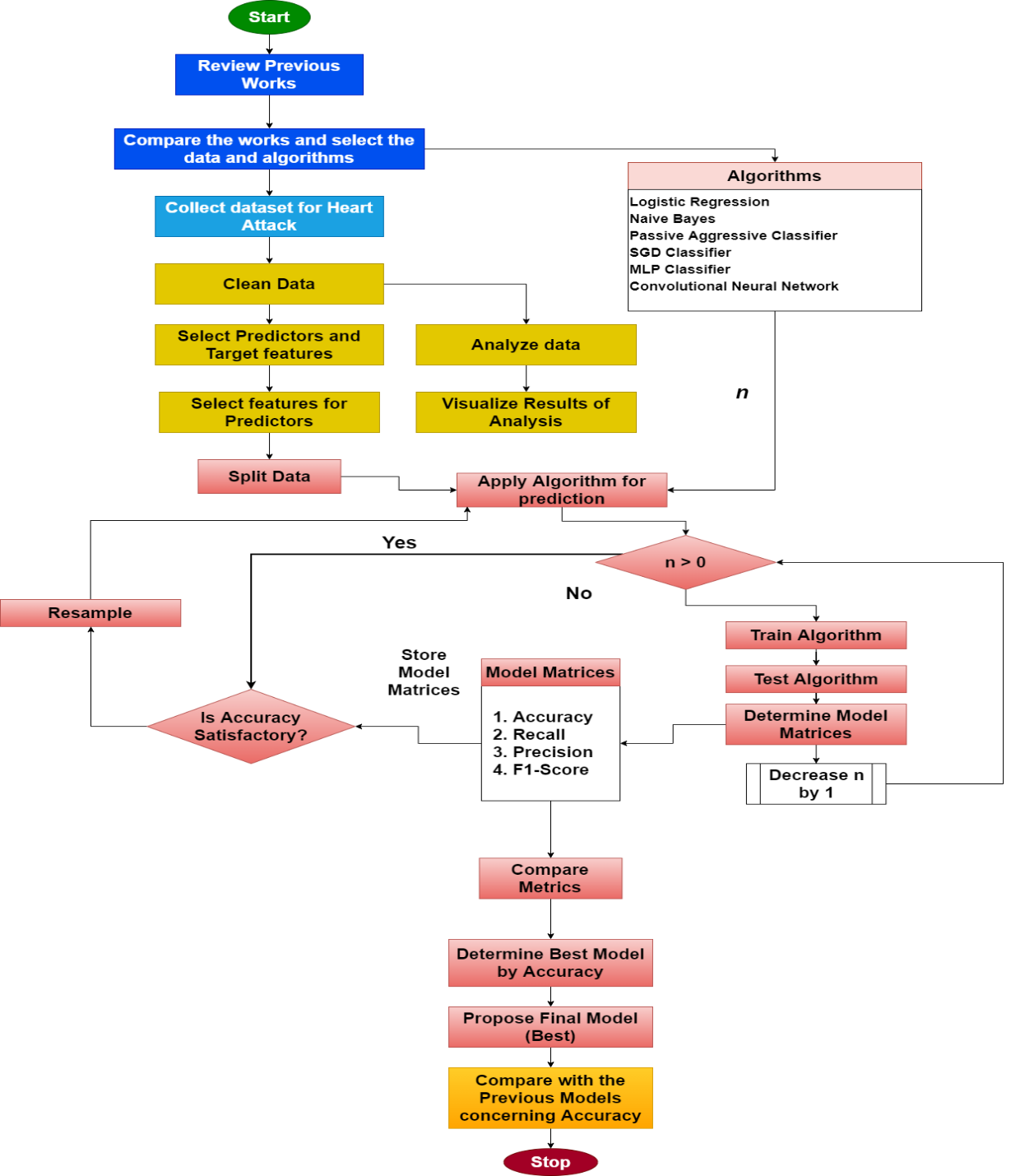
**Recall**

The recall is another metric of classification to judge the effectiveness of the classification by identifying how the classifier is good to detect classes over false values. The recall can be measured as follows:

**F1-Score**

This is one of the most important measures for judging the effectiveness of the classifier for imbalanced classification. This is the combined measure of precision and recall which means, this metric is incorporated by the measures of precision and recall (Gogi & M.N., 2018). The F1-Score of the classifier can be measured as follows:

## Proposed Methodology

The method that will be applied in this project for the detection of heart attack based on symptoms is shown below:

**Figure 2:Methodology for the Detection of Heart Attack**

# Chapter-4 Analysis and Result

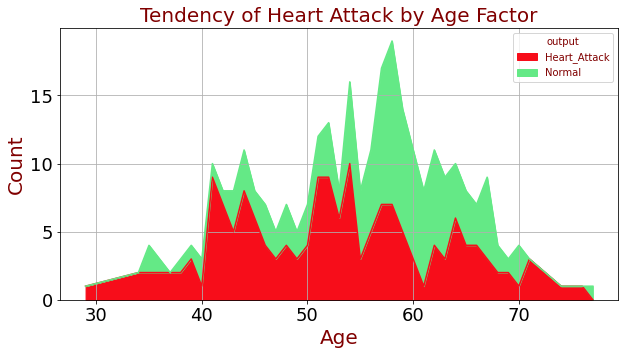
This chapter will be dealing with the discussion of the result o the analysis and the detection of a heart attack. The analysis will be done using python by taking the selected data into concern. The result of the data analysis and the practical output of the application of machine learning and deep learning classifiers will be shown and demonstrated here.

## Analysis of Heart Attack Data

In this section, the discussion and demonstration of the analytical outcome will be done regarding the feature of the data.

**The tendency of Heart Attack by Age Factor**

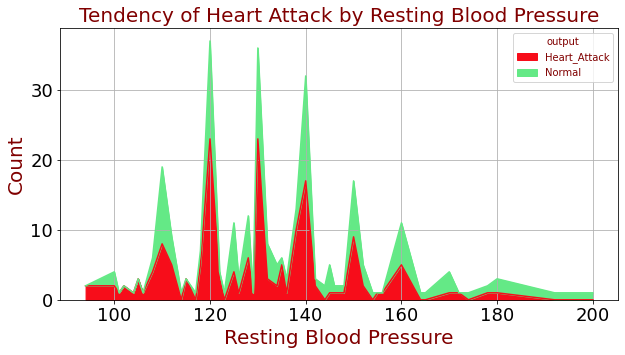
The age factor has been analyzed, and it has been seen that the higher tendency of having a heart attack is from 50-60 years. The outcome is presented as follows:



**Figure 3:The tendency of Heart Attack by Age Factor**

**The tendency of Heart Attack by Resting Blood Pressure**

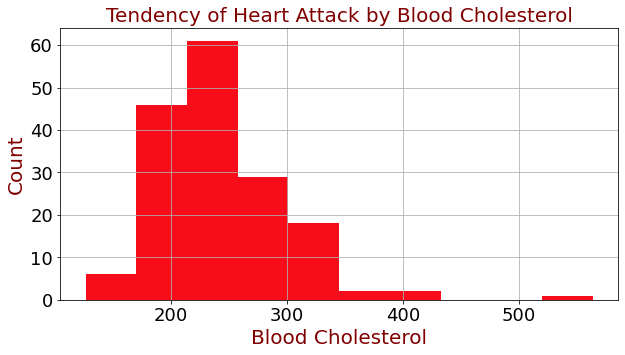
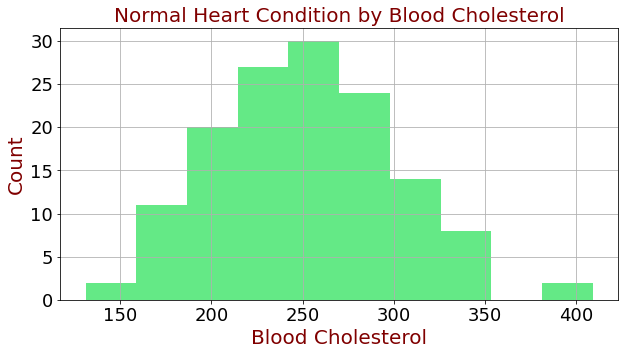
The Resting Blood Pressure has been analyzed, and it has been seen that the higher tendency of having a heart attack are the value of blood pressure such as 120 and 130. The outcome is presented as follows (Chen, et al., 2017):



**Figure 4:The tendency of Heart Attack by Resting Blood Pressure**

**The tendency of Heart Attack by Blood Cholesterol**

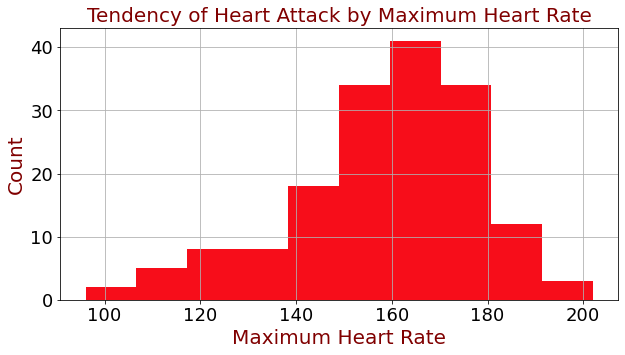
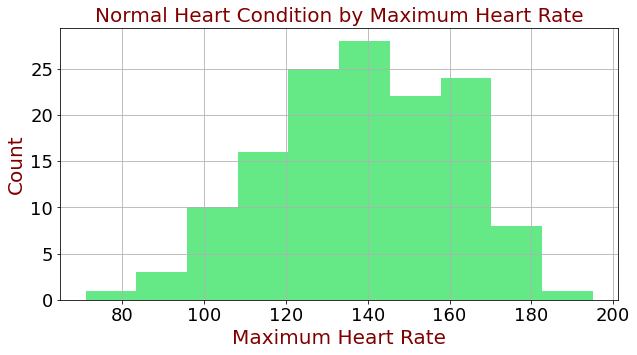
The amount of cholesterol in the blood has been analyzed, and it has been seen that the higher tendency of having a heart attack is from 200-300. The outcome is presented as follows:



**Figure 5:The tendency of Heart Attack by Blood Cholesterol**

**The tendency of Heart Attack by Maximum Heart Rate**

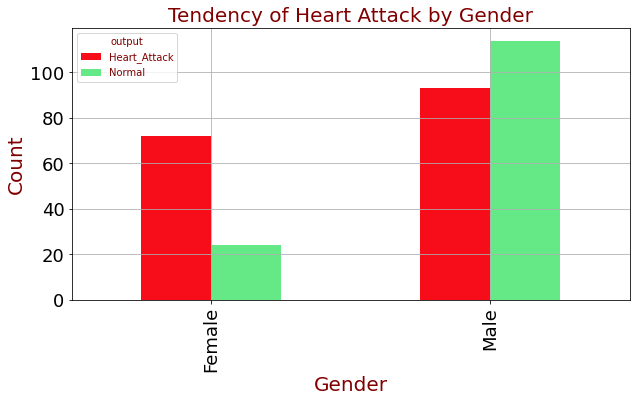
The Maximum Heart Rate of the patients have been analyzed, and it has been seen that the higher tendency of having a heart attack is from 140-180. The outcome is presented as follows:



**Figure 6:The tendency of Heart Attack by Maximum Heart Rate**

**The tendency of Heart Attack by Gender**

Gender is one of the important factors for the identification of the possibility of a heart attack (Haq, et al., 2021). From the analysis shown below, it can be observed that the possibility of occurrence of a heart attack is higher in the case of Male compared to females.



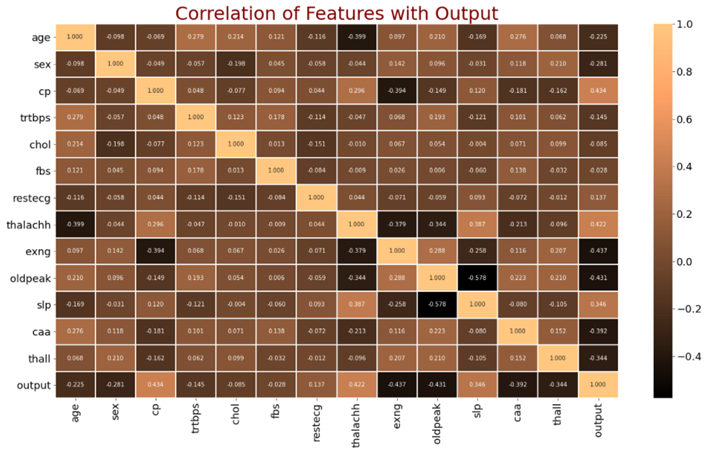
**Figure 7:The tendency of Heart Attack by Gender**

## Heart Attack Prediction

The outcome of the classification will be presented in this section.

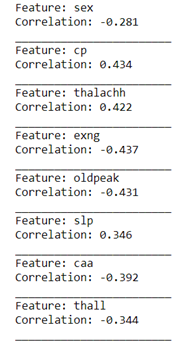
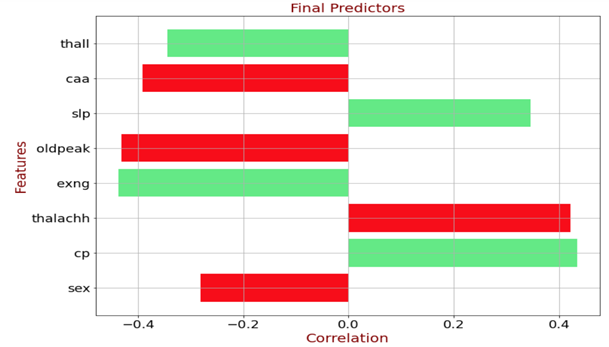
### Feature Selection

The selected data contains 13 symptoms or features, out of which the most important symptoms will be found out by applying correlation (Ismaeel, et al., 2015). The feature of importance will be selected for those symptoms that have a higher correlation with the "output" feature (target). For doing this, the correlation has been done on the features, and the outcome heatmap has been shown below:



**Figure 8:Correlation of Features with Outcome**

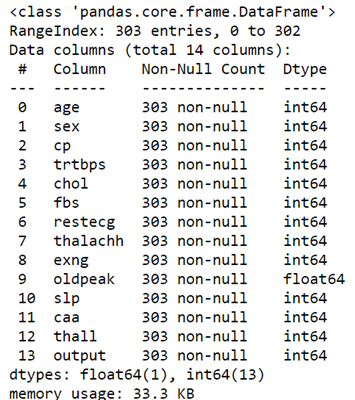
After executing the correlation, the most important symptoms have been found, and those are namely sex, cp, thelachh, exng, oldpeak, spl, caa and thall. The outcome of the selection of the important symptoms are shown below:

**Figure 9:Selection of Features from Correlation**

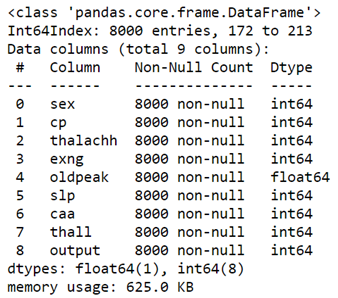
### Data Resampling

The selected data contains 14 features, and each of the features contains 303 instances. This small number of instances will not be suitable for the classifiers of machine learning and deep learning due to the problem called underfitting (Khan & Mondal, 2021). The data information of the selected dataset is shown below:



**Figure 10:Data Information before Resampling**

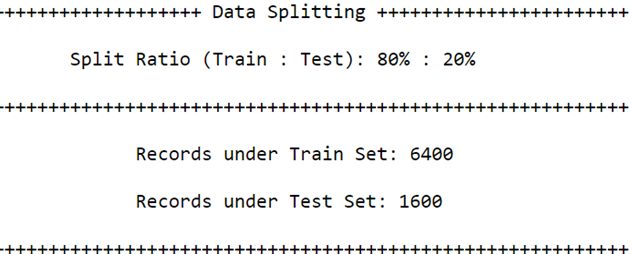
So, the data has been resampled, and 8000 rows have been generated from the existing instances, and the data information for the present resampled data is shown below:



**Figure 11:Data Information after Resampling**

### Creating Data Segments

Now, the data has been segregated to form two separate datasets, out of which the first segregation will be used for training the classifiers, and the other segregation will be sued for testing the model (Kamata, et al., 2015). Here, the segregation has been done with the ratio of 80%-20% and shown below:

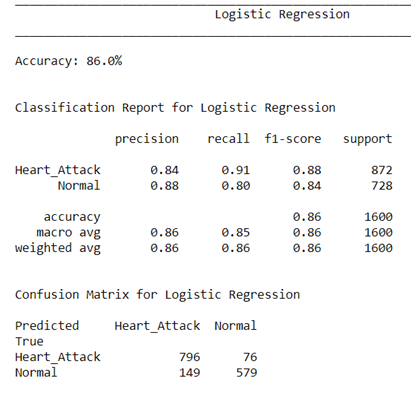
****

**Figure 12:Data Segregation**

### Prediction Using Machine Learning

**Logistic Regression**

This classifier has been applied to the heart attack data for detection and prediction purposes. The classifier has been fitted with the train data segment, and thereby the relevant model has been generated. This model has been predicted by the test data segment to determine the heart attack (Chitnis, et al., 2020). In order to assess the effectiveness of the classifier, the metrics of the classification have been found, and the confusion matrix has been generated. The outcome of the application of this classifier is shown below:

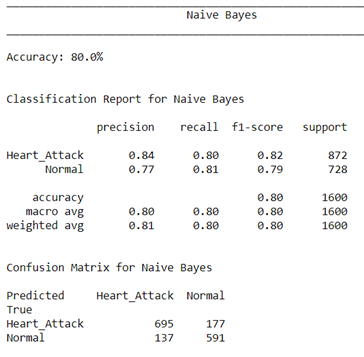


**Figure 13:Outcome of Classification by Logistic Regression**

The outcomes of the classification have been shown above, which provides insight into this classifier. The classification report shows that this classifier has produced 84% precision and 88% f1-score in heart attack and 88% precision and 84% f1-score for the normal patients. On the other hand, the accuracy of this classifier is 86%. From the view of the confusion matrix, it can be seen that 1375 observations have been predicted correctly, and 225 observations have been misclassified out of 1600 instances in the test data segment. So, the classifier can be said to be good/bad for the classification and detection of a heart attack (Keya, et al., 2021).

**Naive Bayes**

The detection of a heart attack has been done by applying this classifier. The classifier has been sequentially trained and tested using the train, and test data segments and the outcome of the classification has been achieved. The outcome of the classifier has been evaluated using metrics like accuracy, f1-score, precision etc., by classes such as heart attack and normal. The classification report and the confusion matrix have been generated to get an accurate view of the outcome (Geetha & Arunachalam, 2021). The result of the classification is shown below:

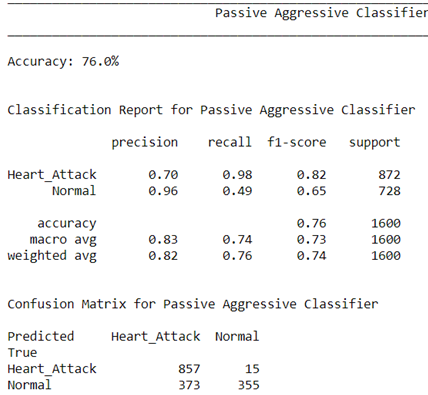


**Figure 14:Outcome of Classification by Naïve Bayes**

The outcomes of the classification have been shown above that provides insight into this classifier (Raja, et al., 2021). The classification report shows that this classifier has produced 84% precision and 82% f1-score in heart attack and 77% precision and 79% f1-score for the normal patients. On the other hand, the accuracy of this classifier is 80%. From the view of the confusion matrix, it can be seen that 1286 observations have been predicted correctly, and 314 observations have been misclassified out of 1600 instances in the test data segment (Kanchan & Kishor, 2016). So, the classifier can be said to be good/bad for the classification and detection of a heart attack.

**Passive Aggressive Classifier**

This classifier has been applied to the heart attack data for detection and prediction purposes. The classifier has been fitted with the train data segment ,and thereby, the relevant model has been generated (Banowati, et al., 2019). This model has been predicted by the test data segment to determine the heart attack. To check the effectiveness of the classifier, the metrics of the classification have been found, and the confusion matrix has been generated. The outcome of the application of this classifier is shown below:

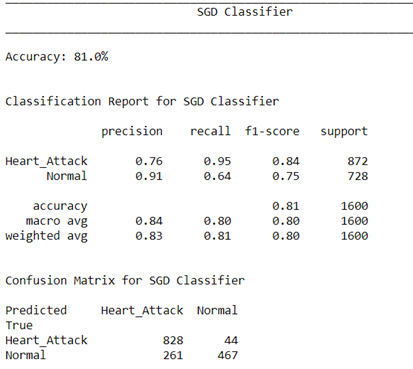


**Figure 15:Outcome of Classification by Passive Aggressive Classifier**

The outcomes of the classification have been shown above , which provides insight into this classifier. The classification report shows that this classifier has produced 70% precision and 82% f1-score in heart attack and 96% precision and 65% f1-score for the normal patients. On the other hand, the accuracy of this classifier is 76%. From the view of the confusion matrix, it can be seen that 1212 observations have been predicted correctly, and 388 observations have been misclassified out of 1600 instances in the test data segment (Akhtar, et al., 2021). So, the classifier can be said to be good/bad for the classification and detection of a heart attack.

**SGD Classifier**

This classifier has been applied to the heart attack data for detection and prediction purposes. The classifier has been fitted with the train data segment , generating the relevant model (Ahmed, et al., 2021). This model has been predicted by the test data segment to determine the heart attack. To check the effectiveness of the classifier, the metrics of the classification have been found, and the confusion matrix has been generated. The outcome of the application of this classifier is shown below:

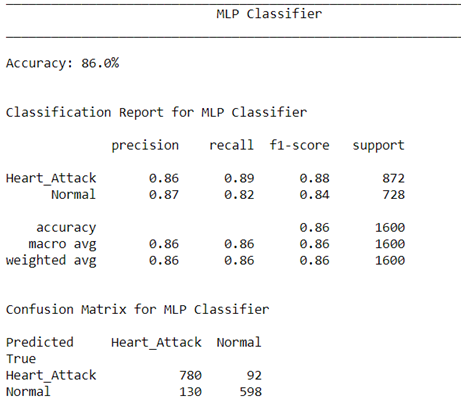


**Figure 16:Outcome of Classification by SGD Classifier**

The outcomes of the classification have been shown above, which provides insight into this classifier. The classification report shows that this classifier has produced 76% precision and 84% f1-score in heart attack and 91% precision and 75% f1-score for the normal patients. On the other hand, the accuracy of this classifier is 81%. From the view of the confusion matrix, it can be seen that 1296 observations have been predicted correctly, and 304 observations have been misclassified out of 1600 instances in the test data segment. So, the classifier can be said to be good/bad for the classification and detection of a heart attack (Elstad & Walløe, 2015).

**MLP Classifier**

The detection of a heart attack has been done by applying this classifier. The classifier has been sequentially trained and tested using the train and test data segments, and the outcome of the classification has been achieved (Faiz, et al., 2019). The outcome of the classifier has been evaluated using metrics like accuracy, f1-score, precision etc., by classes such as heart attack and normal. The classification report and the confusion matrix have been generated to get an accurate view of the outcome. The result of the classification is shown below:



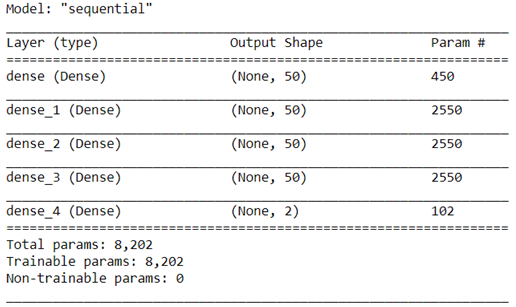
**Figure 17:Outcome of Classification by MLP Classifier**

The outcomes of the classification have been shown above that provides insight into this classifier (Banowati, et al., 2019). The classification report shows that this classifier has produced 86% precision and 88% f1-score in heart attack and 87% precision and 84% f1-score for the normal patients. On the other hand, the accuracy of this classifier is 86%. From the view of the confusion matrix, it can be seen that 1378 observations have been predicted correctly, and 222 observations have been misclassified out of 1600 instances in the test data segment. So, the classifier can be said to be good/bad for the classification and detection of a heart attack (Schulte, et al., 2018).

### Prediction using Deep Learning

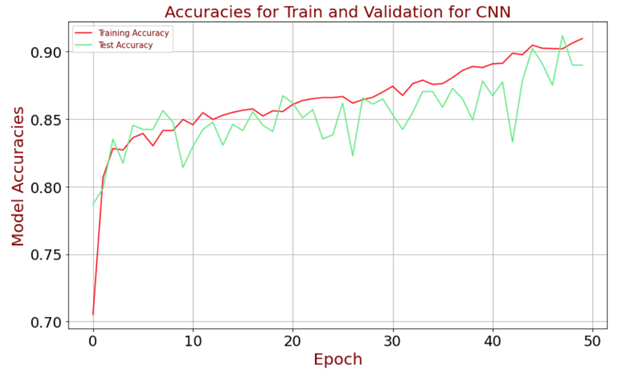
**Convolutional Neural Network**

The detection of a heart attack has been done by applying a Convolutional Neural Network. The classifier has been sequentially trained and tested using the train, and test data segments and the outcome of the classification has been achieved (Kumari, et al., 2019). The outcome of the classifier has been evaluated using metrics like accuracy, f1-score, precision etc., by classes such as heart attack and normal. The model that has been created for the Convolutional Neural Network is shown below:



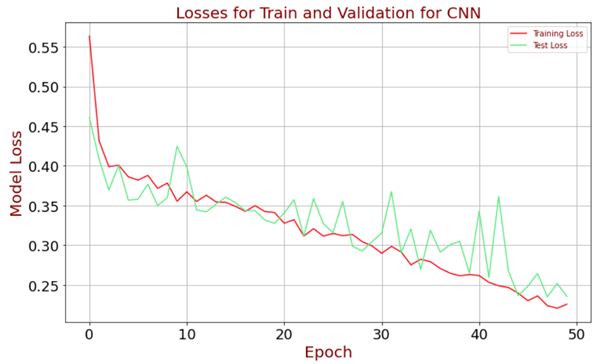
**Figure 18:Model for Convolutional Neural Network**

After the application of the Convolutional Neural Network, the accuracy has been compared for the training phase and testing phase and is shown below:



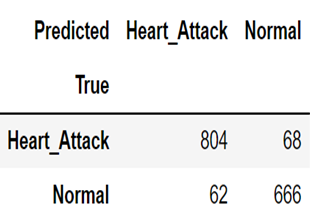
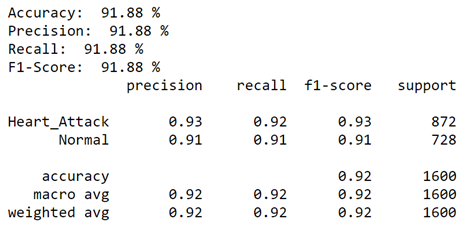
**Figure 19:Accuracy Comparison for CNN**

After the application of the Convolutional Neural Network, the loss in the model has been compared for the training phase and testing phase and is shown below:



**Figure 20:Model Loss Comparison for CNN**

The classification report and the confusion matrix have been generated to get an accurate view of the outcome. The result of the classification is shown below:

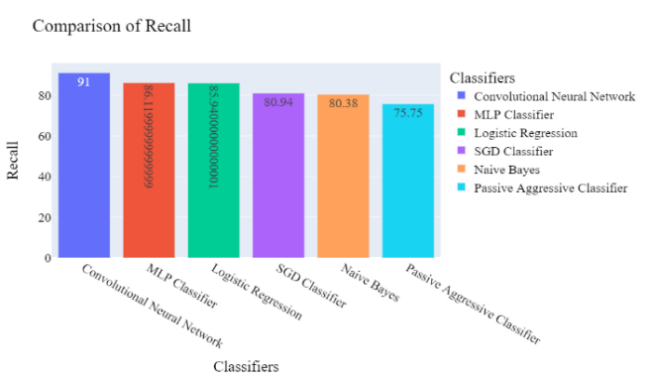
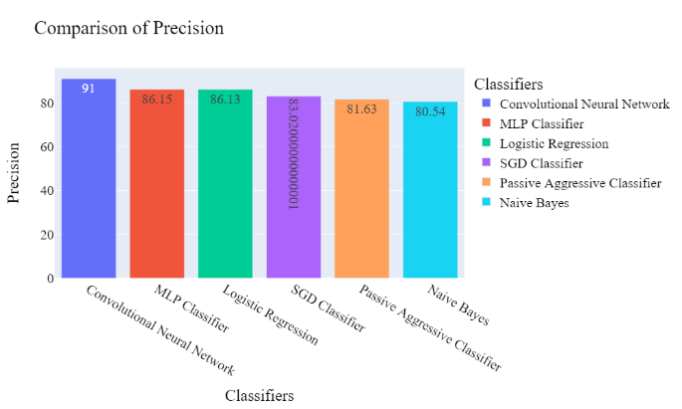
 

**Figure 21:Model Loss Comparison for CNN**

The outcomes of the classification have been shown above that provides insight into this classifier (Kodama, et al., 2019). The classification report shows that this classifier has produced 93% precision and 93% f1-score in heart attack and 91% precision and 91% f1-score for the normal patients. On the other hand, the accuracy of this classifier is 91.88%. From the view of the confusion matrix, it can be seen that 1470 observations have been predicted correctly, and 130 observations have been misclassified out of 1600 instances in the test data segment. So, the classifier can be said to be good/bad for the classification and detection of a heart attack (Kamata, et al., 2015).

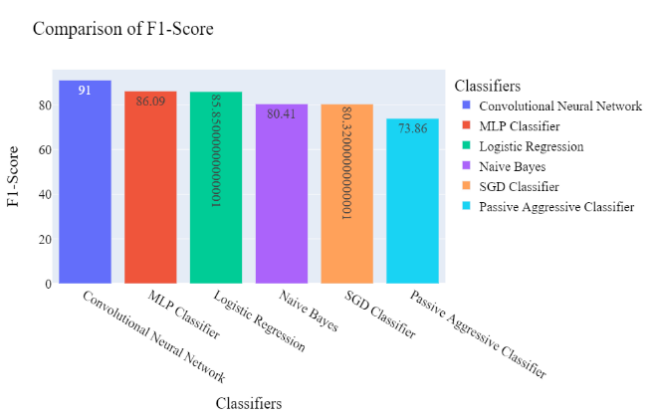
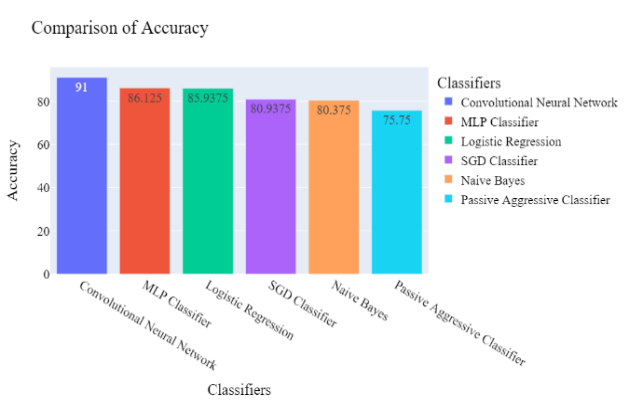
## Comparison of Performances

Now, the performance of the classifiers will be compared to find the best one as defined in the proposed methodology section. In this context, all four classification metrics will be compared, and the comparison of precision and recall are shown below:



**Figure 22:Comparison of Precisions and Recalls**

In the same way, the comparison has been done for accuracy and f1-score, and these are shown below:

**Figure 23:Comparison of Accuracies and F1-Scores**

So, the comparison shows that the Convolutional Neural Network has produced the highest accuracy along with the f1-score, recall and precision. So, Convolutional Neural Network will be taken as the best classifier for the detection of a heart attack.

# Chapter-5 Discussion and Conclusion

The project has been done for the detection of heart disease by emphasizing the prediction of a heart attack. In the present world, a heart attack is one of the most seen diseases for which a good number of people dies. So, it is necessary to identify the symptoms and predict the possibility of a heart attack. Previous literature works have been reviewed in order to design the aim, research question and project objectives. Atallah & Al-Mousa (2021) have used the classifiers from the domain of machine learning for the detection of heart disease and got 89% accuracy. Ed-Daoudy & Maalmi (2021) have used the classifiers from the domain of machine learning for the detection of heart disease and got 87% accuracy in detection. Keya et al. (2021) have used the classifiers from the domain of machine learning for the detection of heart disease and got 87% accuracy in the detection of heart disease. Raja et al. (2021) have used the classifiers from the domain of machine learning for the detection of heart disease and got 85% accuracy in their experiment. Haq et al. (2021) have used the classifiers from the domain of machine learning for the detection of heart disease and got 87.45% accuracy in their experiment. Khan & Mondal (2021) have used the classifiers from the domain of machine learning for the detection of heart disease and got 73.87% accuracy. Ahmed et al. (2021) have used the classifiers from the domain of machine learning for the detection of heart disease and got 89.56% accuracy. So, from the literature review, it has been identified that those researchers have used machine learning and deep learning classifiers to detect a heart attack and got the average accuracy. Most of them have used the heart attack dataset from UCI. By following the approaches and increasing the accuracy of detection, the data has been collected to detect heart disease. The data mainly contains the symptoms for which the heart disease has been observed. However, the data shows both the people with and without heart disease corresponding to the symptoms. The data has been collected from the Kaggle data repository. The original dataset contains 76 attributes that are the symptoms of a heart attack. Later for the feasibility of research, only 13 important attributes, along with the attribute with the name "output", have been taken into consideration. So, the stored 13 attributes of features of the data are the selected symptoms of the heart attack. For heart attack detection, five classifiers have been chosen, namely Logistic Regression, Naive Bayes, Passive Aggressive Classifier, SGD Classifier and MLP Classifier, and one classifier has been chosen from deep learning, namely Convolutional Neural Network. By applying all six classifiers, it has been observed that Convolutional Neural Network has produced 91.8% accuracy for predicting a heart attack, which is higher than the previous models. So, Convolutional Neural Network will be taken as the best classifier regarding the higher accuracy, f1-score, precision and recall.

## Addressing the Research Questions

In this project, three questions have been formulated based on the aim of the research, and those will be addressed below:

While addressing question-1, it can be said that the heart attack can be predicted precisely using a deep learning classifier with a 91% average precision over two classes. While addressing question-2, it can be said that the symptoms and features namely sex, cp, thelachh, exng, oldpeak, spl, caa and thall, are high;y useful for the prediction of a heart attack. While addressing question-3, it can be said that Convolutional Neural Network has produced 91.8% accuracy for predicting a heart attack, which is higher than the previous models.

## Limitation of the Project

The present limitation of the project are as follows:

1. The project has been done by collecting data that contains only the observations of 303 patients.
2. Fewer features or symptoms have been used here as the data implied so.

## Future Scope

The future scope of the project are as follows:

1. The final model that has been selected is by Convolutional Neural Network. This model can be used for the detection of other types of diseases.

# References

A. S. Abdullah, R. R. R., 2012. A data mining model for predicting coronary heart disease using random forest classifier. *Proc. Int. Conf. Recent Trends Comput. Methods Commun. Controls,* pp. 22-25.

Abdulhadi, N. & Al-Mousa, A., 2021. Diabetes Detection Using Machine Learning Classification Methods. *IEEE Access,* pp. 715-719.

Ahmed, M. R. et al., 2021. *A Cloud Based Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms.* [Online]

Available at: https://ieeexplore.ieee.org/document/8781022

Akhtar, F. et al., 2021. Plant Disease Detection based on Deep Learning Approach. *International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE),* pp. 1-5.

Akshai, K. & Anitha, J., 2021. Plant disease classification using deep learning. *3rd International Conference on Signal Processing and Communication (ICPSC),* pp. 1-6.

Atallah, R. & Al-Mousa, A., 2021. Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method. *IEEE Access,* 9 Nov.pp. 660-669.

Banowati, C., Novianty, A. & Setianingsih, C., 2019. Cholesterol Level Detection Based on Iris Recognition Using Convolutional Neural Network Method. *IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET),* pp. 1-7.

BHF, 2021. *Facts and figures.* [Online]

Available at: https://www.bhf.org.uk/what-we-do/news-from-the-bhf/contact-the-press-office/facts-and-figures

Biswas, R. et al., 2020. Identification of Pathological Disease in Plants using Deep Neural Networks - Powered by Intel® Distribution of OpenVINO. *Computing Analytics and Networks (Indo-Taiwan ICAN) 2020 Indo – Taiwan 2nd International Conference,* pp. 45-48.

Bose, B., Priya, J. & Welekar, S., 2020. Hemp Disease Detection and Classification Using Machine Learning and Deep Learning. *IEEE Access,* pp. 915-919.

Chaves, R. & Ramírez, J., 2010. Effective Diagnosis of Alzheimer's Disease by Means of Association Rules. *Hybrid Artificial Intelligence Systems,* Volume 1, pp. 452-459.

Chen, M. et al., 2017. Disease Prediction by Machine Learning Over Big Data From Healthcare Communities. *IEEE Access,* pp. 8869 - 8879.

Chitnis, G., Bhanushali, V. & Ranade, A., 2020. A Review of Machine Learning Methodologies for Dental Disease Detection. *IEEE Access,* pp. 145-151.

Ed-Daoudy, A. & Maalmi, K., 2021. *Real-time machine learning for early detection of heart disease using big data approach.* [Online]

Available at: https://.ieee.org/document/8723839

Elstad, M. & Walløe, L., 2015. Heart rate variability and stroke volume variability to detect central hypovolemia in spontaneously breathing, young, healthy volunteers. *8th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO),* pp. 1-4.

Faiz, M., Naseem, S. A., Murad, M. & khalid, R., 2019. Cost effective monitoring system for the early detection of ischemic stroke in smoker. *International Conference on Information Science and Communication Technology (ICISCT),* pp. 1-4.

Geetha, C. & Arunachalam, A., 2021. Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms. *International Conference on Computer Communication and Informatics (ICCCI),* pp. 1-6.

Gogi, V. J. & M.N., V., 2018. Prognosis of Liver Disease: Using Machine Learning Algorithms. *International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE),* pp. 198-203.

Haq, A. U. et al., 2021. Identifying the Predictive Capability of Machine Learning Classifiers for Designing Heart Disease Detection System. *IEEE Access,* 9 Nov.pp. 1-5.

Islam, M. A. et al., 2020. Risk Factor Prediction of Chronic Kidney Disease based on Machine Learning Algorithms. *3rd International Conference on Intelligent Sustainable Systems (ICISS),* pp. 103-106.

Ismaeel, S., Miri, A. & Chourishi, D., 2015. Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis. *IEEE Canada International Humanitarian Technology Conference (IHTC2015),* pp. 1-3.

Kamata, K. et al., 2015. Development of stroke detection method by heart rate variability analysis and support vector machine. *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA),* pp. 876-879.

Kanchan, B. D. & Kishor, M. M., 2016. Study of machine learning algorithms for special disease prediction using principal of component analysis. *International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC),* pp. 134-138.

Katarya, R. & Jain, S., 2020. Comparison of Different Machine Learning Models for diabetes detection. *IEEE Access,* pp. 682-694.

Katarya, R. & Srinivas, P., 2020. Predicting heart disease at early stages using machine learning: A survey. *In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE.,* pp. pp. 302-305.

Keya, M. S. et al., 2021. Measuring the Heart Attack Possibility using Different Types of Machine Learning Algorithms. *IEEE Access,* 9 Nov.pp. 852-863.

Khan, M. I. H. & Mondal, M. R. H., 2021. Effectiveness of Data Driven Diagnosis of Heart Disease. *IEEE Access,* 9 Nov.pp. 3998-4011.

Khan, P., Kader, M. F. & Islam, S. M. R., 2021. Machine Learning and Deep Learning Approaches for Brain Disease Diagnosis: Principles and Recent Advances. *IEEE Access,* pp. 214-223.

Kodama, T. et al., 2019. Ischemic Stroke Detection by Analyzing Heart Rate Variability in Rat Middle Cerebral Artery Occlusion Model. *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* pp. 1-7.

Kohli, P. S. & Arora, S., 2018. *Application of Machine Learning in Disease Prediction.* s.l.:s.n.

Korkut, U. B., Göktürk, Ö. B. & Yildiz, O., 2018. Detection of plant diseases by machine learning. *IEEE Access,* pp. 1-5.

Kumari, D., Saini, I. & Sood, N., 2019. Detection of Heart rate through Speech Using Mel Frequency cepstrum coefficients. *3rd International Conference on Trends in Electronics and Informatics (ICOEI),* pp. 1-5.

Raja, M. S., Anurag, M., Reddy, C. P. & Sirisala, N. R., 2021. Machine Learning Based Heart Disease Prediction System. *IEEE Access,* 9 Nov.pp. 492-498.

Ronit, 2020. *Heart Disease UCI.* [Online]

Available at: https://www.kaggle.com/ronitf/heart-disease-uci

[Accessed 2021].

Schulte, R., Krug, J. & Rose, G., 2018. Identification of a signal for an optimal heart beat detection in multimodal physiological datasets. *Computing in Cardiology,* pp. 1-6.

Tao, R. et al., 2018. Magnetocardiography-based ischemic heart disease detection and localization using machine learning methods. *IEEE Transactions on Biomedical Engineering, 66(6).,* pp. 1658-1667.

Tarik, M. I., Akter, S. & Mamun, A. A., 2021. Potato Disease Detection Using Machine Learning. *IEEE Access,* pp. 15567-15581.

# Appendix

**Code**

**Import Libraries**

import warnings

warnings.filterwarnings("ignore")

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import matplotlib

matplotlib.rcParams['axes.labelsize'] = 19

matplotlib.rcParams['xtick.labelsize'] = 18

matplotlib.rcParams['ytick.labelsize'] = 18

matplotlib.rcParams['text.color'] = '#800000'

lbl="#800000"

bicol=["#F70D1A","#64E986"]

import seaborn as sns

import plotly.express as px

import plotly.figure\_factory as ff

from sklearn.linear\_model import LogisticRegression

from sklearn.linear\_model import PassiveAggressiveClassifier

from sklearn.naive\_bayes import BernoulliNB

from sklearn.linear\_model import SGDClassifier

from sklearn.neural\_network import MLPClassifier

from sklearn.utils import resample

from sklearn.utils import shuffle

import tensorflow as tf

from tensorflow.keras.applications import ResNet50

from tensorflow.keras.layers import Dense, Flatten, Dropout,BatchNormalization ,Activation

from tensorflow.keras.models import Model, Sequential

from tensorflow.keras.optimizers import Adam

from tensorflow.keras.callbacks import ReduceLROnPlateau, ModelCheckpoint, EarlyStopping

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import precision\_score, recall\_score, f1\_score, classification\_report,accuracy\_score

**Reading Heart Attack Data**

heart1=pd.read\_csv("heart.csv")

heart1.head()

heart1.shape

heart1['output'].unique()

heart1.info()

**Formating Features**

heart=heart1.copy()

heart['output']=heart['output'].replace(heart['output'].unique(),["Heart\_Attack","Normal"])

heart['exng']=heart['exng'].replace(heart['exng'].unique(),["No","Yes"])

heart['cp']=heart['cp'].replace(heart['cp'].unique(),["Asymptomatic","Typical Angina","Non-anginal Pain","Atypical Angina"])

heart['fbs']=heart['fbs'].replace(heart['fbs'].unique(),["> 120 mg/dl","< 120 mg/dl"])

heart['sex']=heart['sex'].replace(heart['sex'].unique(),["Male","Female"])

heart=heart.sample(frac=1)

heart=heart.reset\_index(drop=True)

heart.head()

**Data Attributes**

print("+++++++++++++++++++++++++++++++++++++++++++")

print("\t Rows in Data: {}\n\t Columns in Data: {}".format(heart.shape[0],heart.shape[1]))

print("+++++++++++++++++++++++++++++++++++++++++++")

print("++++++++++++++++++++++++++++++++++")

print(" Observations By Gneder")

print("++++++++++++++++++++++++++++++++++")

print(" Male: ",heart['sex'].value\_counts().tolist()[0])

print(" Female: ",heart['sex'].value\_counts().tolist()[1])

print("++++++++++++++++++++++++++++++++++")

**Data Analysis**

pd.crosstab(heart.age,heart['output']).plot(kind="area",figsize=(9,6),color=bicol)

plt.title('Tendency of Heart Attack by Age Factor', fontsize=19,color=lbl)

plt.xlabel('Age', fontsize=19,color=lbl)

plt.ylabel('Count', fontsize=19,color=lbl)

plt.grid()

plt.savefig("age.jpg")

plt.show()

pd.crosstab(heart.trtbps,heart['output']).plot(kind="area",figsize=(9,6),color=bicol)

plt.title('Tendency of Heart Attack by Resting Blood Pressure', fontsize=19,color=lbl)

plt.xlabel('Resting Blood Pressure', fontsize=19,color=lbl)

plt.ylabel('Count', fontsize=19,color=lbl)

plt.grid()

plt.savefig("trtbps.jpg")

plt.show()

yes=heart[heart['output']=="Heart\_Attack"]

no=heart[heart['output']=="Normal"]

plt.figure(figsize=(10,6))

plt.hist(yes.chol,color=bicol[0])

plt.title('Tendency of Heart Attack by Blood Cholesterol', fontsize=19,color=lbl)

plt.xlabel('Blood Cholesterol', fontsize=19,color=lbl)

plt.ylabel('Count', fontsize=19,color=lbl)

plt.grid()

plt.savefig("cholyes.jpg")

plt.show()

plt.figure(figsize=(10,6))

plt.hist(no.chol,color=bicol[1])

plt.title('Normal Heart Condition by Blood Cholesterol', fontsize=19,color=lbl)

plt.xlabel('Blood Cholesterol', fontsize=19,color=lbl)

plt.ylabel('Count', fontsize=19,color=lbl)

plt.grid()

plt.savefig("cholno.jpg")

plt.show()

plt.figure(figsize=(10,6))

plt.hist(yes.thalachh,color=bicol[0])

plt.title('Tendency of Heart Attack by Maximum Heart Rate', fontsize=19,color=lbl)

plt.xlabel('Maximum Heart Rate', fontsize=21,color=lbl)

plt.ylabel('Count', fontsize=19,color=lbl)

plt.grid()

plt.savefig("thalachhyes.jpg")

plt.show()

plt.figure(figsize=(10,5))

plt.hist(no.thalachh,color=bicol[1])

plt.title('Normal Heart Condition by Maximum Heart Rate', fontsize=19,color=lbl)

plt.xlabel('Maximum Heart Rate', fontsize=20,color=lbl)

plt.ylabel('Count', fontsize=20,color=lbl)

plt.grid()

plt.savefig("thalachhno.jpg")

plt.show()

pd.crosstab(heart.sex,heart['output']).plot(kind="bar",figsize=(10,5),color=bicol)

plt.title('Tendency of Heart Attack by Gender', fontsize=20,color=lbl)

plt.xlabel('Gender', fontsize=20,color=lbl)

plt.ylabel('Count', fontsize=20,color=lbl)

plt.grid()

plt.savefig("gender.jpg")

plt.show()

pd.crosstab(heart.cp,heart['output']).plot(kind="bar",figsize=(10,5),color=bicol)

plt.title('Tendency of Heart Attack by Chest Pain', fontsize=19,color=lbl)

plt.xlabel('Chest Pain', fontsize=20,color=lbl)

plt.ylabel('Count', fontsize=19,color=lbl)

plt.grid()

plt.savefig("cp.jpg")

plt.show()

pd.crosstab(heart.fbs,heart['output']).plot(kind="bar",figsize=(10,5),color=bicol)

plt.title('Tendency of Heart Attack by Fasting Sugar Level', fontsize=19,color=lbl)

plt.xlabel('Fasting Sugar Level', fontsize=19,color=lbl)

plt.ylabel('Count', fontsize=20,color=lbl)

plt.grid()

plt.savefig("fbs.jpg")

plt.show()

pd.crosstab(heart.exng,heart['output']).plot(kind="bar",figsize=(10,5),color=bicol)

plt.title('Tendency of Heart Attack by Exercise Induced Angina', fontsize=19,color=lbl)

plt.xlabel('Exercise Induced Angina', fontsize=19,color=lbl)

plt.ylabel('Count', fontsize=19,color=lbl)

plt.grid()

plt.savefig("exng.jpg")

plt.show()

heart.head()

**Detection of Heart Attack by Machine Learning**

plt.figure(figsize=(21,13))

plt.title("Correlation of Features with Output", fontsize=31,color=lbl)

sns.heatmap(heart1.corr(),annot=True,cmap='copper',fmt='.3f',linewidths=1)

plt.savefig("corr.jpg")

plt.show()

corref=heart1.corr()['output']

init\_preds=corref.index.tolist()[:-1]

init\_corrs=corref.tolist()[:-1]

posall=[]

negall=[]

fin\_preds=[]

fin\_corrs=[]

for i in range(len(init\_corrs)):

if init\_corrs[i]<0:

negall.append(init\_corrs[i])

else:

posall.append(init\_corrs[i])

negall=np.array(negall)

posall=np.array(posall)

for i in range(len(init\_corrs)):

if init\_corrs[i]>posall.mean() or init\_corrs[i]<negall.mean():

fin\_preds.append(init\_preds[i])

fin\_corrs.append(init\_corrs[i])

plt.figure(figsize=(13,11))

plt.title("Final Predictors", fontsize=20,color=lbl)

plt.barh(fin\_preds,fin\_corrs,color=bicol)

plt.xlabel("Correlation", fontsize=19,color=lbl)

plt.ylabel("Features", fontsize=19,color=lbl)

plt.grid()

for i in range(len(fin\_preds)):

print("Feature: {}\nCorrelation: {}\n\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_".format(fin\_preds[i],fin\_corrs[i],3)))

**Data Resampling**

heart\_rs = resample(heart1, replace = True, n\_samples = 8000, random\_state = 0)

heart\_rs=pd.concat([heart\_rs[fin\_preds],heart\_rs['output']],axis=1)

heart\_rs.shape

heart\_rs['output'].unique()

heart\_rs.info()

X=heart\_rs[fin\_preds]

y=heart\_rs['output']

**Split Data**

x\_train,x\_test,y\_train,y\_test=train\_test\_split(X,y, train\_size=0.8, random\_state=10)

print("Data Splitting ")

print("\n\tSplit Ratio (Train : Test): {}% : {}%\n".format((len(x\_train)/len(X))\*100),(len(x\_test)/len(X))\*100)))

print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")

print("\n\t Records under Train Set: {}".format(len(X)\*(0.8))))

print("\n\t Records under Test Set: {}\n".format(len(X)\*(1-0.8))))

print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")

**Selection of Machine Learning Classifier**

mlalgo=[

LogisticRegression(),

PassiveAggressiveClassifier(),

BernoulliNB(),

SGDClassifier(),

MLPClassifier()

]

mlnms=[

"Logistic Regression",

"Passive Aggressive Classifier",

"Naive Bayes",

"SGD Classifier",

"MLP Classifier"

]

prec=[]

rcl=[]

f1=[]

acc=[]

print("\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_")

for i in range(len(mlalgo)):

print(" {} ".format(mlnms[i]))

print("\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_")

mlalgo[i].fit(x\_train,y\_train)

y\_pred=mlalgo[i].predict(x\_test)

score=accuracy\_score(y\_test,y\_pred)

prec.append(precision\_score(y\_test, y\_pred, average='weighted'),4)\*100)

rcl.append(recall\_score(y\_test, y\_pred, average='weighted'),4)\*100)

f1.append(f1\_score(y\_test, y\_pred, average='weighted'),4)\*100)

acc.append(score\*100,4))

cm=pd.crosstab(y\_test, y\_pred, rownames=['True'], colnames=['Predicted'], margins=True)

print("\nAccuracy: {}%\n".format(score,2)\*100))

print("\nClassification Report for {} \n\n{}".format(mlnms[i],classification\_report(y\_test, y\_pred)))

print("\nConfusion Matrix for {} \n\n{}\n".format(mlnms[i],cm.iloc[:2,:2]))

print("\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_")

**Detection of heart Attack by Deep Learning**

y1=y.replace(y.unique(),[0,1])

x\_train\_dp,x\_test\_dp,y\_train\_dp,y\_test\_dp=train\_test\_split(X,y1, train\_size=0.8, random\_state=10)

print("Data Splitting ")

print("\n\tSplit Ratio (Train : Test): {}% : {}%\n".format((len(x\_train)/len(X))\*100),(len(x\_test)/len(X))\*100)))

print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")

print("\n\t Records under Train Set: {}".format(len(X)\*(0.8))))

print("\n\t Records under Test Set: {}\n".format(len(X)\*(1-0.8))))

print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")

input\_size = heart\_rs.shape[1]-1

output\_size = 2

hidden\_layer\_size = 50

heart\_deep\_cnn = tf.keras.Sequential([

tf.keras.layers.Dense(hidden\_layer\_size, input\_shape=(input\_size,), activation='relu'), # 1st hidden layer

tf.keras.layers.Dense(hidden\_layer\_size, activation='relu'),

tf.keras.layers.Dense(hidden\_layer\_size, activation='relu'),

tf.keras.layers.Dense(hidden\_layer\_size, activation='relu'),

tf.keras.layers.Dense(output\_size, activation='softmax') # output layer

])

heart\_deep\_cnn.summary()

heart\_deep\_cnn.compile(optimizer='adam', loss='sparse\_categorical\_crossentropy', metrics=['accuracy'])

batch\_size = 16

epochs = 50

heart\_deep\_cnn\_history = heart\_deep\_cnn.fit(x\_train\_dp,y\_train\_dp,batch\_size=batch\_size,epochs=epochs,verbose=1,validation\_split=0.2)

heart\_deep\_cnn\_df=pd.DataFrame({

"Epoch":[i+1 for i in range(epochs)],

"Training Accuracy":heart\_deep\_cnn\_history.history['accuracy'],

"Test Accuracy":heart\_deep\_cnn\_history.history['val\_accuracy'],

"Training Loss":heart\_deep\_cnn\_history.history['loss'],

"Test Loss":heart\_deep\_cnn\_history.history['val\_loss']

})

heart\_deep\_cnn\_df.to\_csv("heart\_deep\_cnn\_model.csv")

heart\_deep\_cnn\_df.head()

heart\_deep\_cnn\_df.iloc[:,1:3].plot(kind="line",figsize=(12,7),color=bicol)

plt.title("Accuracies for Train and Validation for CNN",fontsize=20,color=lbl)

plt.xlabel("Epoch",fontsize=20,color=lbl)

plt.ylabel("Model Accuracies",fontsize=20,color=lbl)

plt.grid()

plt.show()

heart\_deep\_cnn\_df.iloc[:,3:].plot(kind="line",figsize=(12,7),color=bicol)

plt.title("Losses for Train and Validation for CNN",fontsize=20,color=lbl)

plt.xlabel("Epoch",fontsize=20,color=lbl)

plt.ylabel("Model Loss",fontsize=20,color=lbl)

plt.grid()

plt.show()

heart\_deep\_pred=heart\_deep\_cnn.predict(x\_test\_dp)

tf\_prediction1 = tf.nn.softmax(heart\_deep\_pred, axis=1)

heart\_deep\_pred\_label = tf.argmax(tf\_prediction1, axis=1)

heart\_deep\_pred\_label = heart\_deep\_pred\_label.numpy()

y\_test\_dp=y\_test\_dp.replace(y\_test\_dp.unique(),["Normal","Heart\_Attack"])

print(y\_test\_dp.unique())

heart\_deep\_pred\_label=pd.Series(heart\_deep\_pred\_label)

heart\_deep\_pred\_label=heart\_deep\_pred\_label.replace(heart\_deep\_pred\_label.unique(),['Normal','Heart\_Attack'])

y\_test\_dp

cnn\_prd\_outcome=pd.DataFrame({"Actual":y\_test\_dp.tolist(),"Predicted":heart\_deep\_pred\_label.tolist()})

cnnct=pd.crosstab(cnn\_prd\_outcome['Actual'],cnn\_prd\_outcome['Predicted'],rownames=['True'], colnames=['Predicted'], margins=True)

cnnct.iloc[:2,:2]

print("Accuracy: ",accuracy\_score(cnn\_prd\_outcome["Actual"],cnn\_prd\_outcome["Predicted"]),4)\*100,"%")

print("Precision: ",precision\_score(cnn\_prd\_outcome["Actual"],cnn\_prd\_outcome["Predicted"], average='micro'),4)\*100,"%")

print("Recall: ",recall\_score(cnn\_prd\_outcome["Actual"],cnn\_prd\_outcome["Predicted"], average='micro'),4)\*100,"%")

print("F1-Score: ",f1\_score(cnn\_prd\_outcome["Actual"],cnn\_prd\_outcome["Predicted"], average='micro'),4)\*100,"%")

print(classification\_report(cnn\_prd\_outcome["Actual"],cnn\_prd\_outcome["Predicted"]))

**Performance Comparison of Classifiers**

prec.append(precision\_score(cnn\_prd\_outcome["Actual"],cnn\_prd\_outcome["Predicted"], average='micro'),4)\*100)

rcl.append(recall\_score(cnn\_prd\_outcome["Actual"],cnn\_prd\_outcome["Predicted"], average='micro'),4)\*100)

f1.append(f1\_score(cnn\_prd\_outcome["Actual"],cnn\_prd\_outcome["Predicted"], average='micro'),4)\*100)

acc.append(accuracy\_score(cnn\_prd\_outcome["Actual"],cnn\_prd\_outcome["Predicted"]),4)\*100)

mlnms.append("Convolutional Neural Network")

heart\_clf\_out=pd.DataFrame({

"Classifiers":mlnms,

"Accuracy":acc,

"Precision":prec,

"Recall":rcl,

"F1-Score":f1,

})

for i in heart\_clf\_out.columns.tolist()[1:]:

heart\_clf\_out=heart\_clf\_out.sort\_values(by=i,ascending=False)

fig = px.bar(heart\_clf\_out, y=i, x="Classifiers",text=i,color="Classifiers",

title="Comparison of {}".format(i),height=600,width=1000)

fig.show()

**Heart Attack DataSet**

****

**Data Source Link**

https://www.kaggle.com/rashikrahmanpritom/heart-attack-analysis-prediction-dataset