

An Analysis of Heart Disease Prediction Using Machine Learning and Deep Learning Techniques

Submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering

By

**CH RAJA SHAKER
REG. NO. 38110091**

**ANISETTI SIDHARTHA
REG. NO. 38110531**



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade "A" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119**

MAY 2022



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **CH RAJA SHAKER (Reg. No: 38110091), ANISETTI SIDHARTHA (Reg. No: 38110531)** who carried out the project entitled **"An Analysis of Heart Disease Prediction Using Machine Learning and Deep Learning Techniques"** under my supervision from December 2021 to March 2022.

Internal Guide

Mrs. M. D. Anto Praveena M.C.A., M.E., (Ph.D.),

Head of the Department

Dr. L. Lakshmanan., M.E., Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I, CH RAHA SHAKER (Reg No 38110091), ANISETTI SIDHARTHA (Reg No 38110531) hereby declare that the project report entitled “**An Analysis of Heart Disease Prediction Using Machine Learning and Deep Learning Techniques**” done by me under the guidance **Mrs. M. D. Anto Praveena M.C.A., M.E., (Ph.D.,)** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering.

DATE: 14-04-2022

The image shows two handwritten signatures in blue ink. The first signature is 'A. Sidhartha' and the second is 'Rajshukan'.

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

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

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D, Dean**, School of Computing, **Dr. S. Vigneshwari, M.E., Ph.D.** and **Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department of Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Mrs. M. D. Anto Praveena M.C.A., M.E., (Ph.D.,)** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

The significance of the heart as the body's most vital organ cannot be stressed. Heart disease is the leading cause of death worldwide. Heart failure (HF) is a main cause of death that must be successfully predicted (HF). Angiography, the gold standard for clinical diagnosis of HF, is expensive and can have catastrophic repercussions, according to research. In this scenario, machine learning and deep learning are applied. Machine learning and deep learning techniques can be used to forecast the whole range of hazards associated with this project. This dataset is created by combining previously available datasets with eleven distinct categories. This investigation would not be possible without this information. According to the findings, machine learning approaches exceeded deep learning in the diagnosis of cardiovascular diseases. We utilized the PCA approach to estimate the relative relevance of each of the dataset's 11 fields. When sample approaches were applied, accuracy and recall rates increased. According to the data, Random Forest Classifiers, Decision Tree Classifiers, and Nave Bayes algorithms surpass other ML algorithms.

TABLE OF CONTENTS

CHAPTER.NO	TITLE	PAGE.NO
	ABSTRACT	v
	LIST OF FIGURES	viii
	ABBREVIATIONS	x
1	INTRODUCTION	1
1.1	HEART DISEASES	1
1.2	WHAT IS MACHINE LEARNING	4
2	LITERATURE SURVEY	8
2.1	PREDICTION OF HEART DISEASE USING MACHINE LEARNING ALGRITHMS	8
2.1	EFFECTIVE HEART DISEASE PREDICTION USING HYBRID MACHINE LEARNING TECHNIQUES	9
2.3	APPLICATION OF MACHINE LEARNING IN DISEASES	9
2.4	CLASSIFICATION OF HEART DISEASE USING K-NEAREST NEIGHBOR AND GENETIC ALGORITHM	10
2.5	EARLY AND ACCURATE PREDICTION OF HEART DISEASE USING MACHINE LEARNING MODEL	10
3	AIM AND SCOPE OF THE INVESTIGATION	11
3.1	MOTIVATION	11
3.2	PROBLEM DEFINITION	11
3.3	OBJECTIVE OF THE PROJECT	12
3.4	EXISTING SYSTEM	12
3.5	PROPOSED SYSTEM	13
3.6	LIMITATIONS OF PROJECT	13
3.7	FEASIBILITY STUDY	14
3.8	SYSTEM REQUIREMENTS	14
3.9	SYSTEM ARCHITECTURE DIAGRAM	15
4	MODULES AND ALGORITHMS	16

4.1	DESCRIPTION OF THE DATASET	16
4.2	DATA PREPARATION	17
4.3	MACHINE LEARNING CLASSIFIERS PROPOSED	18
4.4	DEEP LEARNING ALGORITHMS PROPOSED	24
4.5	UML DIAGRAMS	26
5	RESULT AND DISCUSSION PERFORMANCE ANALYSIS	34
5.1	PERFORMANCE METRICS	35
5.2	PERFORMANCE ANALYSIS	35
6	SUMMARY AND CONCLUSIONS	38
	REFERENCES	39
	APPENDIX	
	A. SCREENSHOTS	41
	B. SOURCE CODE	43

LIST OF FIGURES

FIGURE NO	FIGURE NAME	PAGE NO
1.1	AI, ML, DL	3
3.1	System Architecture Diagram	15
4.1	Block diagram of the proposed method	18
4.2	Support Vector Machine Algorithm	19
4.3	Random forest Algorithm	20
4.4	Decision tree Algorithm	20
4.5	KNN Algorithm	20
4.6	Voting classifier	21
4.7	Naive Bayes algorithm	21
4.8	SVM linear kernel	22
4.9	SVM Polynomial kernel	22
4.10	SVM Gaussian Kernel	22
4.11	LSTM	23
4.12	CNN	23
4.13	Use Case	27
4.14	Class Diagram	28
4.15	Object Diagram	29
4.16	State Diagram	30

4.17	Activity Diagram	31
4.18	Sequence Diagram	32
4.19	Collaboration Diagram	33
5.1	Without PCA	36
5.2	With PCA & Feature Selection	36
5.3	With Lasso selected feature	37
5.4	Comparison Graph	37

ABBREVIATIONS

CNN	Concurrent Nueral Network
DL	Deep Learning
FN	False Negative
FP	False Positive
HF	Heart Failure
LSTM	Long Short Term Memory
ML	Machine Learning
PCA	Principal Component Analysis
RNN	Recurrent Nueral Network
TP	True Positive
TN	True Negative
UML	Unified Modeling Language

CHAPTER 1

INTRODUCTION

1.1 Heart Diseases

The most common cause of mortality and disability worldwide is cardiovascular disease, notably coronary artery disease, there are around 3,50,000 deaths annually as result of this ailment. This illness is responsible for the deaths of about 6,100,000, Americans per year. An estimated 7,35,000 Americans have a heart attack each year, with 5,25,000 having their first one and 2,10,000 having had one previously. In 22% of Asian nations, heart disease is the most common cause of death (in total heart disease mortality). Hypertensions, diabetes, smoking, and high cholesterol are just a few of the risk factors that can exacerbate heart disease. Heart disease can be difficult to identify as result of this. To forecast the severity of human heart disease, a variety of data mining and neural network algorithms have been used. Because of its uncertain nature, CHD should be approached with caution. There are serious ramifications for the heart and possibly death if the illness is not recognized early enough.

Shortness of breath, heart palpitations, and light-headedness are among the most prevalent symptoms. Blood clots, high blood pressure, diabetes, and the use of cigarettes and other medications are all risk factors for heart disease. An infection of the heart's inner membrane can cause fever, exhaustion, a dry cough, and skin rashes in patients who are experiencing heart problems. Bacteria, viruses, and parasites are the three most prevalent species that cause heart infections. Heart disease, heart failure, and congenital heart disease are only a few of the disorders that fall under the umbrella phrase "cardiovascular disease," yet they all fall under that umbrella category. Automated ways for detecting heart disease have evolved in recent years, anyone interested in learning more about machine learning techniques should start with this paper. Data for machine learning is prepared using the available resources. Heart disease can be predicted using a variety of risk factors. You are more prone to develop heart disease if you have a family history of heart disease, are

older, have high blood pressure, or have high cholesterol. Tobacco and alcohol usage, as well as obesity, have a significant impact. The lungs are inflamed.

A vast range of metabolic abnormalities can be discovered through medical research and data mining. The system can learn to live with CHD without coronary artery disease (CAD), which is the most frequent kind of heart disease (CHD). Estimated 3,50,000 persons per year succumb to this sickness and die as a result. Approximately 6,100,000 people in the United States die each year from the disease. Every year, around 7,35,000 people in the United States suffer a heart attack; 5,25,000 of these people have never had a heart attack before, while 2,10,000 have. When it comes to all-cause mortality, heart disease is the leading killer in 22% of Asian countries (in total heart disease mortality). Heart disease can be exacerbated by a number of risk factors, including hypertension, diabetes, smoking, and elevated cholesterol. It is difficult to diagnose heart problems as a result. Numerous data mining and neural network techniques have been used to determine how severe an individual's cardiac condition is. Because of its uncertain character, CHD should be treated with caution. Even death can result if the condition is not discovered in its earliest stages, which can have fatal consequences for the heart.

Shortness of breath, heart palpitations, and light-headedness are among the most prevalent symptoms. You'll need to address these concerns if you're at risk for heart disease due to high blood pressure, diabetes, or the use of cigarettes and other drugs. An infection of the heart's inner membrane can cause fever, exhaustion, dry cough, and skin rashes in patients with heart disease. Bacteria, viruses, and parasites are the most common culprits in cases of heart infection. Heart disease, heart failure, congenital heart disease, and angina pectoris are all examples of cardiovascular disease. Several automated methods for detecting heart disease have emerged in the last several years. Anyone interested in learning more about machine learning techniques should start with this paper. To prepare datasets for machine learning, we make use of the resources at our disposal. A number of risk variables can be used to predict heart disease. Heart disease risk factors include age, gender, blood pressure and cholesterol levels, as well as a family history of heart disease. Smoking, drinking,

and obesity all contribute to the problem. The lungs are inflamed.

Researchers can uncover a wide range of metabolic issues using medical research and data mining. Machine learning approaches enable a system to learn without having to be explicitly coded. Based on previous data, machine learning reasoning is possible. Machine learning can be applied in a wide number of industries. According to this study, the ability to detect heart problems is also hampered. Artificial intelligence (AI) is another term for machine learning, which includes deep learning (AI). In-depth study can aid a wide range of academic fields. Using this test, heart disease can also be anticipated. Machine learning approaches must be explicitly coded. Based on previous data, machine learning can make intelligent decisions. Machine learning can be used in a variety of industries. As this study reveals, the detection of heart disease is also compromised. Artificial intelligence is also known as machine learning, which includes deep learning (AI). Many academic fields can benefit from in-depth research. This test can also indicate heart disease, which is why it's so important.

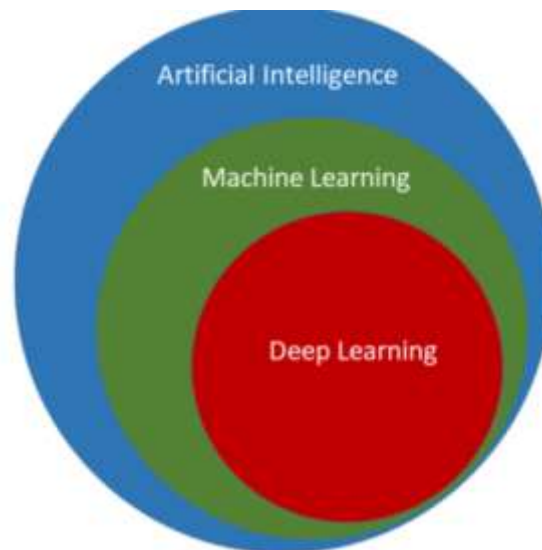


Fig:1.1: This diagram shows how AI, ML, and DL all work together. In the same way that DL is an area of ML, so too is ML an area of AI. Artificial Intelligence (AI), Deep Learning (DL), and Machine Learning (ML).

1.2 What is Machine Learning?

Before we take a look into the details of various machine learning methods, let's start by looking at what machine learning is, and what it isn't. Machine learning is often categorized as a subfield of artificial intelligence, but I find that categorization can often be misleading at first brush. The study of machine learning certainly arose from research in this context, but in the data science application of machine learning methods, it's more helpful to think of machine learning as a means of *building models of data*.

Fundamentally, machine learning involves building mathematical models to help understand data. "Learning" enters the fray when we give these models *tunable parameters* that can be adapted to observed data; in this way the program can be considered "learning" from the data. Once these models have been fit to previously seen data, they can be used to predict and understand aspects of newly observed data. I'll leave to the reader the more philosophical digression regarding the extent to which this type of mathematical, model-based "learning" is similar to the "learning" exhibited by the human brain. Understanding the problem setting in machine learning is essential to using these tools effectively, and so we will start with some broad categorizations of the types of approaches we'll discuss here.

1.2.1 Types of Machine Learning

- *Supervised Learning* – This involves learning from a training dataset with labeled data using classification and regression models. This learning process continues until the required level of performance is achieved.
- *Unsupervised Learning* – This involves using unlabeled data and then finding the underlying structure in the data in order to learn more and more about the data itself using factor and cluster analysis models.
- *Semi-supervised Learning* – This involves using unlabeled data like Unsupervised Learning with a small amount of labeled data. Using labeled data vastly increases the learning accuracy and is also more cost-effective than Supervised Learning.

- *Reinforcement Learning* – This involves learning optimal actions through trial and error. The next action is decided by learning behaviors that are based on the current state and that will maximize the reward in the future.

1.2.2 Need for Machine Learning

Human beings, at this moment, are the most intelligent and advanced species on earth because they can think, evaluate and solve complex problems. On the other side, AI is still in its initial stage and haven't surpassed human intelligence in many aspects. Then the question is that what is the need to make machine learn? The most suitable reason for doing this is, "to make decisions, based on data, with efficiency and scale".

Lately, organizations are investing heavily in newer technologies like Artificial Intelligence, Machine Learning and Deep Learning to get the key information from data to perform several real-world tasks and solve problems. We can call it data-driven decisions taken by machines, particularly to automate the process. These data-driven decisions can be used, instead of using programming logic, in the problems that cannot be programmed inherently. The fact is that we can't do without human intelligence, but other aspect is that we all need to solve real-world problems with efficiency at a huge scale. That is why the need for machine learning arises.

1.2.3 Advantages of Machine learning

Easily identifies trends and patterns - Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviors and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.

No human intervention needed (automation) - With ML, you don't need to babysit your project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on their own. A

common example of this is anti-virus software; they learn to filter new threats as they are recognized. ML is also good at recognizing spam.

Continuous Improvement - As ML algorithms gain experience, they keep improving in accuracy and efficiency. This lets them make better decisions. Say you need to make a weather forecast model. As the amount of data you have keeps growing, your algorithms learn to make more accurate predictions faster.

Handling multi-dimensional and multi-variety data - Machine Learning algorithms are good at handling data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments.

Wide Applications - You could be an e-tailer or a healthcare provider and make ML work for you. Where it does apply, it holds the capability to help deliver a much more personal experience to customers while also targeting the right customers.

1.2.4 Challenges in Machines Learning

While Machine Learning is rapidly evolving, making significant strides with cybersecurity and autonomous cars, this segment of AI as whole still has a long way to go. The reason behind is that ML has not been able to overcome number of challenges. The challenges that ML is facing currently are –

Quality of data – Having good-quality data for ML algorithms is one of the biggest challenges. Use of low-quality data leads to the problems related to data preprocessing and feature extraction.

Time-Consuming task – Another challenge faced by ML models is the consumption of time especially for data acquisition, feature extraction and retrieval.

Lack of specialist persons – As ML technology is still in its infancy stage, availability of expert resources is a tough job.

No clear objective for formulating business problems – Having no clear objective and well-defined goal for business problems is another key challenge for ML because this technology is not that mature yet.

Issue of overfitting & underfitting – If the model is overfitting or underfitting, it cannot be represented well for the problem.

Curse of dimensionality – Another challenge ML model faces is too many features of data points. This can be a real hindrance.

Difficulty in deployment – Complexity of the ML model makes it quite difficult to be deployed in real life.

1.2.5 Applications of Machines Learning

Machine Learning is the most rapidly growing technology and according to researchers we are in the golden year of AI and ML. It is used to solve many real-world complex problems which cannot be solved with traditional approach. Following are some real-world applications of ML –

- Emotion analysis
- Sentiment analysis
- Error detection and prevention
- Weather forecasting and prediction
- Stock market analysis and forecasting
- Speech synthesis
- Speech recognition
- Customer segmentation
- Object recognition
- Fraud detection
- Fraud prevention
- Recommendation of products to customer in online shopping

CHAPTER 2

LITERATURE SURVEY

2. 1 Prediction of heart disease using machine learning algorithms

AUTHORS: Santhana Krishnan J and Geetha S,

ABSTRACT: Health care field has a vast amount of data, for processing those data certain techniques are used. Data mining is one of the techniques often used. Heart disease is the Leading cause of death worldwide. This System predicts the arising possibilities of Heart Disease. The outcomes of this system provide the chances of occurring heart disease in terms of percentage. The datasets used are classified in terms of medical parameters. This system evaluates those parameters using data mining classification technique. The datasets are processed in python programming using two main Machine Learning Algorithm namely Decision Tree Algorithm and Naive Bayes Algorithm which shows the best algorithm among these two in terms of accuracy level of heart disease.

2.2 Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

AUTHORS: Senthil kumar mohan, chandrasegar thirumalai and Gautam Srivastva,

ABSTRACT: Heart disease is one of the most significant causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. We have also seen ML techniques being used in recent developments in different areas of the Internet of Things (IoT). Various studies give only a glimpse into predicting heart disease with ML techniques. In this paper, we propose a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques. We produce an enhanced performance level with an accuracy level of 88.7% through the prediction

model for heart disease with the hybrid random forest with a linear model (HRFLM).

2.3 Application of Machine Learning in Diseases Prediction

AUTHORS: Pahulpreet Singh Kohli and Shriya Arora

ABSTRACT: The application of machine learning in the field of medical diagnosis is increasing gradually. This can be contributed primarily to the improvement in the classification and recognition systems used in disease diagnosis which is able to provide data that aids medical experts in early detection of fatal diseases and therefore, increase the survival rate of patients significantly. In this paper, we apply different classification algorithms, each with its own advantage on three separate databases of disease (Heart, Breast cancer, Diabetes) available in UCI repository for disease prediction. The feature selection for each dataset was accomplished by backward modeling using the p-value test. The results of the study strengthen the idea of the application of machine learning in early detection of diseases.

2.4 Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm

AUTHORS: M. Akhil, B. L. Deekshatulu, and P. Chandra

ABSTRACT: Data mining techniques have been widely used to mine knowledgeable information from medical data bases. In data mining classification is a supervised learning that can be used to design models describing important data classes, where class attribute is involved in the construction of the classifier. Nearest neighbor (KNN) is very simple, most popular, highly efficient and effective algorithm for pattern recognition. KNN is a straight forward classifier, where samples are classified based on the class of their nearest neighbor. Medical data bases are high volume in nature. If the data set contains redundant and irrelevant attributes, classification may produce less accurate result. Heart disease is the leading cause of death in INDIA. In Andhra Pradesh heart disease was the leading cause of mortality accounting for 32% of all deaths, a rate as high as Canada (35%) and USA. Hence there is a need to define a decision support system that helps clinicians decide to take precautionary steps. In

this paper we propose a new algorithm which combines KNN with genetic algorithm for effective classification. Genetic algorithms perform global search in complex large and multimodal landscapes and provide optimal solution. Experimental results shows that our algorithm enhance the accuracy in diagnosis of heart disease.

2.5 Early and Accurate Prediction of Heart Disease Using Machine Learning Model

AUTHORS: B Padmaja, Chintala Srinidhi, Kotha Sindhu, Kalali Vanaja

ABSTRACT: Heart disease is one of the critical health issues and many people across the world are suffering with this disease. It is important to identify this disease in early stages to save many lives. The purpose of this article is to design a model to predict the heart diseases using machine learning techniques. This model is developed using classification algorithms, as they play important role in prediction. The model is developed using different classification algorithms which include Logistic Regression, Random Forest, Support vector machine, Gaussian Naïve Bayes, Gradient boosting, K-nearest neighbors, Multinomial Naïve bayes and Decision trees. Cleveland data repository is used to train and test the classifiers. In addition to this, feature selection algorithm named chi square is used to select key features from the input data set, which will decrease the execution time and increases the performance of the classifiers. Out of all the classifiers evaluated using performance metrics, Random forest is giving good accuracy. So, the model built using Random forest is efficient and feasible solution in identifying heart diseases and it can be implemented in healthcare which plays key role in the stream of cardiology.

CHAPTER 3

AIM AND SCOPE OF THE INVESTIGATION

3.1 MOTIVATION

Machine learning, which incorporates deep learning, is referred to as artificial intelligence (AI). In-depth research can help in a variety of academic subjects. Heart disease can also be predicted with this test. Machine learning may make intelligent conclusions based on past data. In-depth study might be beneficial in a variety of academic subjects. This test can also reveal the presence of heart problems, that's why it's so essential.

3.2 PROBLEM DEFINITION

Diagnosis and prediction of heart related diseases requires more precision, perfection and correctness because a little mistake can cause fatigue problem or death of the person, there are numerous death cases related to heart and their counting is increasing exponentially day by day. To deal with the problem there is essential need of prediction system for awareness about diseases.

3.3 OBJECTIVE OF PROJECT

By combining the machine learning with ensemble learning. In training and testing, the use of a 303-occurrence dataset for cardiac illnesses from Cleveland proved to be crucial. It was utilized to obtain more data than would otherwise be possible. We employed tenfold cross validation. Decision Trees, Nave-Bayes, K-Nearest Neighbor, a Radial Basis Function, and Support Vector Machines were among the classifiers utilized. The classification of the dataset was predicted using an ensemble of classifiers.

3.4 EXISTING SYSTEM

The obtained results are compared with the results of existing models within the same domain and found to be improved. The data of heart disease patients collected from the UCI laboratory is used to discover patterns with NN, DT, Support Vector

machines SVM, and Naive Bayes. The results are compared for performance and accuracy with these algorithms. The proposed hybrid method returns results of 86.8% for F-measure, competing with the other existing methods. The classification without segmentation of Convolutional Neural Networks (CNN) is introduced. This method considers the heart cycles with various start positions from the Electrocardiogram (ECG) signals in the training phase.

3.4.1 DISADVANTAGES OF EXISTING SYSTEM

Less importance was given to determining the relationships between these features and to identifying their level of priority within the prediction model.

3.5 PROPOSED SYSTEM

In this scenario, machine learning and deep learning are applied. Machine learning and deep learning techniques can be used to forecast the whole range of hazards associated with this project. This dataset is created by combining previously available datasets. For your convenience, we've sorted them into eleven distinct categories. Our investigation would not be possible without this information. According to the findings, machine learning approaches exceeded deep learning in the diagnosis of cardiovascular diseases. We utilized the PCA approach to estimate the relative relevance of each of the dataset's 11 fields. When sample approaches were applied, accuracy and recall rates increased. According to the data, Random Forest Classifiers, Decision Tree Classifiers, and Nave Bayes algorithms surpass other ML algorithms.

3.5.1 ADVANTAGES OF PROPOSED SYSTEM

- There is essential need of prediction system for awareness about diseases.
- Better accuracy and precision.

3.6 LIMITATIONS OF PROJECT

Medical research and data mining can be used to identify a wide range of metabolic disorders. Machine learning approaches allow a system to learn without having to

code it explicitly. From previous data it is evident that machine learning reasoning is conceivable. Machine learning, which incorporates deep learning, is referred to as artificial intelligence (AI). In-depth research can help in a variety of academic subjects. Heart disease can also be predicted with this test. Machine learning may make intelligent conclusions based on past data. In-depth study might be beneficial in a variety of academic subjects. This test can also reveal the presence of heart problems, that's why it's so essential.

3.7 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are:

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

3.7.1 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

3.7.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The

developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

3.7.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

3.8 SYSTEM REQUIREMENTS

3.8.1 SOFTWARE REQUIREMENTS

The functional requirements or the overall description documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraints and user documentation.

The appropriation of requirements and implementation constraints gives the general overview of the project in regard to what the areas of strength and deficit are and how to tackle them.

- Python idle 3.7 version(or)
- Anaconda 3.7(or)
- Jupiter(or)
- Google Colab

3.8.2 HARDWARE REQUIREMENTS

Minimum hardware requirements are very dependent on the particular software being developed by a given Enthought Python / Canopy / VS Code user. Applications that need to store large arrays/objects in memory will require more RAM, whereas applications that need to perform numerous calculations or tasks more quickly will require a faster processor.

- Operating system : windows, linux
- Processor : minimum intel i3
- Ram : minimum 4gb
- Hard disk : minimum 250gb

3.9 SYSTEM ARCHITECTURE DIAGRAM

A system architecture diagram is the distribution of the functional correspondences. These are formal elements, the embodiment of concepts and information. Architecture defines the relations between elements, amongst features, and the surrounding elements. For architecture diagram of this model refer to figure 3.1. The main purpose of this diagram is to facilitate collaboration, to increase communication, and to provide vision and guidance.

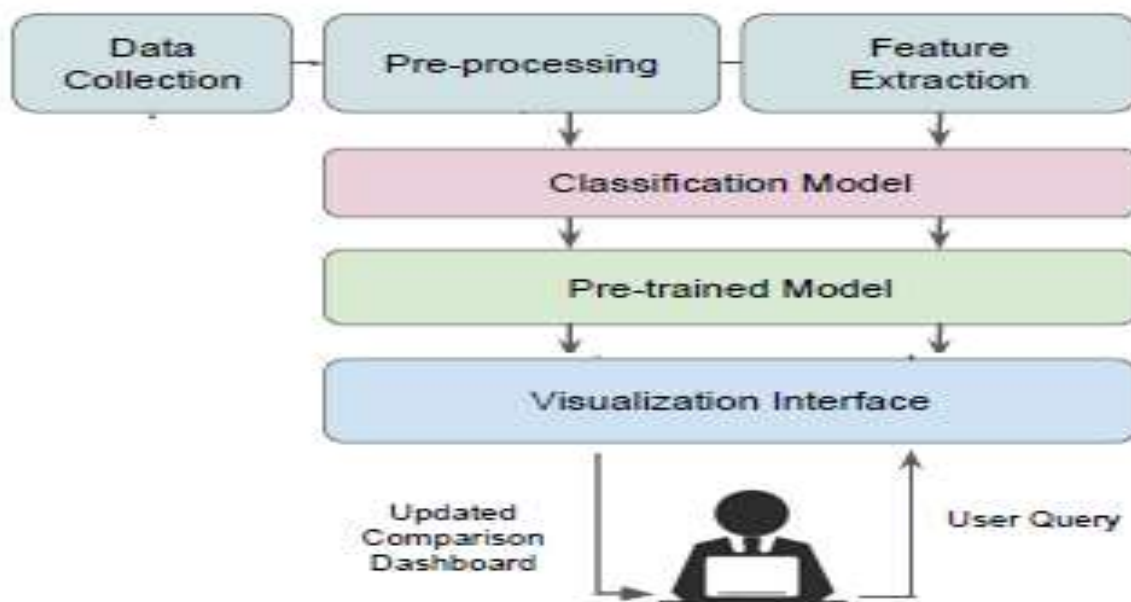


Fig:3.1: System Architecture Diagram

CHAPTER 4

MODULES AND ALGORITHMS

4.1 Description of the Dataset

People were recruited for the study based on their history of cardiac issues as well as other medical concerns. This dataset provides much-needed information on the patients' medical features, such as age, sex, resting blood pressure, fasting blood sugar level, and so on, allowing us to identify whether the patient has been diagnosed with heart disease. The attributes information is summarized in table 4.1. We combined them based on 11 similar criteria, resulting in the world's largest heart disease dataset. This dataset reveals a pattern that may be used to identify persons who are at risk for heart disease.

The five datasets used for its curation are:

- i. Cleveland: 303
- ii. Hungarian: 294
- iii. Switzerland: 123
- iv. Long Beach VA: 200
- v. Statlog (Heart) Data Set: 270

TABLE 4.1: Dataset Attributes Description.

Observation	Description
age	The patient's age is expressed in years. (Numeric)
sex	The patient's gender will be denoted by a 1 for men and a 0 for females. (Nominal)
chest pain type	Chest pain can take four different forms. Asymptomatic angina is equals to 0, atypical angina is equals to 1, non-angina pain equals to 2, and classic angina equals to 3.

resting bp	Level of blood sugar at resting mode in mm/hg (Numerical) the level of blood pressure in the resting state. The standard range is 120/80.
cholesterol	The amount of triglycerides in the blood is measured in milligrams per deciliter (mg/dl) (numerical).
Fasting blood sugar	In actual circumstances, fasting blood sugar levels > 120 mg/dl are represented as 1; in false cases, they are represented as 0.
Max heart rate	Attained maximum heart rate.
Restingecg (values -0,1,2)	An ECG taken while at rest can be displayed in three distinct ways.
exercise angina	Exercise-induced angina is a type of chest pain caused by a decrease in blood flow to the heart (1=yes; 0=no).
Old peak	Activity produces ST depression when compared to rest.
ST slope	The slope of the peak exercise's ST portion.
target	A heart risk of 1 indicates coronary artery disease, while a heart risk of 0 indicates a healthy heart.

4.2 Data Preparation

There were no significant outliers or null values in the dataset, and no major outliers were observed. Both PCA and a feature selection technique, as well as applying the data directly to machine learning and deep learning algorithms, produced identical results. As a result, we decided to stick with the raw data we had. Utilized the PCA approach to determine the relative relevance of each of the dataset's 11 fields, and then narrow the list down to the most critical ones. The Lasso algorithm, an embedded technique, is used to select only the most important qualities when performing feature selection. Due to our model's excellent correlation with the

objective variable, feature selection does not greatly improve our results. Through the collection and analysis of data on certain variables in an established system, it is feasible to answer relevant questions and measure outcomes. Feature selection helps in the building of predictive models by decreasing the number of input variables. Since processing costs are reduced and some model performance may be improved, cutting input variables is favorable. For block diagram of the proposed method refer to figure 4.1.

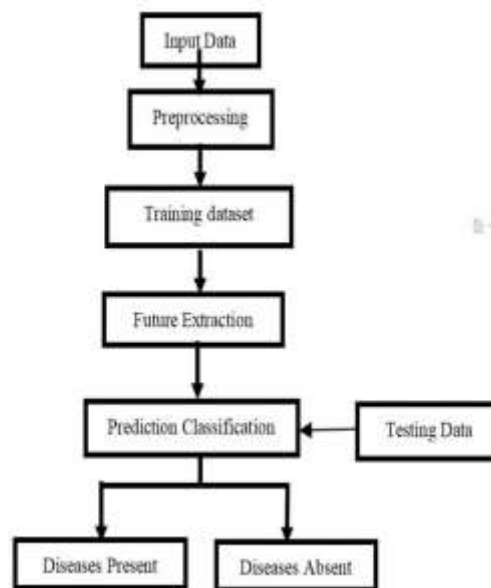


Fig: 4.1: Block diagram of the proposed method

4.3 Machine Learning Classifiers Proposed

Machine Learning may be applied to a wide range of circumstances, thanks to a variety of different techniques (ML). The following components typically make up an ML pipeline:

- Data Collection
- Data Preparation
- Feature Extraction
- Feature Selection
- Supervised/Unsupervised/Reinforcement learning task

Most ML goals require a good representation of features. Because of their ability to learn feature representations at a number of abstraction levels, these representations were often generated by specialists in their field prior to the emergence of Deep Learning (DL) technologies.

In this context, ML Algorithms used are:

1. Support Vector Machine
2. Random Forest Classifier
3. Decision Tree Classifier
4. KNeighbors Classifier
5. Voting Classifier
6. Naive Bayes
7. SVM - Linear Kernel
8. SVM - Polynomial Kernel
9. SVM - Gaussian Kernel
10. SVM - Sigmoid Kernel

4.3.1 Support Vector Machine

A supervised machine learning technique employs target-labeled training data to discover the optimum hyper-plane. We are trying to find a vantage point from which socioeconomic classes can be discerned in our study. It's important to select the hyperplane with the minimum margin of uncertainty that best optimizes class section distances. The hyper-plane outlined before can be used to classify a new data point.

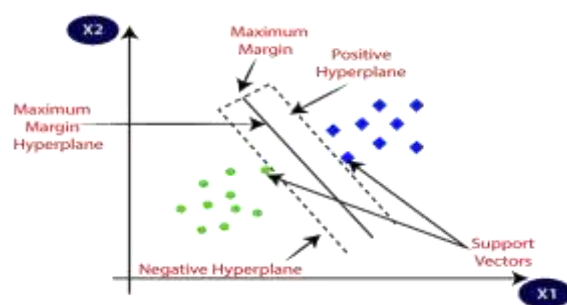


Fig:4.2: Support Vector Machine Algorithm

4.3.2 Random Forest Classifier

Meta-estimator Random Forests are meta-estimators because they combine many decision tree classifiers into a single model. This allows for greater accuracy and better control over over-fitting.

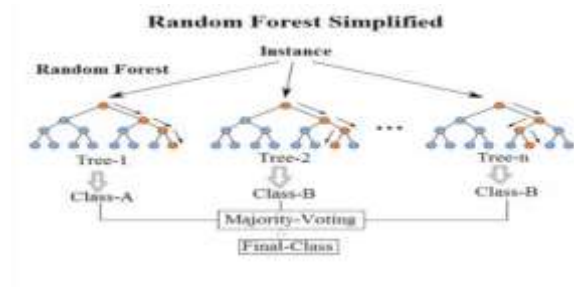


Fig:4.3: Random forest Algorithm

4.3.3 Decision Tree

It is possible to use Decision Tree as a supervised learning strategy for both continuous and categorical variables. Once the dataset has been separated into categories based on the most important predictors, classification begins. Each attribute's entropy is then calculated. The attribute with the greatest information gain or the least amount of entropy is selected as a splitting attribute. The rest of the properties are treated in the same way.

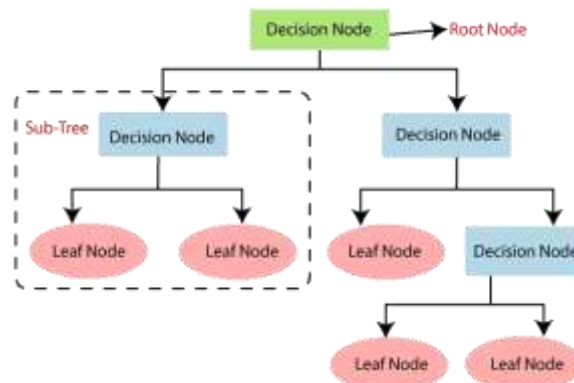


Fig:4.4: Decision tree Algorithm

4.3.4 KNeighbors Classifier

Simple and widely used supervised classification is the KNN model developed by Fix and Hodge. If the test and training data points are far apart, it uses this information to find its closest neighbors. Thereafter, it will be assigned to the class of its nearest neighbor. KNN stands for the number of nearest neighbors. "K" has a significant impact on the grading process. KNN uses a range of metrics, including Euclidean, Manhattan, and Minkowsky, to analyze numerical data.

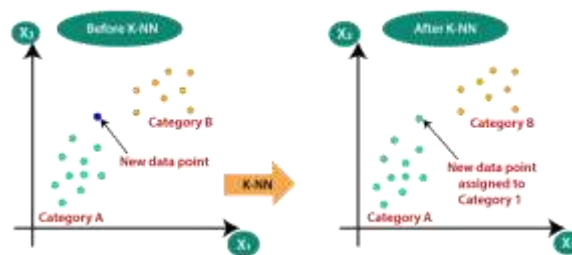


Fig:4.5: KNN Algorithm

4.3.5 Voting Classifier

Based on the outcomes of the various base models, a voting classifier can make predictions about the outcome of an election. Voting decisions can be used to combine the aggregating criteria for each estimator output.

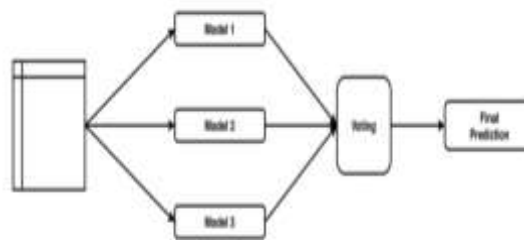


Fig:4.6: Voting classifier

4.3.6 Naïve bayes

Based on the Bayes Theorem, a set of classification methods known as naive Bayes classifiers has been developed. The essential assumption of this algorithmic family is that each pair of features being classified is distinct from the others.

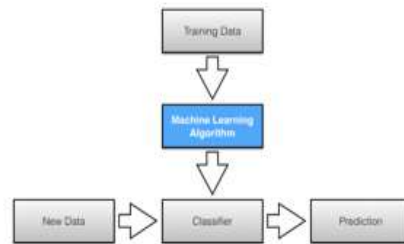


Fig:4.7: Naïve bayes algorithm

4.3.7 SVM-Linear Kernel

The Linear Kernel is the best choice for an SVM model if the data can be divided by a single line. Another commonly used kernel is bzip2. Data sets with a greater number of features benefit more from these techniques.

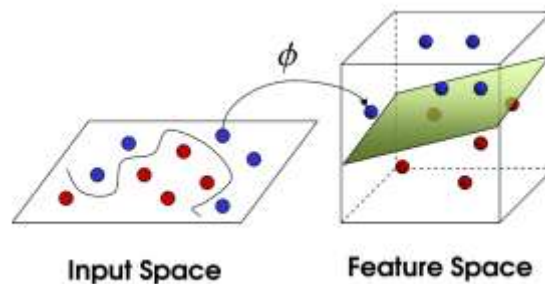


Fig:4.8: SVM linear kernel

4.3.8 SVM-Polynomial Kernel

Since it reflects the similarity of vectors in a feature space over polynomials of the original variables, the polynomial kernel can be used to train non-linear models. Support vector machines (SVMs), a type of kernelized model, are often used in conjunction with this method.

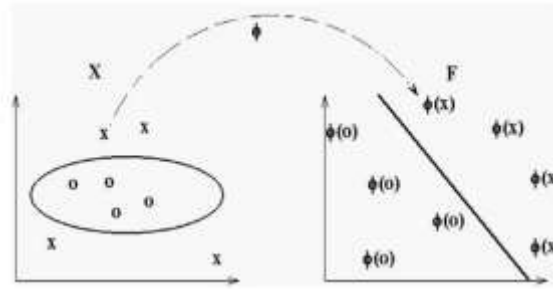


Fig:4.9: SVM polynomial kernel

4.3.9 SVM-Gaussian Kernel

Another prominent SVM kernel method is the Gaussian RBF kernel (Radial Basis Function). As the distance from the origin to a particular location grows, the RBF kernel's value decreases. The format of a Gaussian Kernel is as follows: $\|X_1 - X_2\|$ = Euclidean distance between X_1 and X_2 .

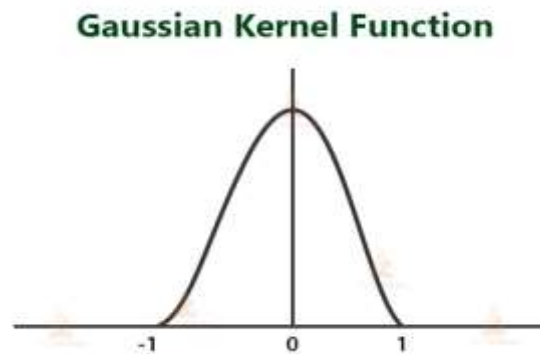


Fig:4.10: SVM Gaussian kernel

4.3.10 SVM-Sigmoid Kernel

As an artificial neuron activation function in neural networks, the Sigmoid Kernel uses the bipolar sigmoid function. A SVM model with a sigmoid kernel function is the same as a two-layer perceptron neural network.

4.4 Deep Learning Algorithms Proposed

Because DL tasks might be computationally costly, Deep Learning is advantageous.

Many of the original concepts of DL were discovered years before they were put into practice. The availability of less costly processing resources has aided the development of DL applications.

In this context, DL Algorithms used are:

1. LSTM
2. CNN
3. RNN

4.4.1 LSTM

Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, can be used by neural networks capable of learning sequence prediction challenges. What is the major objective of this project? For example, detecting changes or foreseeing difficulties that may arise as result of the incorrect data.

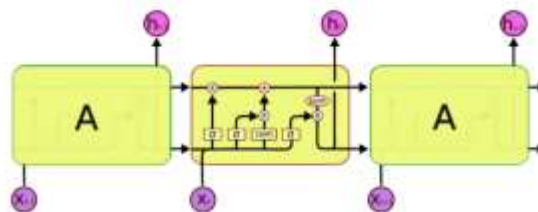


Fig:4.11: LSTM diagram

4.4.2 CNN

One of the most used neural networks for processing organized arrays of data, such as images, is the convolutional neural network (CNN).

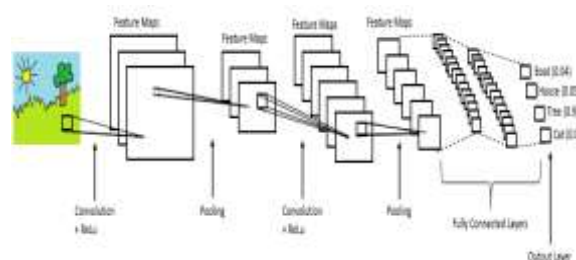


Fig:4.12: CNN Diagram

A sort of neural network known as recurrent neural networks (RNN) can make it easier to model sequence data. Because they are built on feed forward networks, RNNs are an exact replica of what happens in the human brain. Recurrent neural networks, on the other hand, were able to predict sequential data in a way that other algorithms are unable to.

4.5 UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object-oriented computer software. In its current form UML is comprised of two major components: A Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing object oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

4.5.1 GOALS

The Primary goals in the design of the UML are as follows:

- i. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
- ii. Provide extendibility and specialization mechanisms to extend the core concepts.

- iii. Be independent of programming languages and development process.
- iv. Provide a formal basis for understanding the modeling language.
- v. Encourage the growth of OO tools market.
- vi. Support higher level development concepts such as collaborations, frameworks, patterns and components.
- vii. Integrate best practices.

4.5.2 Use case diagram

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted. For use case diagram of the model refer to figure 4.12.

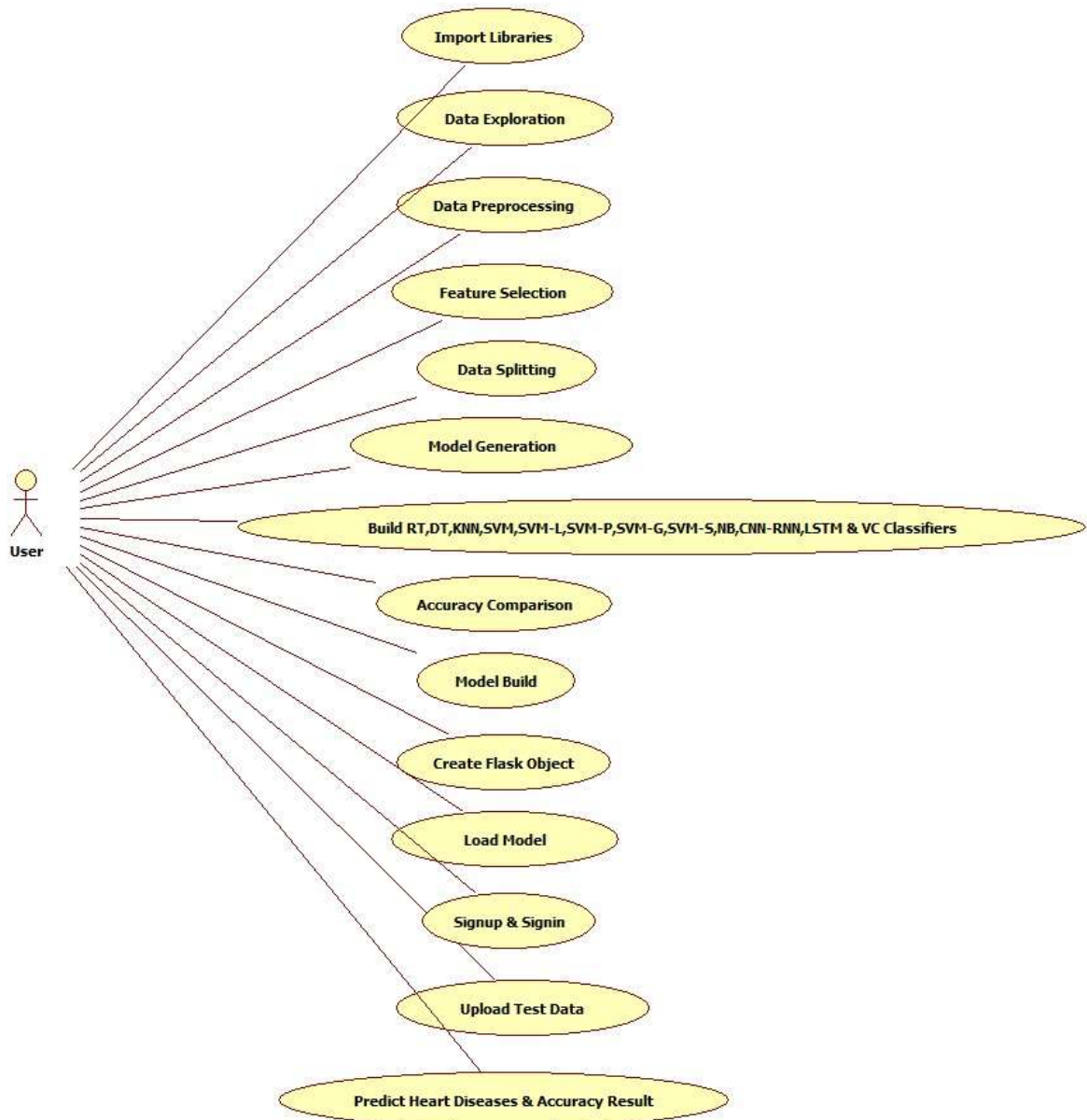


Fig:4.12: Use Case Diagram

4.5.3 Class diagram

The class diagram is used to refine the use case diagram and define a detailed design of the system. The class diagram classifies the actors defined in the use case diagram into a set of interrelated classes. The relationship or association between the classes can be either an "is-a" or "has-a" relationship. Each class in the class diagram

may be capable of providing certain functionalities. These functionalities provided by the class are termed "methods" of the class. Apart from this, each class may have certain "attributes" that uniquely identify the class. For class diagram of the model refer to figure 4.13.

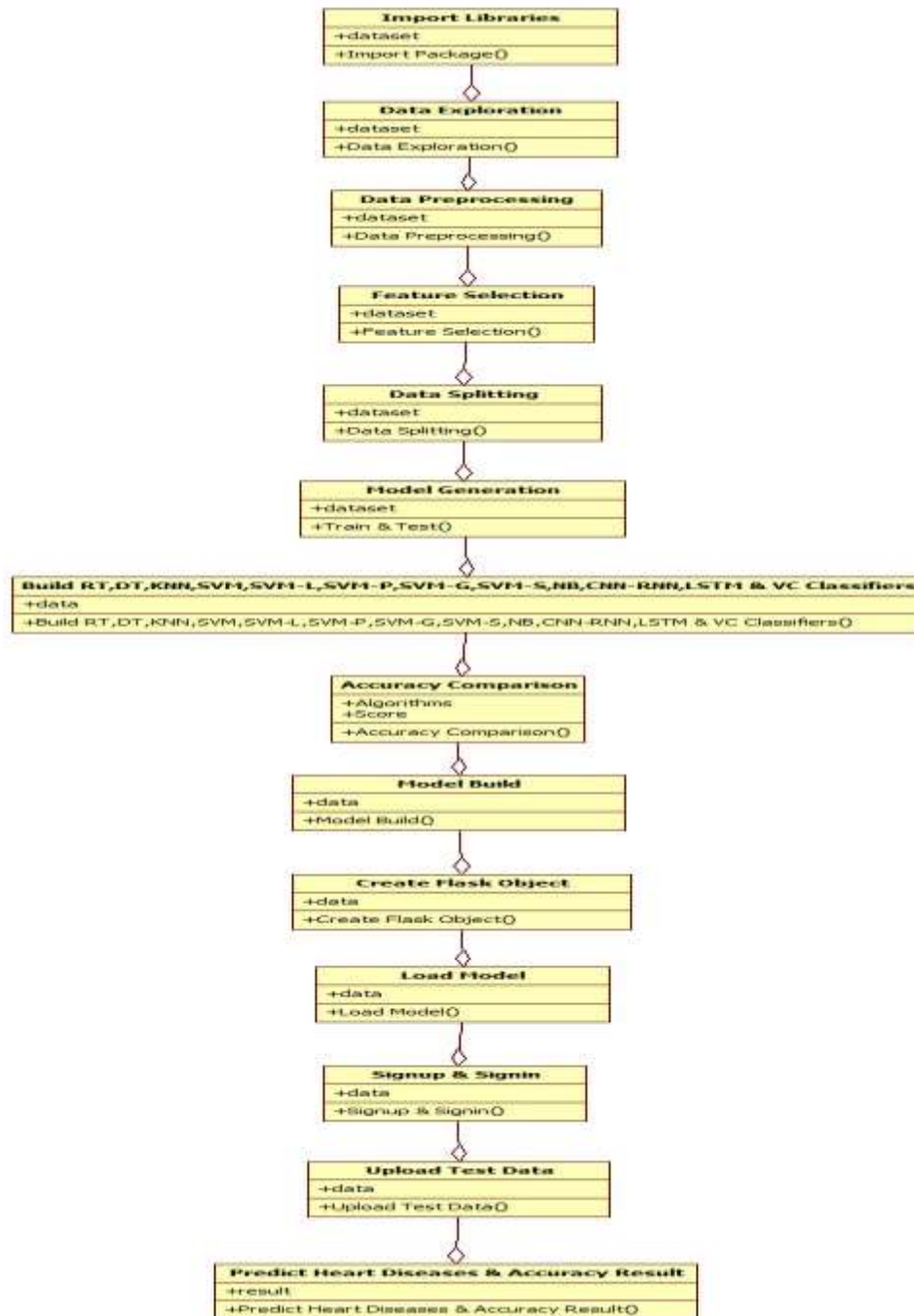


Fig:4.13: Class Diagram

4.5.4 Object diagram

The object diagram is a special kind of class diagram. An object is an instance of a class. This essentially means that an object represents the state of a class at a given point of time while the system is running. The object diagram captures the state of different classes in the system and their relationships or associations at a given point of time. For object diagram of the model refer to figure 4.14.

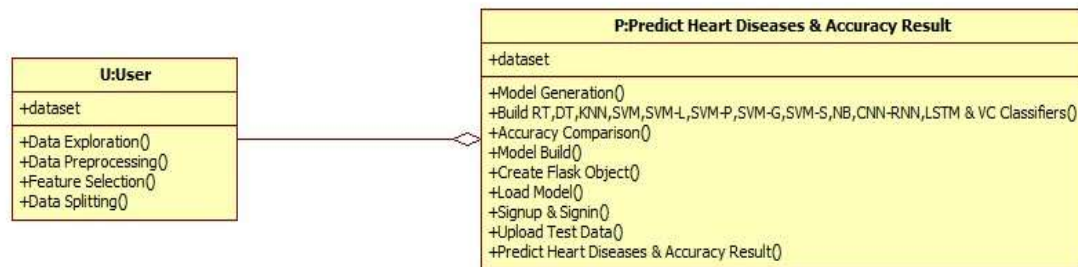


Fig: 4.14: Object Diagram

4.5.5 State diagram

A state diagram, as the name suggests, represents the different states that objects in the system undergo during their life cycle. Objects in the system change states in response to events. In addition to this, a state diagram also captures the transition of the object's state from an initial state to a final state in response to events affecting the system. For state diagram of the model refer to figure 4.15.

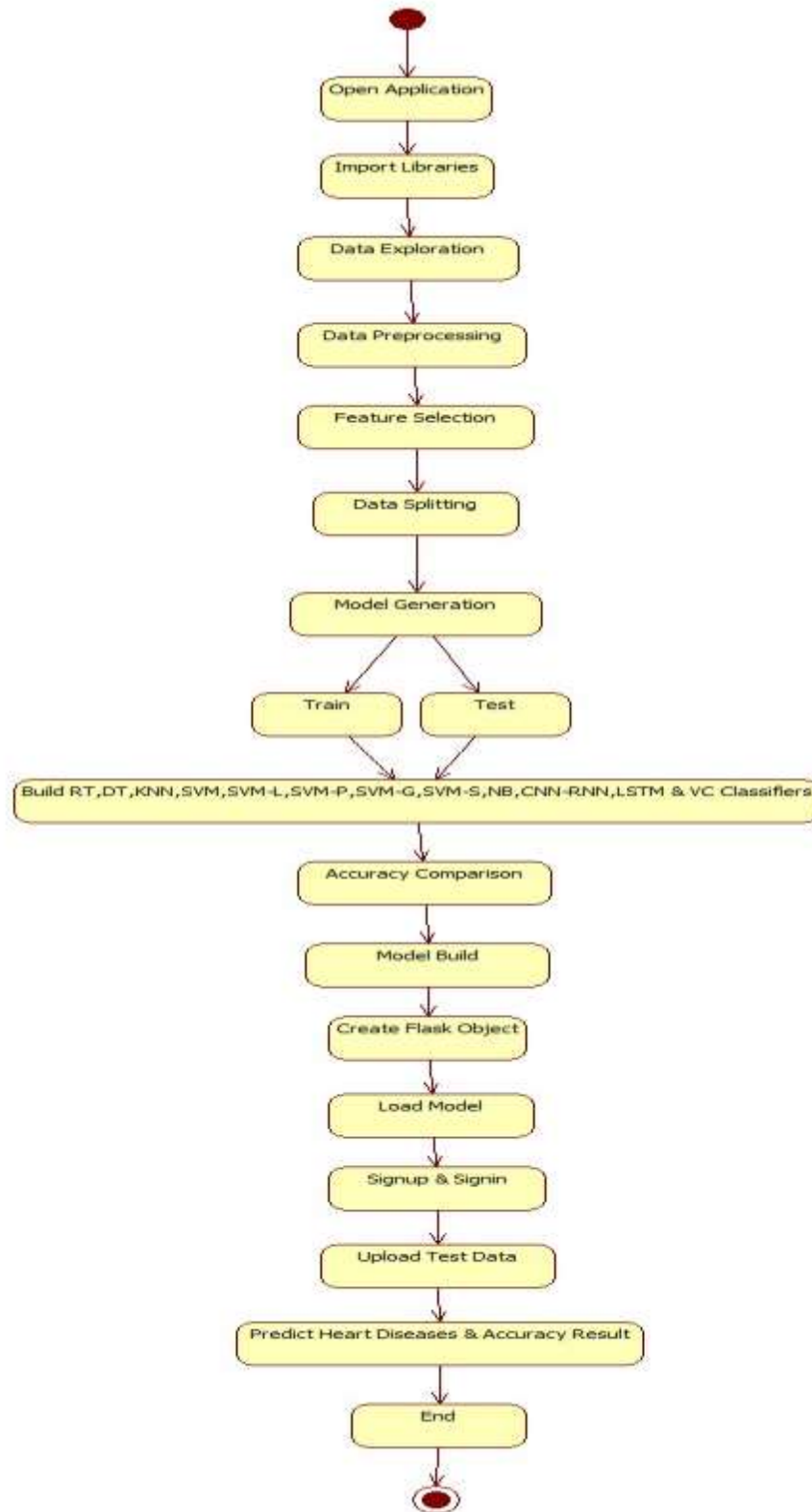


Fig:4.15: State Diagram

4.5.6 Activity diagram

The process flows in the system are captured in the activity diagram. Similar to a state diagram, an activity diagram also consists of activities, actions, transitions, initial and final states, and guard conditions. For activity diagram of the model refer to figure 4.16.

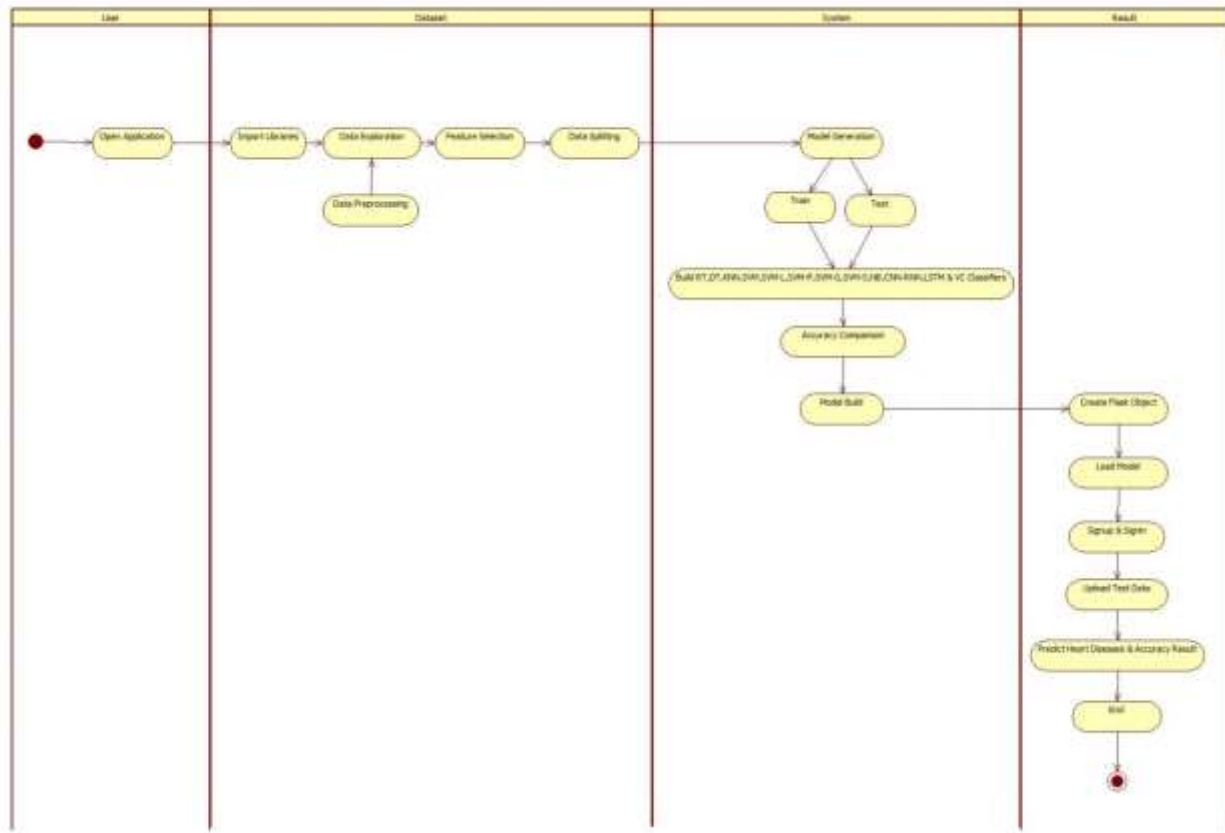


Fig:4.16: Activity Diagram

4.5.7 Sequence diagram

A sequence diagram represents the interaction between different objects in the system. The important aspect of a sequence diagram is that it is time-ordered. This means that the exact sequence of the interactions between the objects is represented step by step. Different objects in the sequence diagram interact with each other by passing "messages". For sequence diagram of the model refer to figure 4.17.

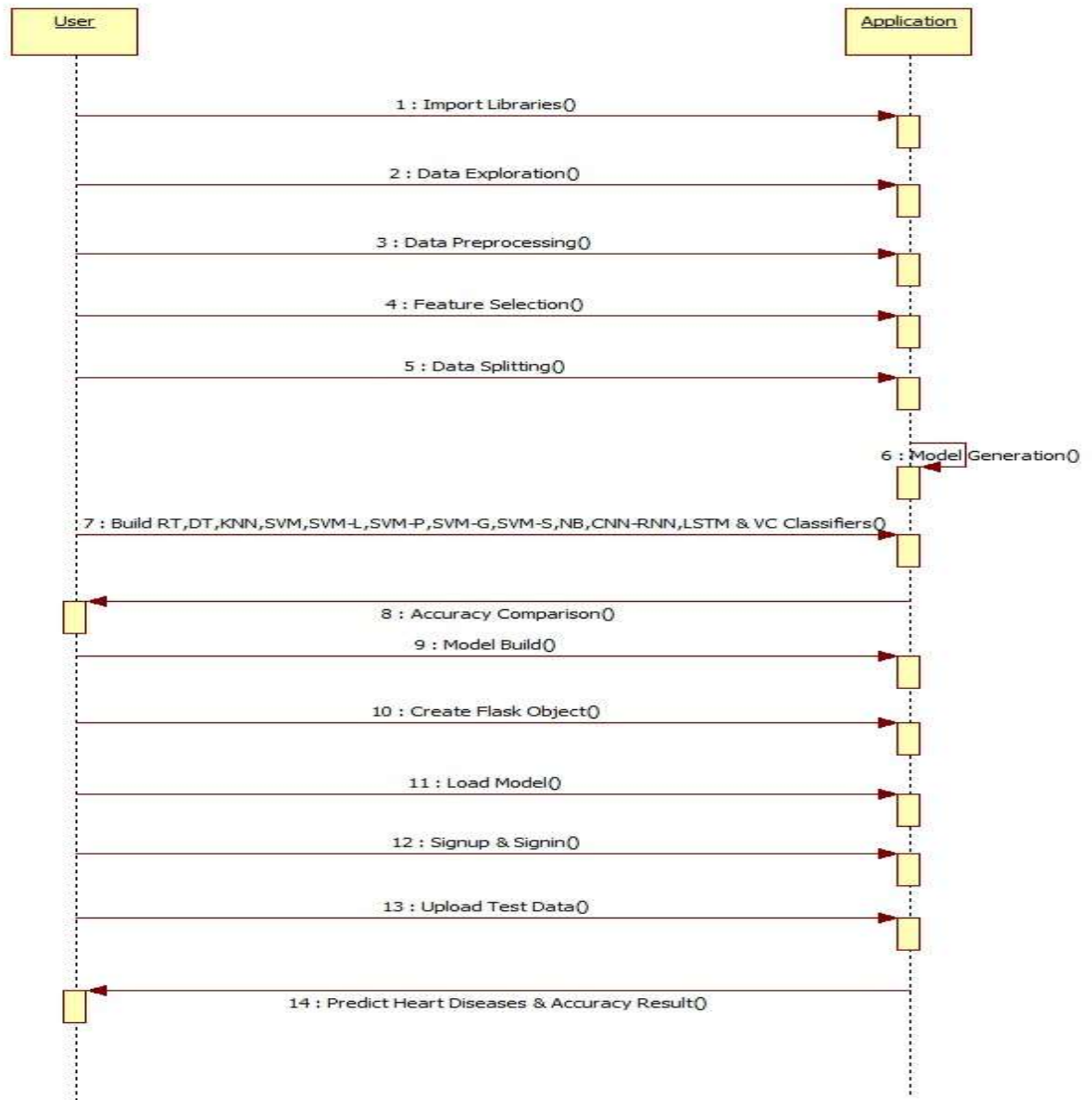


Fig: 4.17: Sequence Diagram

4.5.8 Collaboration diagram

A collaboration diagram groups together the interactions between different objects. The interactions are listed as numbered interactions that help to trace the sequence of the interactions. The collaboration diagram helps to identify all the possible interactions that each object has with other objects. For collaboration diagram of the model refer to figure 4.18.

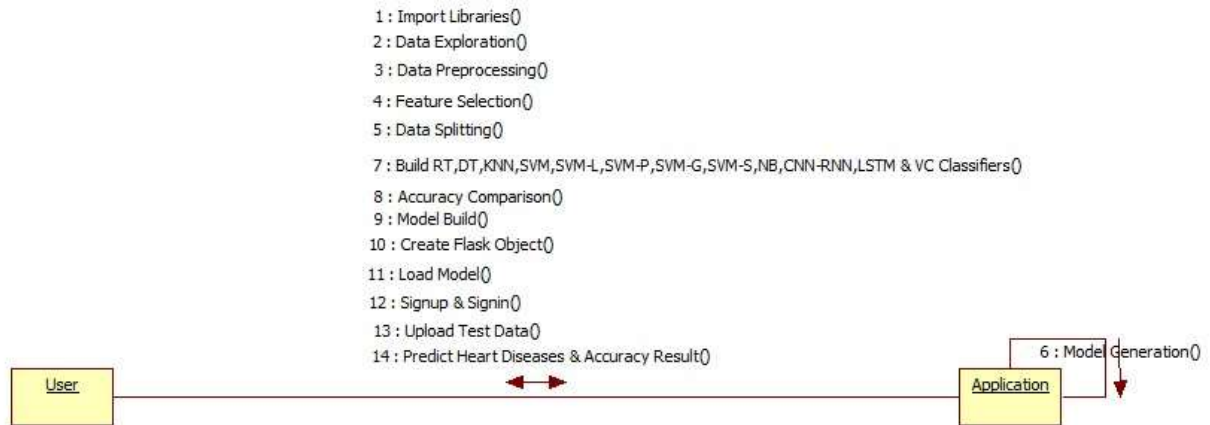


Fig:4.18: Collaboration diagram

CHAPTER 5

RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS

5.1 Performance Metrics

Four performance criteria were used to evaluate the proposed and current approaches such as accuracy, preciseness, recall and f1. Here FN stands for false negative, FP for false positive, and TN stands for true negative, TP stands for True Positive as seen in the instances in the below formula. The test results are summarized in table 5.1.

5.1.1 Accuracy

Accuracy is the measure of trueness in the prediction instance, which provides the correctness of the prediction.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

TABLE 5.1: Performance evaluation in light of accuracy.

Technique	Accuracy (%)
Support Vector Machine	69
Random Forest Classifier	93
Decision Tree Classifier	86
KNeighbors Classifier	68
Voting Classifier	88
Naive Bayes	85
SVM – Linear Kernel	80
SVM – Polynomial Kernel	68

SVM – Gaussian Kernel	69
SVM – Sigmoid Kernel	46
LSTM	62
CNN - RNN	63

5.1.2 Precision

The ratio of accurately anticipated positive observations to total positive observations expected is known as precision.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

4.1.3 Recall (Sensitivity)

The percentage of the class's actual observations that were predicted to be positive.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

4.1.4 F1 score

In other words, the F score or F Measure is a measure that strikes a healthy balance between accuracy and recall. When false positives and false negatives are crucial, it is useful to have.

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

5.2 Performance Analysis

In the prediction of heart disease, the Random Forest Classifier, Nave Bayes, and Decision Tree Classifiers beat other algorithms like SVC and KNN. We used algorithms that were more precise, cost-effective, and faster than those used in

previous studies. Additionally, machine learning techniques outperform deep learning in this study. Using the dataset's updated medical information, we can say that our accuracy has increased. Predicting heart disease patients is easier with machine learning than with deep learning, as well. This reveals that machine learning is better at detecting problems in the heart than human experts. For accuracies of algorithms applied on heart disease datasets along with other techniques refer to figure 5.1, 5.2, 5.3 and 5.4.

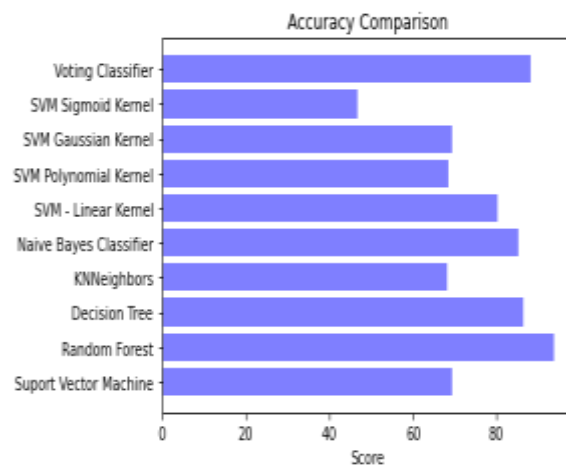


Fig:5.1: Without PCA.

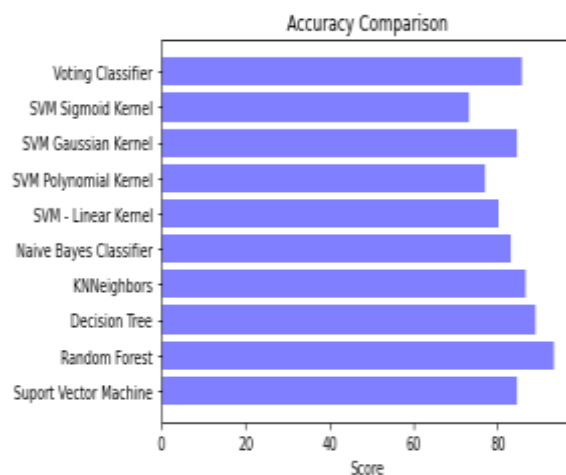


Fig:5.2: With PCA & Feature Selection.

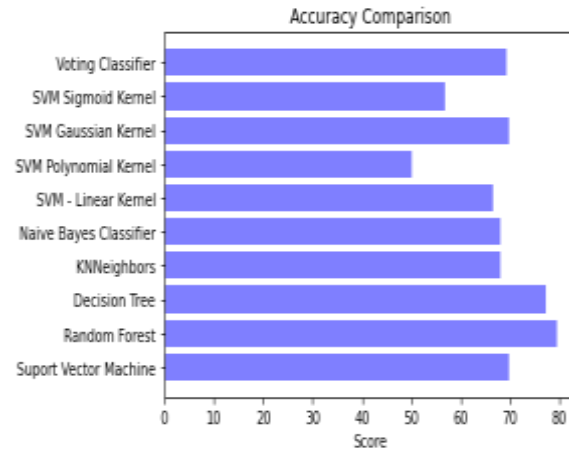


Fig:5.3: With LASSO selected features.

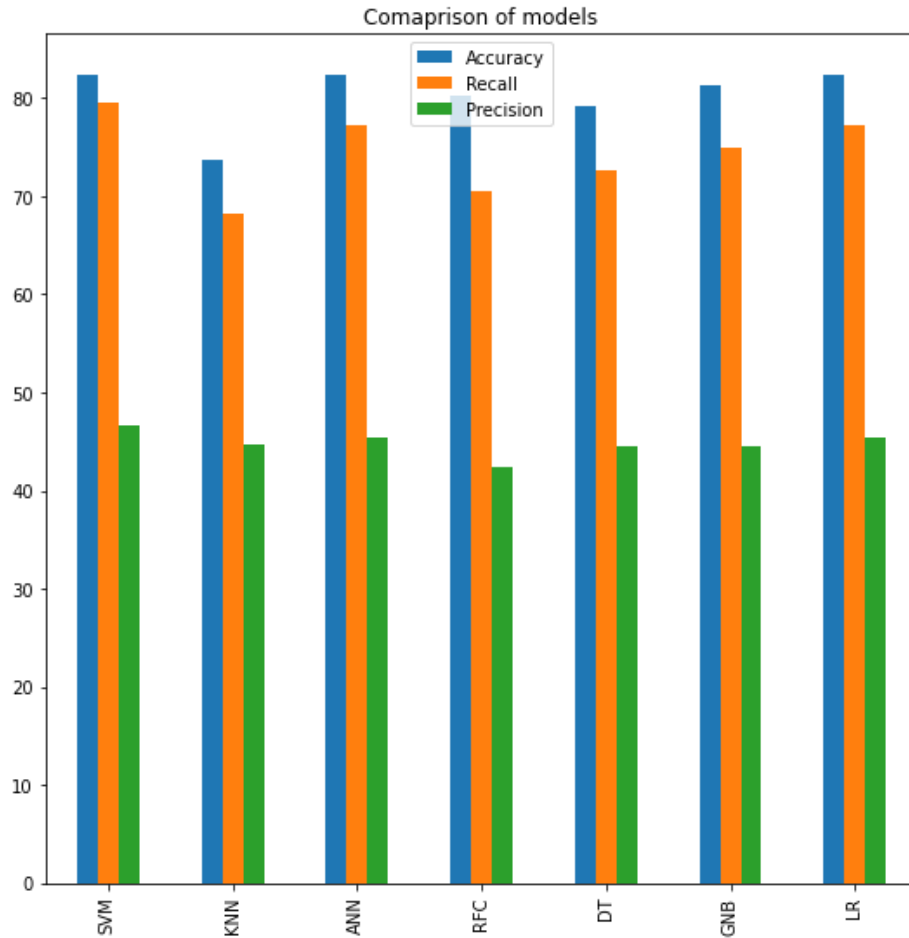


Fig:5.4: Comparison graph.

CHAPTER 6

SUMMARY AND CONCLUSIONS

CONCLUSION

To identify cardiac illness, we used machine learning techniques. It is possible to use Vector Machines, Classifiers like KNeighbors, Voting and Naive Bayes are just a few of the many options available. When sample protocols were used, accuracy and recall rates skyrocketed. Machine learning algorithms were evaluated to discover the best one. In diverse scenarios, each strategy has produced different results. Machine learning algorithms beat deep learning models in terms of accuracy, while Random Forest Classifiers, Decision Tree Classifiers, and Nave Bayes algorithms outperform other ML algorithms, according to the findings.

FUTURE SCOPE

It is our goal in the future to give a cost-effective and simple method for predicting heart disease early on to reduce the death rates in the world.

REFERENCES

Journals:

- [1] Fahd Saleh Alotaibi 2019. "Implementation of Machine Learning Model to Predict Heart Failure Disease", International Journal of Advanced Computer Science and Applications, 10(6).
- [2] Ahmed M. AlaalD1, Thomas Bolton, Emanuele Di Angelantonio, James H.F. RuddID, Mihaela van der Schaar, — "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participantsII", PLOS ONE 14(5): e0213653. <https://doi.org/10.1371/journal>, May 15, 2019H. Poor, — A Hypertext History of Multiuser Dimensions, MUD History, <http://www.ccs.neu.edu/home/pb/mudhistory.html>. 1986.
- [3] Stephen F. Weng, Jenna Reps, Joe Kai1, Jonathan M. Garibaldi, Nadeem Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?", PLOS ONE | <https://doi.org/10.1371/journal.pone.0174944> April 4, 2017
- [4] Rine Nakanishi, Damini Dey, Frederic Commandeur, Piotr Slomka, — "Machine Learning in Predicting Coronary Heart Disease and Cardiovascular Disease Events: Results from The Multi-Ethnic Study of Atherosclerosis (Mesa)II", JACC Mar- 20, 2018, Volume 71, Issue 11
- [5] Syed Nawaz Pasha et al 2020 "Cardiovascular disease prediction using deep learning techniques" To cite this article: IOP Conf. Ser.: Mater. Sci. Eng. 981 022006.
- [6] Shan Xu,Tiangang Zhu, Zhen Zang, Daoxian Wang, Junfeng Hu and Xiaohui Duan, 2017 "Cardiovascular Risk Prediction Method Based on CFS Subset Evaluation and Random Forest Classification" Framework Beijing China IEEE 2nd International Conference on Big Data Analysis,228-232.
- [7] SeyedaminPouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia and Juan Gutierrez 2017 "A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease", Heraklion, Greece, IEEE 2nd IEEE Symposium on Computers and Communication.
- [8] Swain, Debabrata, Santosh Pani, and Debabala Swain. 2019, April. "An efficient

system for the prediction of Coronary artery disease using dense neural network with hyper parameter tuning". International Journal of Innovative Technology and Exploring Engineering (IJITEE) 8 (6S). 689-695.

[9] Ghumbre, S.U. and A.A. Ghatol. 2012. "Heart disease diagnosis using machine learning algorithm." In Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012) Visakhapatnam India, Springer, 132,217-225.

[10] J Thomas MR, Lip GY. "Novel risk markers and risk assessments for cardiovascular disease". Circulation research. 2017;120(1):133–149. <https://doi.org/10.1161/CIRCRESAHA.116.309955> PMID: 28057790.

[11] Chen, H., S Y Huang, P S Hong, C H Cheng and E.J. Lin. 2011. "HDPS-Heart disease prediction system". Computing in Cardiology, Hangzhou, 557–560.

[12] Miranda E, Irwansyah E, Amelga AY, Kom S, Maribondang MM, Kom S, Salim M and Kom S.2016. "Detection of cardiovascular disease risk's level for adults using Naive Bayes classifier". Healthc Inf Res. 22(3) 196–205.

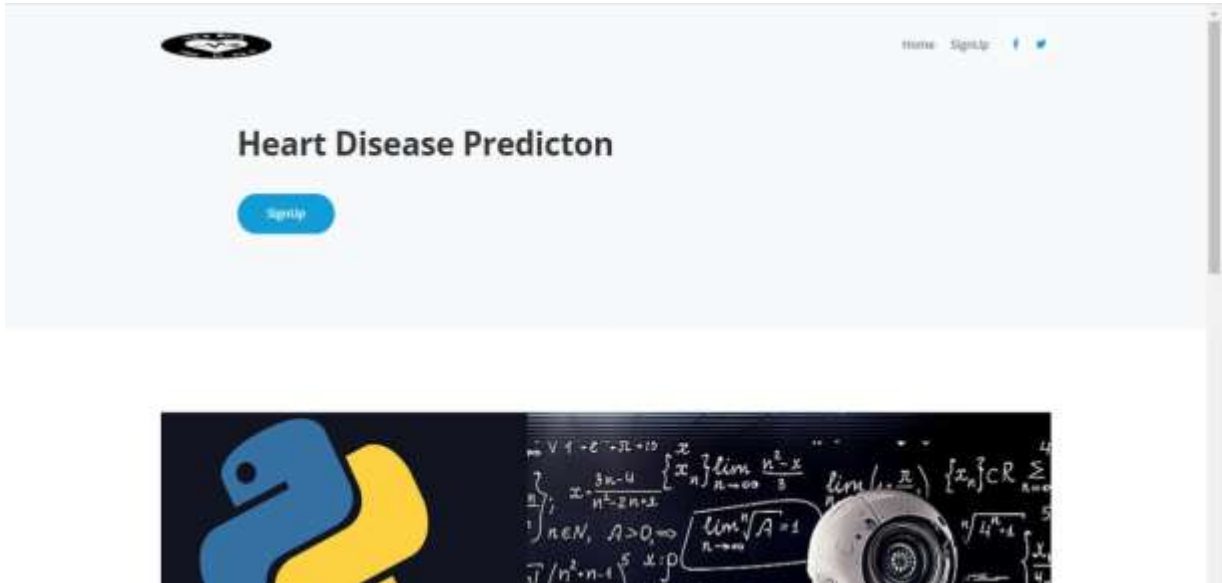
[13] Bharti, S. 2015. "Analytical study of heart disease prediction comparing with different algorithms". International Conference on Computing, Communication and Automation (ICCA2015). Noida,78-82.

[14] Mohammed Ali Shaik 2019 "A Survey on Text Classification methods through Machine Learning Methods". International Journal of Control and Automation 12(6) 390-396.

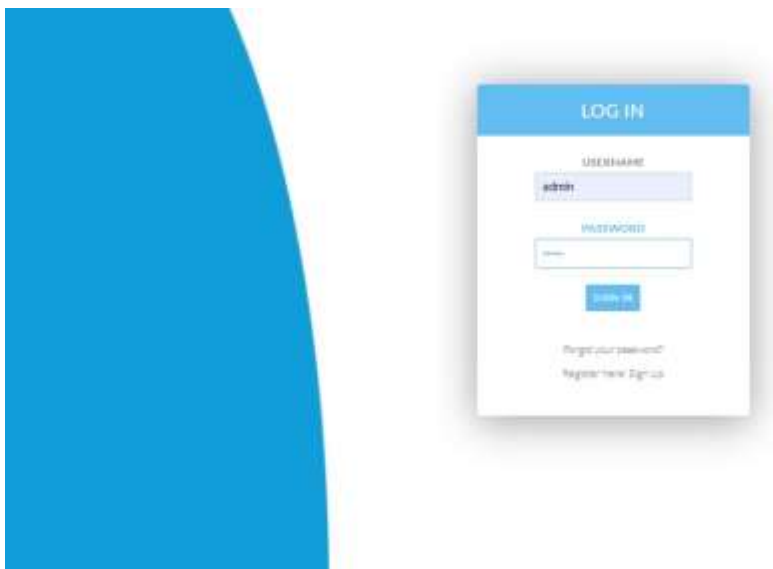
[15] Ravi Kumar R, Babu Reddy M and Praveen, 2019 "An evaluation of feature selection algorithms in machine learning" International journal of scientific & technology research 8(12) 2071-2074.

APPENDIX

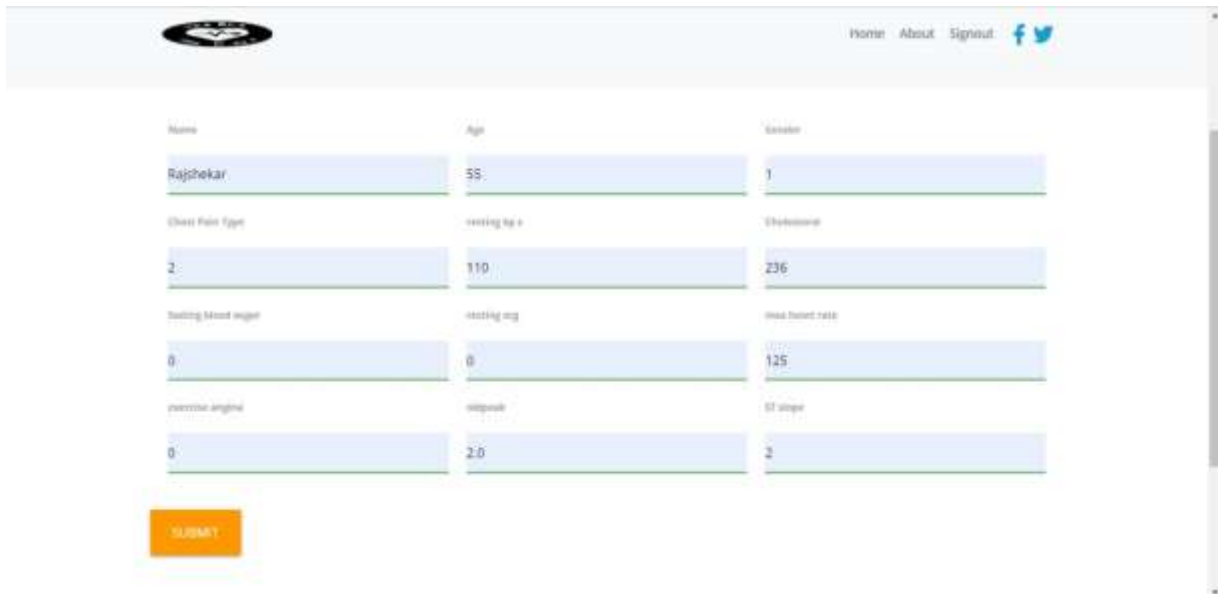
A) SCREENSHOTS



Home Page – Above screen is the home page of our local host, select SignUp to login or create a new account.



LOG IN – Enter your username and password to login as shown in the above screen.

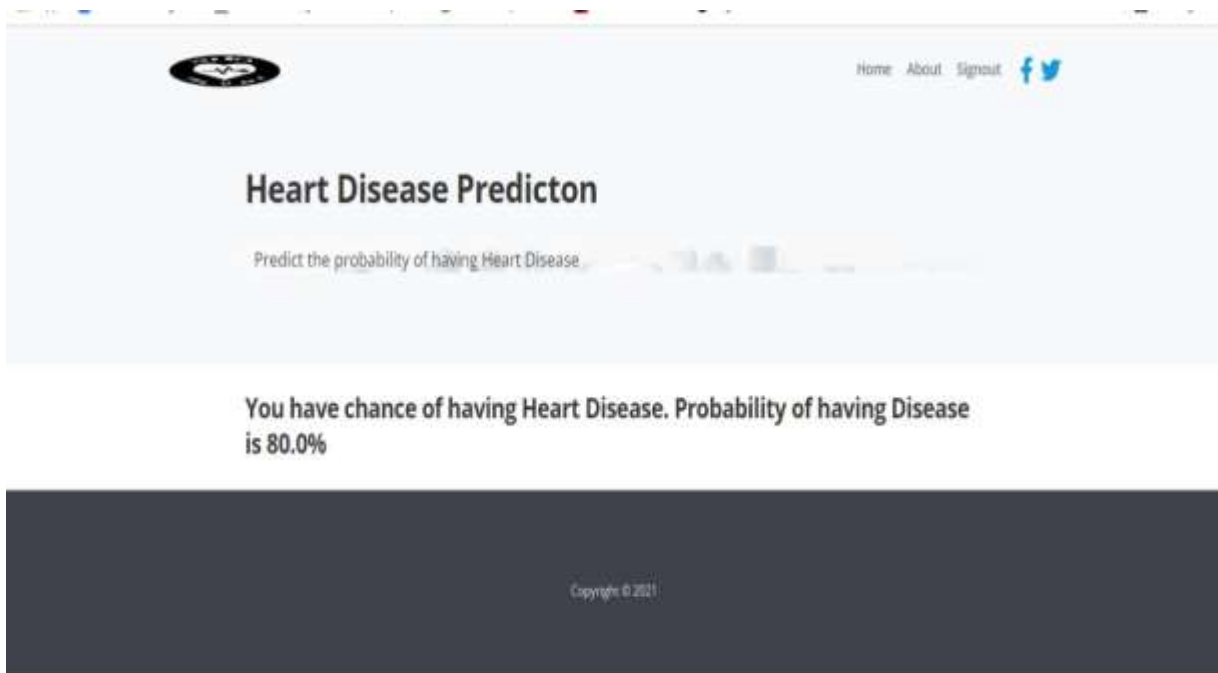


The screenshot shows a web form for heart disease prediction. At the top, there is a logo on the left and navigation links (Home, About, Signup) with social media icons (Facebook, Twitter) on the right. The form consists of nine input fields arranged in a 3x3 grid. Below the grid is an orange 'SUBMIT' button.

Name	Age	Sex
Rajshakar	55	1
Chest Pain Type	resting bp s	cholesterol
2	110	236
fasting blood sugar	resting eeg	max heart rate
0	0	125
exercise angina	oldpeak	st slope
0	2.0	3

SUBMIT

Enter your information for different attributes displayed on the screen after you logged in as shown in the above screen and submit



The screenshot shows the result of the heart disease prediction. The header includes the same logo and navigation links as the previous screen. Below the header, the title 'Heart Disease Prediction' is displayed, followed by the subtitle 'Predict the probability of having Heart Disease'. A bar chart shows a high probability of having the disease. The main result text states: 'You have chance of having Heart Disease. Probability of having Disease is 80.0%'. At the bottom, there is a dark grey footer with the text 'Copyright © 2021'.

Heart Disease Prediction

Predict the probability of having Heart Disease

You have chance of having Heart Disease. Probability of having Disease is 80.0%

Copyright © 2021

The above screen gives us the probability of having heart disease in terms of percentage.

B) SOURCE CODE

```
from flask import Flask,request, url_for, redirect, render_template
import joblib
import pandas as pd
import sqlite3

app = Flask(__name__)

model = joblib.load(open("model.sav", "rb"))

@app.route('/')
def hello_world():
    return render_template("home.html")

@app.route('/logon')
def logon():
    return render_template('signup.html')

@app.route('/login')
def login():
    return render_template('signin.html')

@app.route("/signup")
def signup():

    username = request.args.get('user',")
    name = request.args.get('name',")
    email = request.args.get('email',")
    number = request.args.get('mobile',")
    password = request.args.get('password',")
    con = sqlite3.connect('signup.db')
    cur = con.cursor()
    cur.execute("insert into `info` (`user`,`email`,`password`,`mobile`,`name`) VALUES
(?, ?, ?, ?, ?)",(username,email,password,number,name))
    con.commit()
    con.close()
    return render_template("signin.html")

@app.route("/signin")
def signin():

    mail1 = request.args.get('user',")
    password1 = request.args.get('password',")
```

```

con = sqlite3.connect('signup.db')
cur = con.cursor()
cur.execute("select `user`, `password` from info where `user` = ? AND `password`
= ?",(mail1,password1,))
data = cur.fetchone()

if data == None:
    return render_template("signin.html")

elif mail1 == 'admin' and password1 == 'admin':
    return render_template("index.html")

elif mail1 == str(data[0]) and password1 == str(data[1]):
    return render_template("index.html")
else:
    return render_template("signup.html")

@app.route('/predict',methods=['POST','GET'])
def predict():
    text1 = request.form['1']
    text2 = request.form['2']
    text3 = request.form['3']
    text4 = request.form['4']
    text5 = request.form['5']
    text6 = request.form['6']
    text7 = request.form['7']
    text8 = request.form['8']
    text9 = request.form['9']
    text10 = request.form['10']
    text11 = request.form['11']
    text12 = request.form['12']

    row_df =
pd.DataFrame([pd.Series([text2,text3,text4,text5,text6,text7,text8,text9,text10,text11,t
ext12])])
    print(row_df)
    prediction=model.predict_proba(row_df)
    output='{0:.{1}f}'.format(prediction[0][1], 2)
    output = str(float(output)*100)+'%'
    if output>str(0.5):
        return render_template('result.html',pred=f'You have chance of having Heart
Disease.\nProbability of having Disease is {output}')
    else:
        return render_template('result.html',pred=f'You are safe.\n Probability of having
Disease is {output}')

@app.route('/about')

```

```
def about():  
    return render_template('about.html')
```

```
@app.route('/index')  
def index():  
    return render_template('index.html')
```

```
if __name__ == '__main__':  
    app.run(debug=True)
```