HEART DISEASE PREDICTION USING HYBRID MODEL

A PROJECT REPORT

Submitted by

ANAND KUMAR SONI [Reg No: RA1711003010481]

ROHIT SHRIVASTAVA [Reg No: RA1711003010511]

Under the Guidance of

DR. P. VIGNESHWARAN

(Associate Professor, Department of Computer Science and Engineering)

In Partial Fulfillment of the Requirements

for the Degree of

BACHELOR OF TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR – 603 203

MAY 2021

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR – 603 203

BONAFIDE CERTIFICATE

Certified that this B.Tech project titled “**HEART DISEASE PREDICTION USING HYBRID MODEL**” is the bonafide work of **Mr. Anand Kumar Soni and Mr. Rohit Shrivastava** who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion for this or any other candidate.

<<Signature of the Supervisor>>

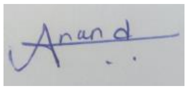
Dr. P. Vigneshwaran
SUPERVISOR
Associate Professor
Department of Computer science and Engineering

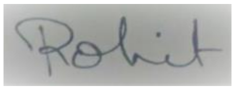
<<Signature>>

Dr. B. AMUTHA
HEAD OF DEPARTMENT
Department of Computer science and Engineering

Signature of the Internal Examiner

Signature of the External Examiner

Sign of students **:** 



OWN WORK DECLARATION

Department of Computer Science and Engineering

SRM Institute of Science & Technology

Own Work* Declaration Form

This sheet must be filled in (each box ticked to show that the condition has been met). It must be signed and datedalong with your student registration number and included with all assignments you submit – work will not be marked unless this is done.

<u>To be completed by the student for all assessments</u>

Degree/ Course : Bachelor of Technology in Computer Science and Engineering

Student Name : Anand Kumar Soni, Rohit Shrivastava

Student Sign: Anand Rohit

Registration Number : RA1711003010481, RA1711003010511

Title of Work : HEART DISEASE PREDICTION USING HYBRID MODEL

I / We hereby certify that this assessment compiles with the University's Rules and Regulations relating to Academic misconduct and plagiarism**, as listed in the University Website, Regulations, and the Education Committee guidelines.

I / We confirm that all the work contained in this assessment is my / our own except where indicated, and that I / Wehave met the following conditions:

- Clearly references / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook / University website

I understand that any false claim for this work will be penalised in accordance with the University policies andregulations.

DECLARATION:
I am aware of and understand the University's policy on Academic misconduct and plagiarism and I certify that this assessment is my / our own work, except where indicated by referring, and that I have followed the good academic practices noted above.
If you are working in a group, please write your registration numbers and sign with the date for every student in your group.

ACKNOWLEDGEMENT

We express our humble gratitude to **C. Muthamizhchelvan**, Vice Chancellor (I/C), SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

We wish to thank **Dr. B. Amutha**, Professor & Head, Department of Computer Science and Engineering, SRM Institute of Science and Technology, for her valuable suggestions and encouragement throughout the period of the project work.

We are extremely grateful to our Academic Advisor **Dr. P. Vishalakshi,** Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, for their great support at all the stages of project work. We would like to convey our thanks to our Panel Head, **Dr. Annapoorani K,** Associate Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, for her inputs during the project reviews.

We register our immeasurable thanks to our Faculty Advisor **Dr.S.Suchitra,** Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to my guide, **Dr. P. Vigneshwaran,** Associate Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, for providing me an opportunity to pursue my project under his mentorship. He provided me the freedom and support to explore the research topics of my interest. His passion for solving the real problems and making a difference in the world has always been inspiring.

We sincerely thank staff and students of the Computer Science and Engineering Department, SRM Institute of Science and Technology, for their help during my research. Finally, we would like to thank my parents, our family members and our friends for their unconditional love, constant support and encouragement.

ANAND KUMAR SONI

ROHIT SHRIVASTAVA

ABSTRACT

Coronary illness is one of the main sources of death on the planet today. Cardiovascular infection expectation is a major test in the field of clinical information investigation. AI (ML) has been found to help in the dynamic and forecast of enormous amounts of data produced by the wellbeing area. In recent developments in various medical industries, machine learning techniques are employed. In this work, we proposed a new method for finding cardiovascular diseases using machine learning techniques. The model prediction uses classification technologies and a dataset is used for Cleveland cardiac diseases. The Decision Tree and Random Forest are applied to machine learning techniques. The new machine learning technology is being applied. The process is implemented by three machine learning algorithms.: 1. decision trees, 2. random forests and 3. (Hybrid of Decision tree and random forest). The prediction model of cardiovascular conditions with the hybrid model shows an accuracy level of 88, 7 percent. The interface is intended to get the client's boundary for anticipating coronary illness, for which we have utilized the Random Forest and the Hybrid Decision Tree model.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ML - Machine learning

CVD - Cardiovascular diseases

CHDD - Cleveland Heart Disease Database

KNN - K-Nearest Neighbor Algorithm

DT - Decision Trees

GA - Genetic algorithm

NB - Naive Bayes

CADSS - Computer Aided Decision Support System

CNN - Convolutional Neural Networks

ECG - Electrocardiogram

DL - Deep learning

CHAPTER 1

INTRODUCTION

1.1 GENERAL

The Data Mining (DM) consists of extracting useful information from large data sets that leads to the prediction or depiction of data by using techniques such as classification, clustering, association etc. Data mining in the health care industry has found extensive applicabilities for the classification of optimal treatment methods, for prediction of risk factors for diseases and for efficient patient care cost structures. Diseases like diabetes, asthma, cardiovascular diseases, AIDS etc. have undergone research using Data Mining Models. Different data mining methods were used to develop health research models such as the naïve classification of Bayesia in general, artificial neural networks, vector support machines, policy areas, regressive logistics and so on.

Every year, approximately 17 million people die as a result of cardiovascular disease. Although these diseases are treatable, their high mortality rates necessitate an early prognosis and assessment of a patient's risk. Common cardiovascular conditions include cardiopathy, cardiovascular disease, heart failure, high blood pressure hearing disease, etc. Tabacco, diabetes, lack of physical activity, higher blood pressure, dietary cholesterol is the common cause.

Cardiovascular infections research utilizing information extraction has been a continuous exertion including high-accuracy expectation, treatment and hazard score examination. The most prominent data set from the Cleveland Heart Clinic has been conducted on multiple CVD surveys. The Data Base on Cleveland Heart Disease (CHDD) was, as such, de facto a database for research into cardiovascular diseases. This paper gives a structure to applying calculated relapse and supporting vector machines to get singular expectations that are thusly utilized in calculations that utilization rules, to suggest these boundaries from this information base. The results of each rule will be compared based on precision, sensitivity and specificity.

Two destinations are achieved: first, a prescient coronary disease structure, and second, the approach, a comparison of fusion efficiency between the results of several models and the use of a single model.

It's difficult to identify cardiac disease due to a variety of factors like diabetes, Blood pressure, cholesterol high, abnormal rate of pulsation, etc. Various techniques for detecting the seriousness of heart disease among men are used in data mining and neural networks. The severity of this condition is categorised using a range, including Naive Bayes Nearest neighbour algorithms, KNN, DT, genetic algorithms (GA) (NB). The nature of cardiac disease is complex and should be carefully treated. This can neither influence the heart nor lead to early death. Various types of syndromes are recognised with a medical and data mining perspective. In cardiovascular prediction and data analysis, data collection plays a major role.

We also saw decision-making trees used to predict cardiac accuracy. The known data extraction methods for predicting cardiovascular disease used different methods of abstraction of knowledge. Many readings were given in this work to produce a model prediction, using two or more techniques as well as separate methods. These new methods are often referred to as hybrid techniques. Preparation of neural networks with a series of the heart rate. This method is used through various medical reports in order to predict exact conditions of the patient related to cardiovascular diseases, such as Left Bundle Branch Block (BBB), Right Bundle Branch Block (RBBB), Atrial Brillation (AFIB) (BII). The data set is used in the radial function network (RBFN) classification, which uses 70% of the data for training and 30% for classification.

We are also introducing the Medicine and Research Computer Aided Decision Support System (CADSS). Data mining techniques have been shown to take less time to forecast the disease in the healthcare sector in previous work[16]. We recommend that cardiovascular disease be diagnosed with the GA. This method uses efficient GA selection association regulations, intersection and mutation of tournament that lead to the new fitness proposed. We use the famous Cleveland data set collected from the study repository for experimental validation. Later we will see how important the results of some of the familiar supervised learning techniques are compared to. PSO is introduced with the most powerful evolutionary algorithm and some heart disease rules are created. The rules were randomly applied with encryption methods to improve overall accuracy. Heart disease based on symptoms such as pulsation, gender, age, etc. is predicted. The ML algorithm, which results are more exact and reliable as Neural Networks, is presented

We saw it. In most cases, neural networks are the best way to anticipate infections, such as heart conditions and cerebrum illnesses. There are 13 cardiovascular properties in the technique we use. In comparison to existing methods in such works, the results indicate an increased level of performance. During these recent years the carotid artery stenting (CAS) has also become a common medical mode of treatment. In patients with older cardiomyopathic conditions, CAS causes major cardiovascular adverse events (MACE). Your evaluation is of great importance. We produce results with an artificial neural ANN network which delivers excellent results in prediction of cardiac diseases. Introducing neural network methods which combine the later probabilities as well as predicted multi-technique values. The accuracy levels of this model are up to 89:01 percent compared to previous works. The Cleveland cardiac set for all experiments is employed by a NN Neural Network to improve cardiac disease performance, as we saw previously. Recent developments in ML technology for Internet material have also been made (IoT). The accurate IoT connected devices have been shown in the Network Traffic Data ML algorithms. Meidan et al. have separately gathered and labelled data on nine IoT, PC and smartphone systems. Through supervised study a multi-phase meta-classification is trained. In the initial phase, the classifier can vary between IoT and non-IoT devices generated by the traffic. Each IoT device has a particular type of IoT device in the second phase. Deep knowledge is a promising approach in complex environments to collect accurate data from IoT devices. Deep learning also is suitable for the edge computer environment thanks to its multi-layer structure.

We present a technique called the Random Hybrid Forest, a linear model in this work (HRFLM). The principle point of this examination is to improve the exactness of cardiovascular illness forecast. Numerous studies have resulted in limits in algorithm selection. The HRFLM method does not use any feature selection limitations, however. Here we experiment with the characteristics of a hybrid study method. The results of our experiment show that our proposed hybrid technique is more likely in comparison to existing methods to predict heart disease.

1.2 PURPOSE

The purpose of the survey is to predict cardiac disorders through an automated, machine-based diagnostic system. We use the hybrid model, which is the best heart disease prediction classification algorithm.

1.2.1 SCOPE

Heart conditions are one of the most common diseases that can reduce the lives of people today. For a variety of medical and health promotion areas an estimated risk for coronary heart disease is important. A longitudinal survey multivariate regression analysis can provide a risk prediction model. Techniques of data mining and algorithms for machine learning play a vital role in the analysis of the various data in medical centres. The techniques and algorithms can be used on a data set directly for the production of certain models or for drawing important conclusions and conclusions.

1.2.1.1 LIMITATIONS

- Due to digital technologies, large amounts of data are being stored in health care centres which are very complex and difficult to analyse in their database.

- Medical diagnosis is considered a major but complex task to be performed accurately and efficiently. Clinical decisions can often be taken based not on data hidden in the database but on the intuition and experience of doctors. This practise results in undesirable prejudices, errors and excessive healthcare costs that have an effect on the quality of the patient's service.

CHAPTER 2

LITERATURE SURVEY

A review of the study is provided in this chapter. This section illustrates some of the researchers' relevant work.

The researchers on the prediction problem of heart disease studied many existing techniques, few of which are discussed below.

In the fields directly associated with this paper there is ample connected work. The highest exact prediction in the medical field was introduced to produce ANN.. ANN's multi-layer background propagation perception (MLP) foreshadows cardiac disease. The results obtained are compared to existing models of the same domain and improved on them. The UCI laboratory data are used to find patterns using Vector Support Machines NN, DT, SVM and Naive Bayes. For their performance and precision the results are compared with these algorithms. For the F-measure, the proposed hybrid approach results in competition with other existing methods of 86:8 per cent. Classifying Convolutionary Neural Networks (CNN) without segmentation is introduced. This method takes into account the heart cycles in which the electrocardiogram (ECG) signals take different start positions in training. During the patient test stage, CNN can generate features with different positions.

A great deal of information has not been effectively used by the medical industry before now. The new approaches here are an easy and effective way to reduce costs and to improve heart disease prediction. The various research techniques for prediction and classification of cardiovascular diseases by means of ML and DL are very accurate to determine their effectiveness.

2.1 [Intelligent heart disease prediction system using random forest and evolutionary approach]

Akhil, Jabbar & Deekshatulu, Bulusu & Chandra, Priti. (2016). Intelligent heart disease prediction system using random forest and evolutionary approach. journal of network and innovative computing. 4. 175-184.

Heart is one of the main sources of early passing around the world. The outcome of the disease is difficult to predict. The extraction of data is used to deduce diagnostic rules automatically and help specialists improve diagnostic reliability. Researchers use various data extraction techniques in the prediction of cardiovascular diseases to help healthcare professionals. One of the most exact and medically suitable learning algorithms is the random forest. The selection chi square measurement is used to assess and determine whether or not the variables are correlated. This document proposes to use a random forest classification model to classify the cardiovascular disorders as a square and genetic algorithm. In comparison to other methods, our approach improves classification accuracy, and the presented model can be used successfully for predicting cardiac diseases .The experimental results show that

2.2 [A Data mining Model for predicting the Coronary Heart Disease using Random Forest Classifier]

A. S. Abdullah and R. R. Rajalaxmi, in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 2225.

Coronary heart disease (CHD) is an important cause and a common form of cardiac disease. Data mining is involved in the discovery from the medical sciences of various types of metabolic syndromes. Data mining classification techniques are an important part of data prediction and analysis. The prediction of accuracy and CHD events was based on classification techniques such as Decision Trees. This paper developed a data mining model to improve predictive accuracy and investigate different CHD-related events by means of the Random Forest classification. This model can help doctors predict how CHD can be linked to various segments of the population and their various events. The events investigated include the AMI, Coronary Artery Bypass Surgery, and Angina, and Percutaneous Coronary Intervention (CABG). Experimental results have shown that the classification of CHD events and risk factors can be successfully used with the algorithm of the random forest classification.

2.3 [Using PSO algorithm for producing best rules in diagnosis of heart disease]

A. H. Alkeshuosh, M. Z. Moghadam, I. A. Mansoori and M. Abdar, "Using PSO Algorithm for Producing Best Rules in Diagnosis of Heart Disease," 2017 International Conference on Computer and Applications (ICCA), Doha, 2017, pp. 306-311.

The global problem of heart disease continues to grow. In the health system, a limitation of the Expertise and human experience in the diagnosis of the manufacture of medical devices leads to improper diagnoses and inadequate or inaccurate information about different conditions. Since it is important to predict a person's condition correctly, medical science is equipped with intelligent tools to diagnose and treat diseases. It can reduce physician failures and economic losses. In this document, the algorithm Particle Swarm Optimization (PSO) is used for the production of rules for cardiac diseases, one of the most powerful algorithms in evolution. The random regulations are first encoded and then optimised by using PSO algorithms based on their accuracy. Finally, the C4.5 algorithm is compared to our results.

2.4 [Backpropogation neural network for prediction of heart disease]

Al-Milli, Nabeel. (2013). Backpropogation neural network for prediction of heart disease. 56. 131-135.

Recently, researchers have proposed multiple software, tools and diverse algorithms for developing effective support systems for medical decision making. In addition, new algorithms and tools are still being developed and represented day by day. One major issue in the diagnostic of heart disease is the development of smart systems for medical decision management to enhance the ability of physicians in many researchers. The neural network is widely used for the diagnosis of cardiomyopathy. A prediction system for cardiac diseases is developed through the neural network in this research paper. 13 medical attributes for predicting heart disease have been used in the system proposed. The experiments in this work have shown that the algorithm is good compared to similar state-of-the-art approaches.

2.5 [Effective Heart Disease Prediction using Distinct Machine Learning Techniques]

N. Suganthi, R. Abinavi, S. Deva Dharshini,V. Haritha (Mar. 2020). International Research Journal of Engineering and Technology.pp.3383

enhanced level of performance through a heart condition prediction model using a linear model in the hybrid random forest; (HRFLM).

2.6 [A Prediction Technique for Heart Disease Based on Long Short Term Memory Recurrent Neural Network]

Manohar Manur,Alok Kumar Pani and Pankaj Kumar. INASS Nov. 2019,pp-31-39.

The proposal includes the proposed two-directional long term random field long memory (BiLSTM-CRF), bidirectional analytical analyses of input health data for efficient analysis and a linear relationship between the features provided by the CRF.

2.7 [Knowledge Discovery in Databases: An Overview. Published by the AAAI Press/ The MIT Press, Menlo Park, C.A 1996.]

Frawley and G. Piatetsky -Shapiro, An Overview. Published by the AAAI Press/ The MIT Press, Menlo Park, C.A 1996.

An intelligent prediction system for heart disease is developed using data extraction techniques. The proposal was made Naive Bays, Network and Trees of Decision Each method has its own strength to achieve correct results. Here are used patterns and connections. It is web-based, user-friendly and extensible.

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

An automatic questionnaire is one of the latest study using neuronal network information to develop a prediction system for heart condition (SAQ). Test results for the "Dundee rank factor score" neural system together with sex and age have been used in order to validate the task of assessing the risk of heart disease. This is based on 3 factors of risk (blood pressure, smoking and blood cholesterol). During the study, a multi-faceted nerve feeding network trained to baking propagation in the algorithm was used.

Drawbacks

The study not only clarifies common risk factors of the disease but also the other data collected in SAQ.

In order to diagnose heart disease the k-NNN has been given higher accuracy by Mai Shouman et al. in application of k- Nearest Neighbors (k-NN).

Drawbacks

Applying integrating voting could not enhance the k-NN accuracy in the diagnosis of heart disease patients, unlike Decision tree classifiers where voting increases accuracy. Voting is an aggregation technique which is used to combine decisions of multiple classifiers. However the accuracy for k-NN with voting reduced to 92.7%.

G Purusothaman et al. examined the different prediction techniques of heart disease and compared them. Instead of a single type like Decision-tree, The Naïve Bayes, Artificial Neural and other hybrid models combine more than a grading technique.

Drawbacks

However, applying hybrid model has difficulties.

3.2 PROPOSED SYSTEM

To meet the need of drawbacks of existing systems and after getting through various surveys we got that Genetic algorithm is not suited for all problems and is expansive since fitness is calculated repeatedly, PSO is very complex in high dimensional space, In Neural Network there is unexplained functionality of network, Naïve bayes assumes all predictors are independent which rarely happens in reality especially in field of Medical, The proposed system is to predict cardiomyopathy through an automatic diagnosis system based on machine learning. Hybrid model used for the prediction system. Hybrid of Decision tree and Random forest we can get advantages of both the algorithms like handeling of both numerical , categorical data and large dataset with higher dimensionality, less effort for data preparation ad handle missing value and maintains accuracy. Hybrid model also gives less errors of R-squared etc which is very essential in sensitive prediction of diseases like heart disease.

The database was used for the predictive system of cardiovascular disease. Because the most common database used by ML researchers is Cleveland.

The data collection contains 303 boxes and 76 characters, but all published studies contain only 14. The target field with variations of values between 0(absence) and 4 indicates whether or not a patient has heart disease. The Cleveland database studies focus on the distinction between absence (value 0) and presence (values range from 1 to 4).

Advantages

The suggested prediction system for heart disease uses a hybrid model.
Patients can be treated on the basis of the generated report
Helps patient to take preventive measures in advance.

3.3 PROBLEM STATEMENT

Heart disease is one of the common conditions which can reduce the lives of people.
Cardiovascular function is affected by the heart condition.

For several aspects of medicine and promotion of health, a risk estimate of coronary heart disease is important.
The risk prediction model may be provided by multivariate regression analysis of a longitudinal survey.
Due to rapidly expanding digital technologies, health centres store a huge number of data in a hig hly complex and difficult to analyse database.
Techniques of data mining and algorithms for machine learning play a vital role in the analysis o f the various data in medical centres.
The techniques and algorithms can be directly used in a production data set of certain models or for drawing important conclusions and conclusions.

3.4 REQUIREMENT ANALYSIS

3.4.1 FUNCTIONAL REQUIREMENTS

Data collection

In order to collect data, quality data has been selected for analysis. This is where we used the data set for cardiovascular disease taken from uci.edu. The task of a data analyst is to identify ways and sources of data collection, interpretation and analysis by means of statistical methods.

Data visualization

A wide range of graphic information can be understood and analysed easier. Some companies specify that a slides, diagrams, charts and templates must be made available by a data analyst. The heart disease rates are shown as part of our approach.

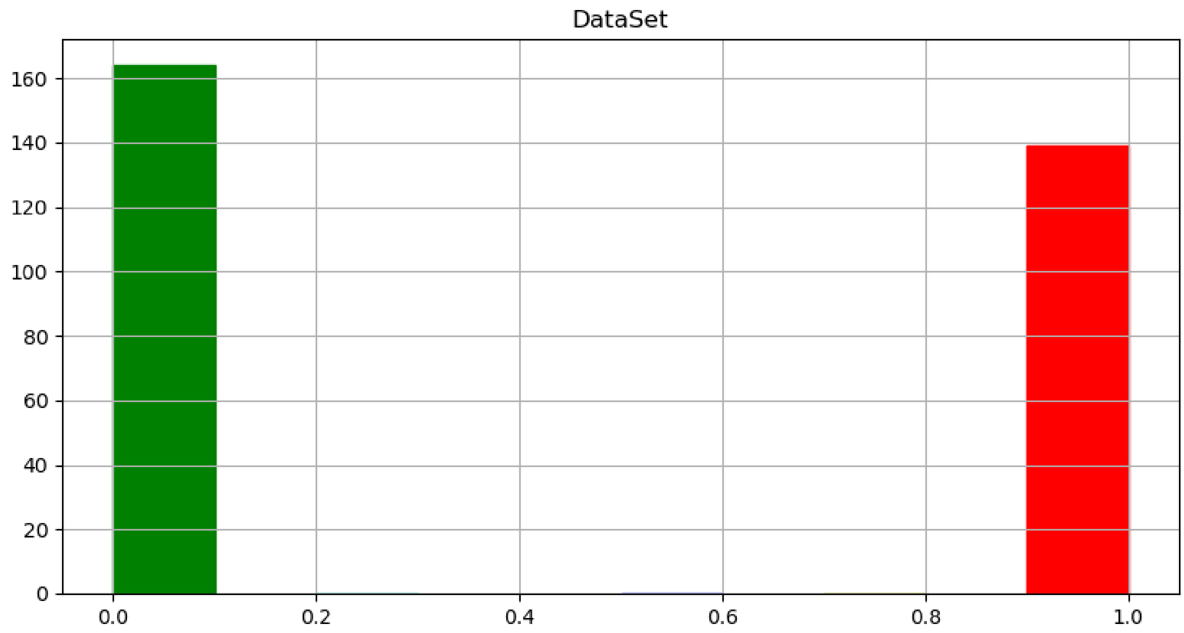


Figure 1: Data Visualization of Heart disease rate

Data preprocessing

Preprocessing is intended to convert raw data into a machine learning form. A data scientist can obtain more exact results through structured and clean data from an applied learning model. Formatization, purification and sampling of data are provided.

Dataset splitting

A machine learning dataset should be divided into 3 sub-sets — training, testing, and validation systems.

Training set. In order to develop and determine its optimal parameters, a data scientific person uses a training package. Test set. For the assessment of the trained model and its generalisation capacity a test set is necessary. The second is the capacity of a model, after training with a training data, to identify patterns of new unseen data. In order to avoid the overpackaging of the model that is unable to generalise it is important that various subsets are used for training and testing.

Model training

A model training can then be performed after a data scientist prepares and splits the collected data into a train. It involves "feeding" the training data to the algorithm. An algorithm processes data and outputs a model that finds a target value (attribute) for your predictive analysis in a new data. The goal of model education is to develop a model.

Model evaluation and testing

This step is aimed at developing the simplest model to formulate an objective value quickly and sufficiently. This can be achieved by a data scientist by tuning the model. The model parameters are optimised in order to achieve the best performance of an algorithm.

Non-functional requirements

A list of non-functional requirements is provided below. Internal stakeholders must define the specific details.

- Response Time

- Availability
- Stability
- Maintainability
- Usability

3.4.2 SYSTEM REQUIREMENTS

Hardware and software as described below include system requirement

3.4.3 HARDWARE REQUIREMENTS

Processor	: Any Processor above 500 MHz.
Ram	: 4 GB
Hard Disk	: 4 GB
Input device	: Standard Keyboard and Mouse.
Output device	: VGA and High Resolution Monitor.

3.4.4 SOFTWARE SPECIFICATION

Operating System	: Windows 7 or higher
Programming	: Python 3.6 and related libraries

CHAPTER 4

SYSTEM DESIGN

SYSTEM DESIGN

This chapter covers UML applications in architecture, implementation datasets, algorithms, and designs.

4.1 ARCHITECTURE DIAGRAM

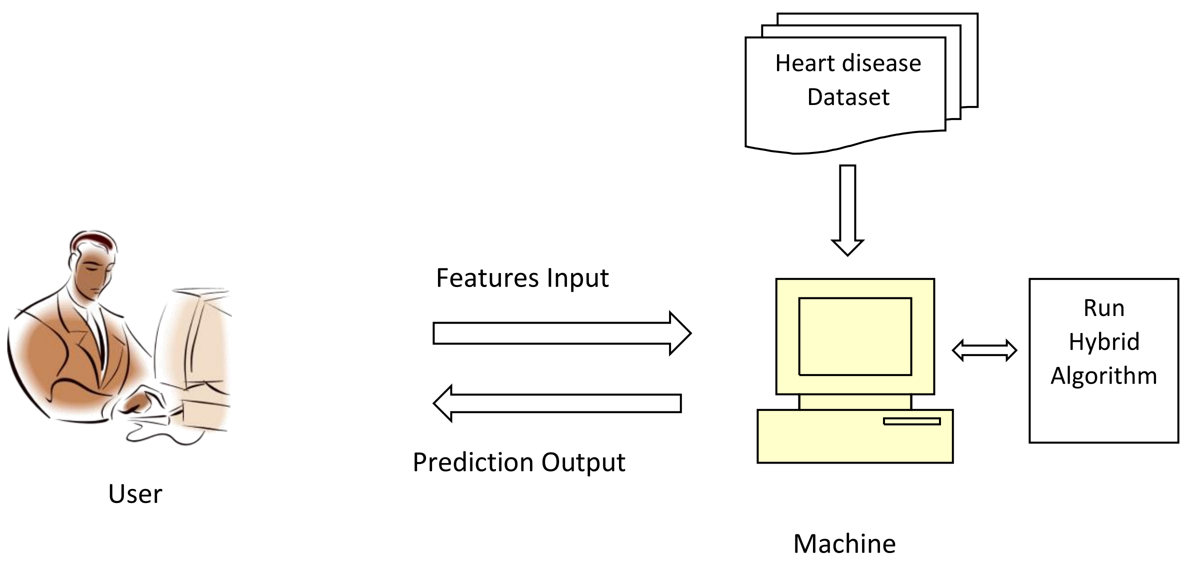


Figure 2: Architecture Diagram

The figure above represents the system architecture in which every work module is represented. User provides data collection input, model training and recognition is mentioned.

FLOW DIAGRAM

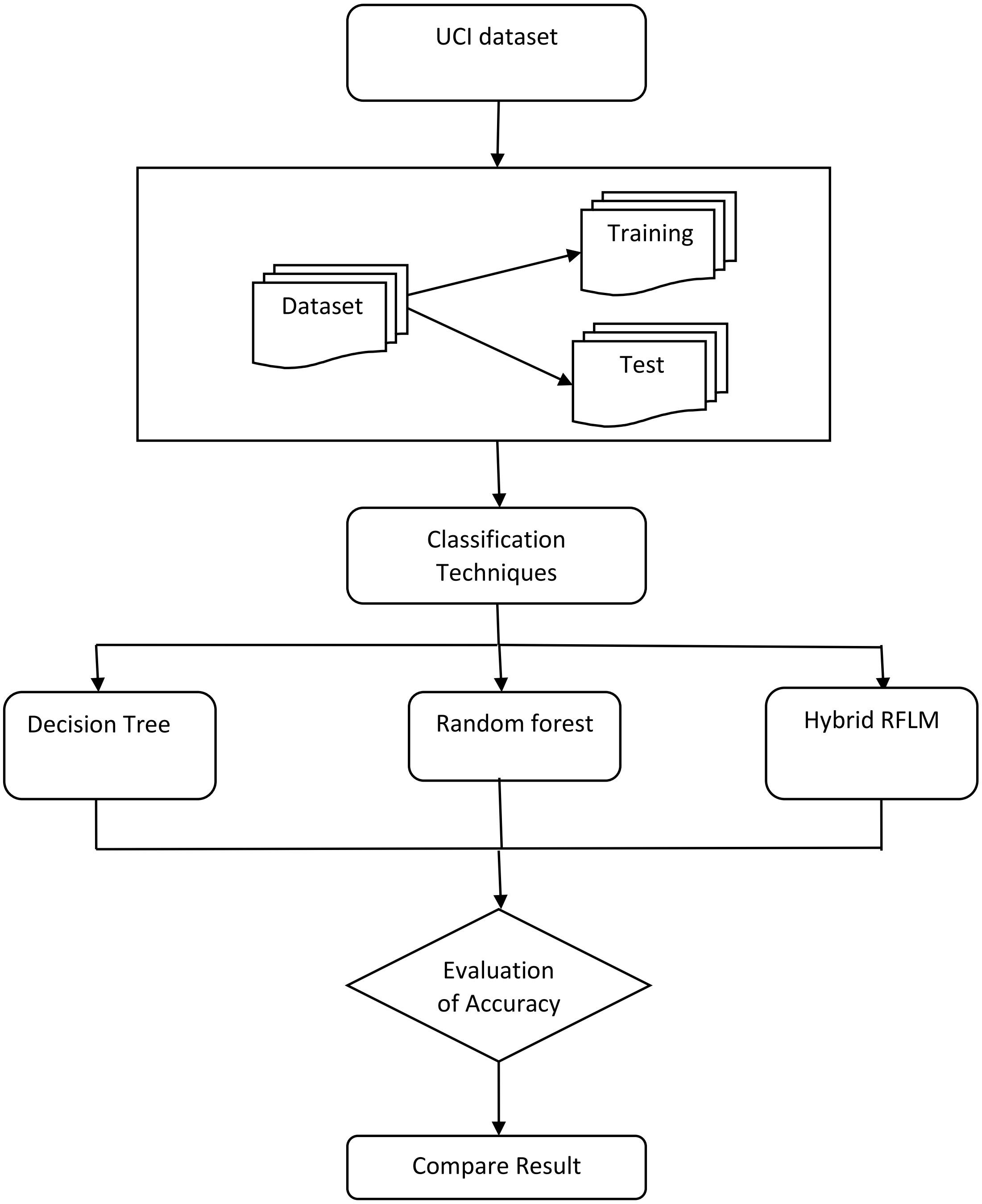


Figure 3: Flow Diagram

The figure above represents the system architecture in which every work module is represented. User will provide the collection of input data, training model and prediction.

4.2 UML DIAGRAMS

The design is a schematic or drawing which shows the appearance and function before the creation of an object. UML is a standard modelling language that enables developers to specify, view, construct and document software artefacts. UML is thus robust, safe and scalable. This is the result. UML is an important element in the development of object-oriented software. Visual models of Software Systems are created using graphical notation.

The following are the various kinds of UML graphs:

- Use Case Diagram
- Class Diagram
- Activity Diagram
- Sequence Diagram
- Collaboration Diagram
- Component Diagram
- Deployment Diagram

4.2.1 USE CASE DIAGRAM

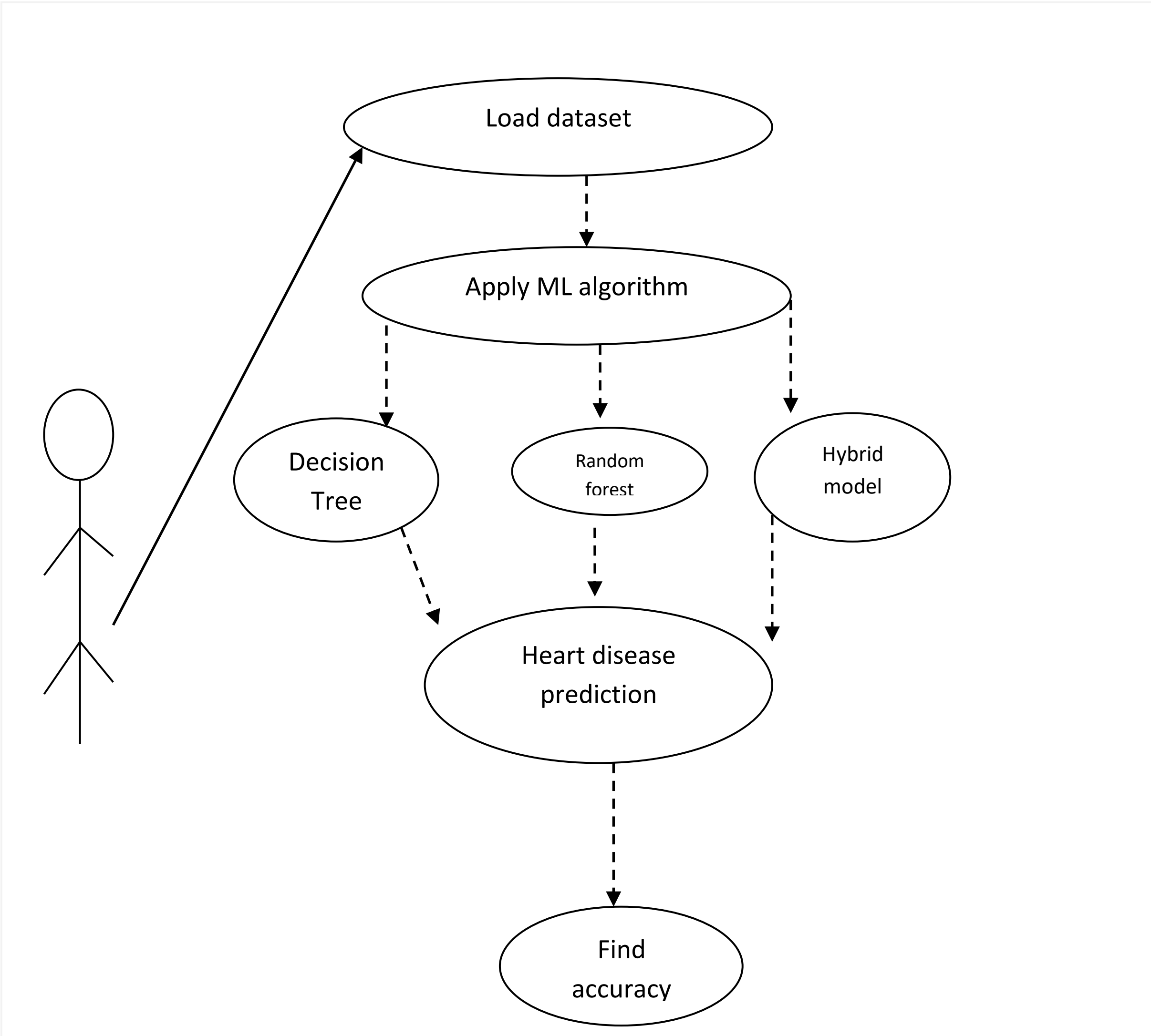


Figure 4: Use case Diagram

The above figure represents usecase diagram, in which user upload dataset is pre-processed and applied algorithm. They captures functionality, here humans/devices are actor and use cases are given in oval shape figures from loading dataset to apply ML algorithms and finding accuracy after heart disease prediction.

4.2.2 SEQUENCE DIAGRAM

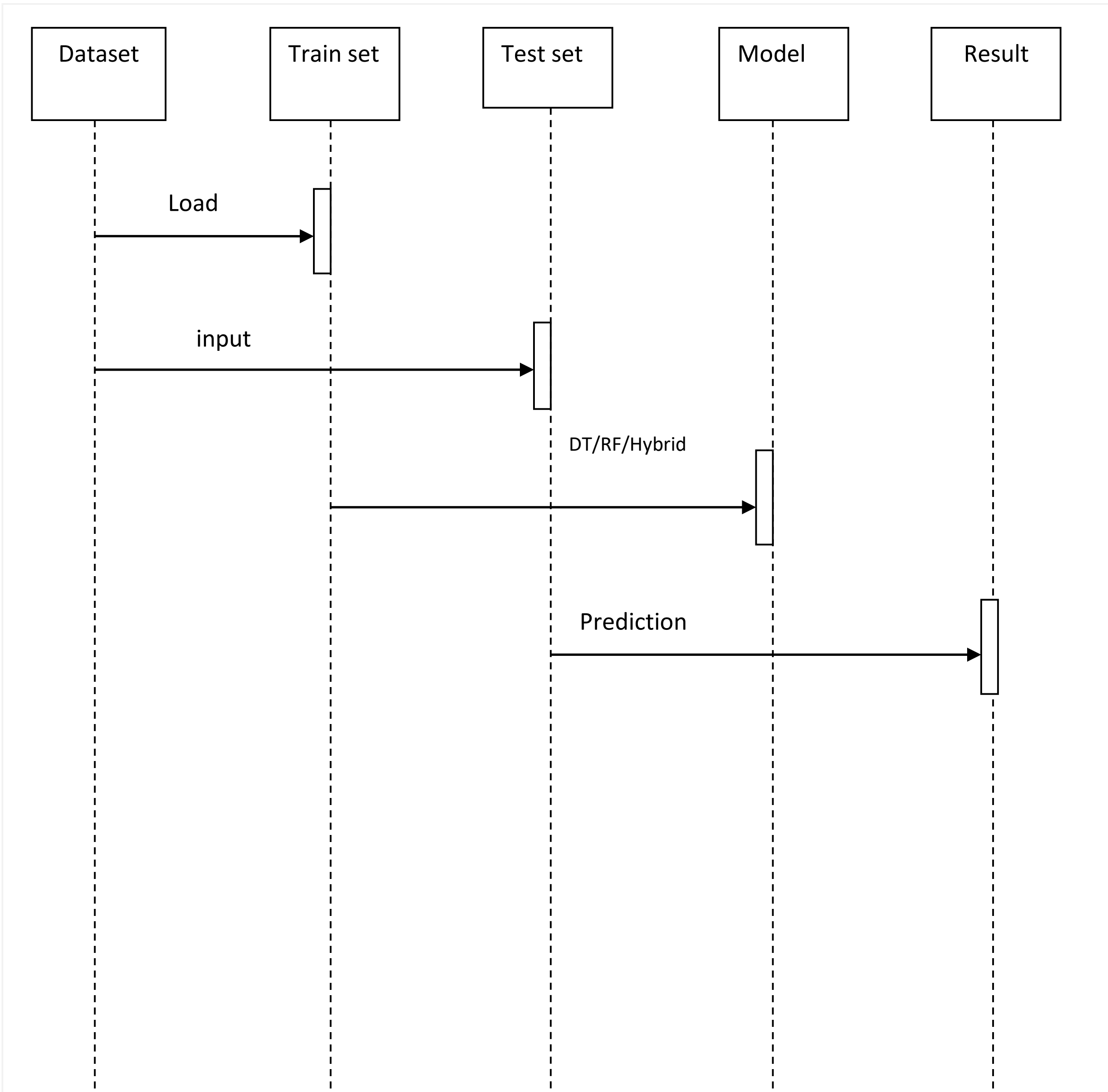


Figure 5: Sequence Diagram

A sequence diagram depicts vertical parallel lines, active processes and objects, and horizontal arrows indicating message exchange. The above figure is a sequence chart and represents the sequence of the flow of data in the proposed system. Here, input from dataset is taken to train and test set than Decision Tree, Random Forest and Hybrid algorithms are applied and model is generated, result is predicted using test set.

4.2.3 ACTIVITY DIAGRAM

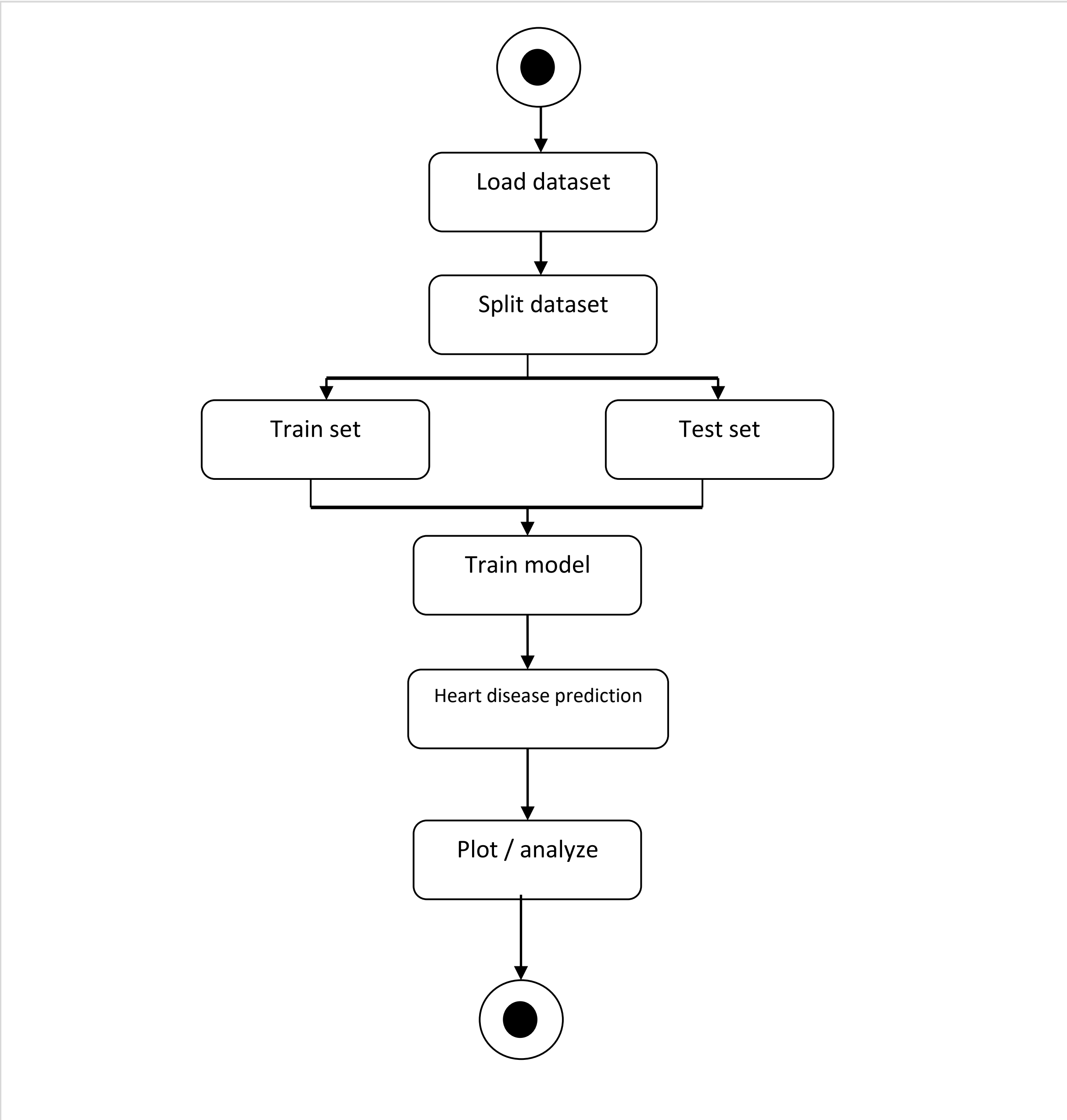


Figure 6: Activity diagram

The above figure represent activity diagram of proposed system. The figure shows complete flow of activity from dataset loading and all sequence of module. It captures dynamic aspects of Heart Disease prediction system as shown connected with fork and joins as shown.

4.2.4 DEPLOYMENT DIAGRAM

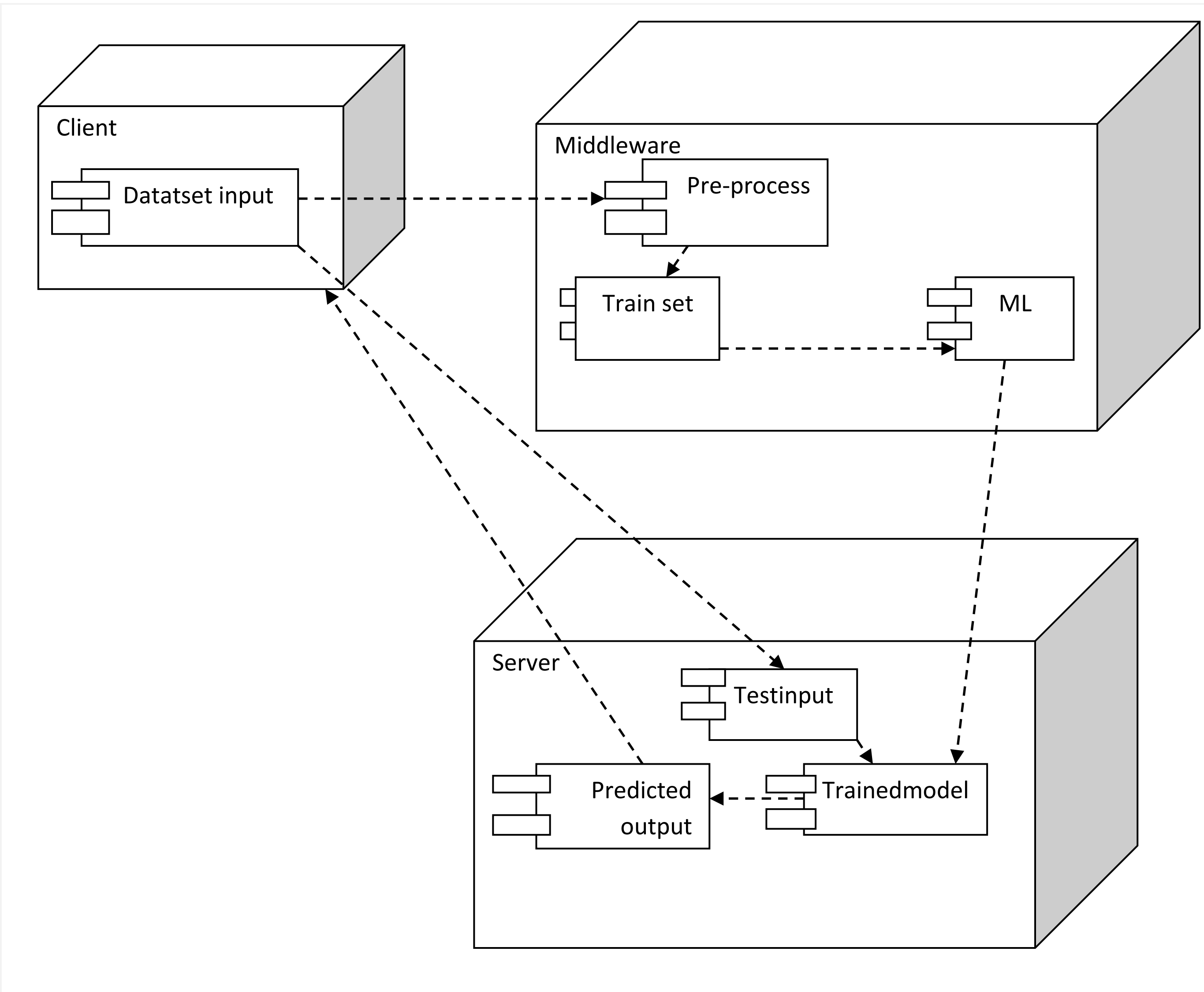


Figure 7: Deployment Diagram

UML model the physical use of artefacts(dataset input, ML etc) at nodes(Client, Middleware and Server) in the deployment diagram. The nodes appear as boxes and the items assigned to every node show up in the boxes as rectangles. Subnodes that appear as nested boxes may be present.

A single node can conceptually represent multiple physical nodes in an implementing diagram, for instance a database cluster.

CHAPTER 5

MODULE DESCRIPTION AND ALGORITHMS

Divided into three steps in prediction of heart disease various algorithms, after browsing data from dataset the interface contains preprocess, Decision Tree, Random Forest, Hybrid and user can Predict based on their own input(figure mentioned in screenshot) ,in each of the modules user can see heart disease present or not based on various aspects like confusion matrix, correlation matrix and accuracy. These modules included in our implementation are as follows

5.1 DECISION TREE

A decision tree is a type of supervised learning algorithm that is frequently used to solve classification problems. It can handle categorical as well as continuous input and output variables. We divide samples into the most significant dividers/differentiation in input variable into two or more homogeneous sets (or subpupulations). In the internal node of the decision tree, the branch shows the results and the leaf represents decisions made after the attribute..

Decision Tree functions as follows

- Set the best dataset attribute as the tree's root.
- Subdivide the training into smaller groups. Subsets should be created with data with the same attribute value in each
- Repeat steps 1 and 2 for each sub-set until the tree's branches have leaf nodes on all branches.

Decision trees begin to predict the class label at the bottom of a tree. Compare the value of the root attribute to the value of the record. Follow the value's branch and move on to the next node based on the comparison.

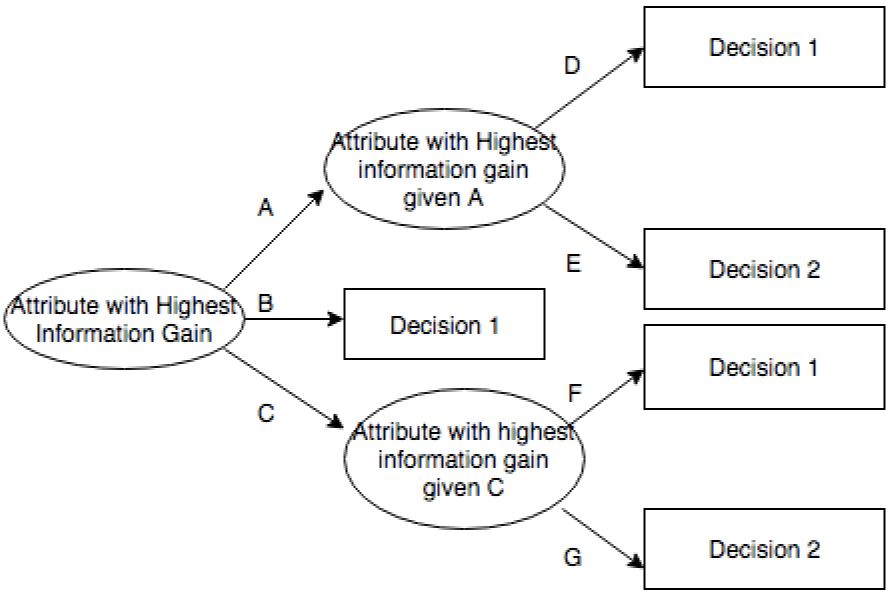


Figure 8: Flow chart Decision Tree algorithm

5.2 RANDOM FOREST MODEL

1. A total of n cases are included in the training dataset. Random substitution is used to select sub-samples from the n cases. These random subsamples from the training data set are used for the construction of individual trees.

2. As long as k variables for input are available, m is chosen to m < k. At each node m variables are randomly selected from k variables. The best division of these m variables is chosen to divide the node. While the forest is growing, its value remains unchanged.

3. Without powdering each tree is cultivated as wide as possible.

4. The new object class is predicted on the basis of a majority of the votes received in all decision-making bodies.

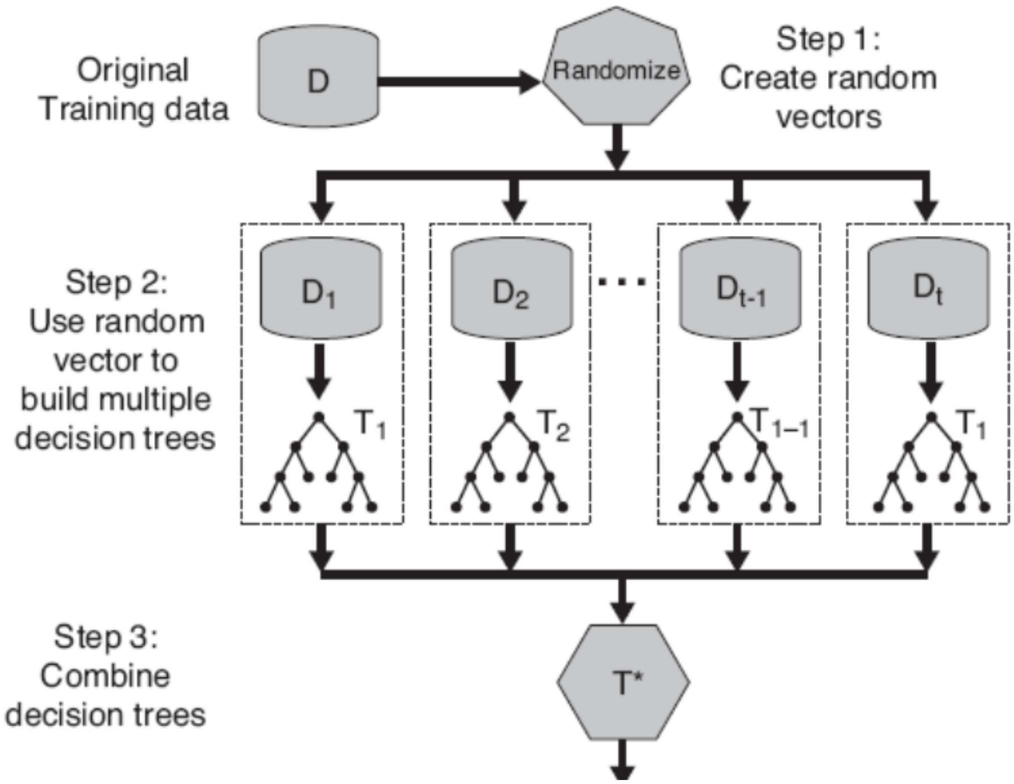


Figure 9: Flow chart of Random Forest

5.3 HYBRID MODEL

The random forest and decision tree algorithms are combined to create a hybrid model. In the combined model, random forest probability is used. The combined model works based probabilities of random forest. The probabilities from random forest is added to train data and fed to decision tree algorithm. Similarly decision tree probabilities are identified and fed to test data. At long last, the figures are known.From the Hybrid of Decision tree and Random forest we can get advantages of both the algorithms like handeling of numerical , categorical data and large dataset with higher dimensionality, less effort for data preparation ad handle missing value and maintains accuracy. Hybrid model also gives less errors of R-squared etc which is very essential in sensitive prediction of diseases like heart disease, which means it is the best classification algorithm for heart disease prediction however applying hybrid model has its own difficulties.

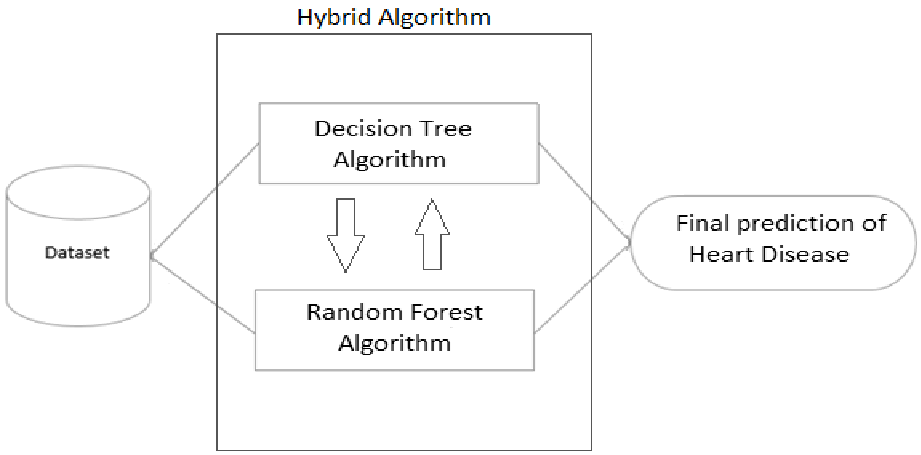


Figure 10: Hybrid Model of Decision Tree and Random Forest

CHAPTER 6

IMPLEMENTATION

6.1 IMPLEMENTATION METHODOLOGY

The work in Python 3.6.4 is conducted with libraries, scientists and physicians. We have downloaded uci.edu dataset. The downloaded data includes binary heart disease classes. An algorithm of machine learning like a decision tree, an arbitrary forest and a hybrid model.

6.2 DATA DICTIONARY

Data collection with attributes such as age, sex, cp, trestbps, chol, fbs, blecg, thalach, exang, oldpeak. This figure shows the samples of the data collected.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slop	ca	thal	pred_attribute
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	2
67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	3
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	2

Figure 11: Dataset used for Study

The dataset variable names are described below

Name of variable	**Short description**	**Name of variable**	**Short description**
Age	Patient Age	thalach	Maximum achieved heart rate
Sex	Male Sex, 1	exang	Induced angina exercise (1 yes)

cp	chest pain	oldpeak	Inducing ST depression. e.g.
trestbps	Blood pressure restoration	slope	Pitch exercise Slope ST
chol	Cholesterol in serum	ca	the number of large vessels
fbs	a fasting blood sugar level of 120mg/dl or higher (1 true)	thal	There is no explanation given, but it is most likely thalassemia (3 normal; 6 fixed defect; 7 reversable defect)
restecg	electroc. result at rest (1 anomaly)	num	heart disease diagnosis (angiographic disease status)

Table: Variable names of dataset

CHAPTER 7

EXPERIMENTAL RESULTS AND ANALYSIS

Scientist learning, pandas, and matplotlib are used in the proposed work in Python 3.6.4. Downloaded from uci.edu the heart disease data set is included in the study. An algorithm for machine learning is used, including decision trees and random forests. This algorithm for learning machines and heart disease has been used. We have applied the Hybrid Decision Tree and Random Improving working quality and innovation, forest models are used. The results show that the detection of heart disease by Random Forest and hybrid models is efficient. The forest random achieves precision 76%, the Decision Tree achieves precision of about 79%, the Hybrid model achieves precision of 76%.

In our experimental study, the following table demonstrates accuracy.

Algorithm	**Accuracy (%)**
Decision Tree	79
Random Forest	76
Hybrid (Decision Tree+ Random Forest)	76

Table: Experimental Results of proposed system

The following figure shows the exact comparison of our work.

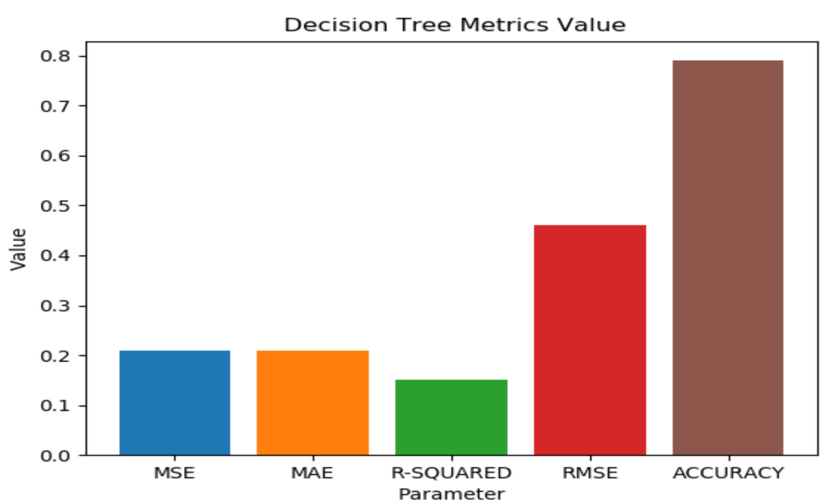


Figure 12: Evaluation metrics for Decision Tree algorithm

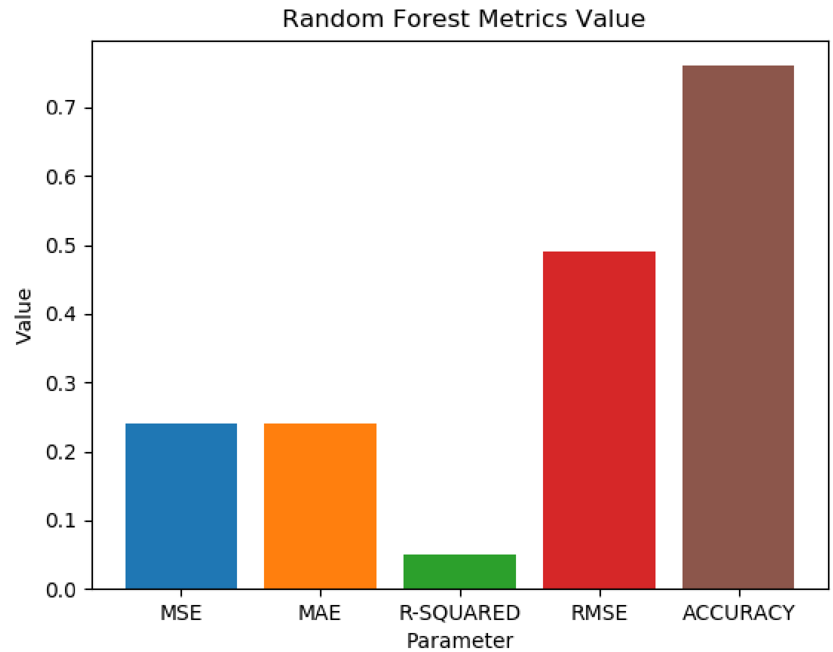


Figure 13: Evaluation metrics for Random Forest algorithm

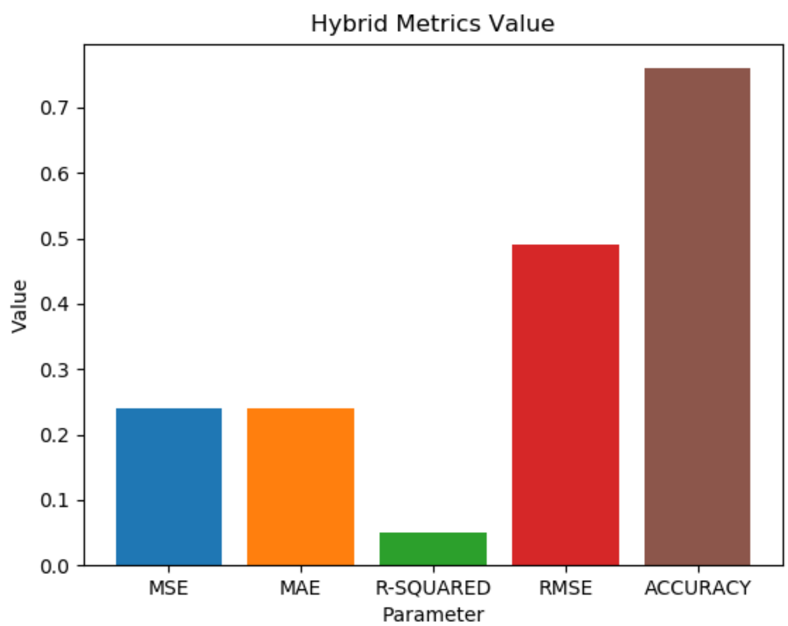


Figure 14: Evaluation metrics for Hybrid algorithm

CHAPTER 8

CONCLUSION AND FUTURE WORK

Finally, as determined by the literature review, it is necessary to develop combination and More complex models to improve the accuracy of cardiovascular early onset disease. The framework proposed using Decision Tree and Random Forest combinations for cardiovascular prediction. The system can use the database of Cleveland Heart Disease to train and test and thus achieve the most efficient model. In future, we are interested to study some of the deep learning models such as CNN or DNN algorithm for heart disease prediction. We are also interested in classifying it as an issue for the multi-class identification of disease level.

CHAPTER 9

REFERENCES

[1] Mackay,J., Mensah,G. 2004 “Atlas of Heart Disease and Stroke” Nonserial Publication, ISBN-13 9789241562768 ISBN-10 9241562765.

[2] Robert Detrano 1989 “Cleveland Heart Disease Database” V.A. Medical Center, Long Beach and Cleveland Clinic Foundation.

[3] Yanwei Xing, Jie Wang and Zhihong Zhao Yonghong Gao 2007 “Combination data mining methods with new medical data to predicting outcome of Coronary Heart Disease” Convergence Information Technology, 2007. International Conference November 2007, pp 868-872.

[4] Jianxin Chen, Guangcheng Xi, Yanwei Xing, Jing Chen, and Jie Wang 2007 “Predicting Syndrome by NEI Specifications: A Comparison of Five Data Mining Algorithms in Coronary Heart Disease” Life System Modeling and Simulation Lecture Notes in Computer Science, pp 129-135.

[5] Jyoti Soni, Ujma Ansari, Dipesh Sharma 2011 “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction” International Journal of Computer Applications, doi 10.5120/2237-2860.

[6] Mai Shouman, Tim Turner, Rob Stocker 2012 “Using Data Mining Techniques In Heart Disease Diagnoses And Treatment“ Electronics, Communications and Computers (JECECC), 2012 Japan-Egypt Conference March 2012, pp 173-177.

[7] Robert Detrano, Andras Janosi, Walter Steinbrunn, Matthias Pfisterer, Johann-Jakob Schmid, Sarbjit Sandhu, Kern H. Guppy, Stella Lee, Victor Froelicher 1989 “International application of a new probability algorithm for the diagnosis of coronary artery disease” The American Journal of Cardiology, pp 304-310.15

[8] Polat, K., S. Sahan, and S. Gunes 2007 “Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn

(nearest neighbour) based weighting preprocessing" Expert Systems with Applications 2007, pp 625-631.

[9] Ozsen, S., Gunes, S. 2009 "Attribute weighting via genetic algorithms for attribute weighted artificial immune system (AWAIS) and its application to heart disease and liver disorders problems" Expert Systems with Applications, pp 386-392.

[10] Resul Das, Ibrahim Turkoglub, and Abdulkadir Sengurb 2009 "Effective diagnosis of heart disease through neural networks ensembles" Expert Systems with Applications, pp 7675–7680.

CHAPTER 10

ANNEXURE

SOURCE CODE

Main.py

```
import tkinter as tk
from tkinter import Message ,Text
from PIL import Image, ImageTk
import pandas as pd

import tkinter.ttk as ttk
import tkinter.font as font
import tkinter.messagebox as tm
import matplotlib.pyplot as plt

import csv
import numpy as np
from PIL import Image, ImageTk
import pandas as pd
import NeuralNetwork as NN
#import predict as pred
from keras.models import load_model
from sklearn.preprocessing import StandardScaler
from tkinter import filedialog


import RFALG as RF
import DTALG as DT
import Hybrid as hy
import predict as pr

def clear():
    print("Clear1")
    txt.delete(0, 'end')
    txt1.delete(0, 'end')
    txt2.delete(0, 'end')
    txt3.delete(0, 'end')
    txt4.delete(0, 'end')
    txt5.delete(0, 'end')
    txt6.delete(0, 'end')
    txt7.delete(0, 'end')
    txt8.delete(0, 'end')
    txt9.delete(0, 'end')
    txt10.delete(0, 'end')
    txt11.delete(0, 'end')
    txt12.delete(0, 'end')
```

```
    txt13.delete(0, 'end')
    txt14.delete(0, 'end')



window = tk.Tk()
window.title("Heart Disease Prediction")


window.geometry('1280x720')
bgcolor="#ffe6e6"
bgcolor1="#e60000"
fgcolor="#660000"

window.configure(background="#ffe6e6")
#window.attributes('-fullscreen', True)

window.grid_rowconfigure(0, weight=1)
window.grid_columnconfigure(0, weight=1)


message1 = tk.Label(window, text="Heart Disease Prediction"
,bg=bgcolor  ,fg=fgcolor  ,width=50  ,height=2,font=('times', 30,
'italic bold underline'))
message1.place(x=100, y=20)

lbl = tk.Label(window, text="Data Set",width=15  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
lbl.place(x=350, y=150)

txt = tk.Entry(window,width=15,bg=bgcolor ,fg=fgcolor,font=('times',
15, ' bold '))
txt.place(x=600, y=150)


lbl1 = tk.Label(window, text="Age",width=15  ,height=1  ,fg=fgcolor
,bg=bgcolor ,font=('times', 15, ' bold ') )
lbl1.place(x=50, y=200)

txt1 = tk.Entry(window,width=15,bg=bgcolor ,fg=fgcolor,font=('times',
15, ' bold '))
txt1.place(x=300, y=200)

lbl2 = tk.Label(window, text="Sex",width=15  ,fg=fgcolor  ,bg=bgcolor
,height=1 ,font=('times', 15, ' bold '))
lbl2.place(x=50, y=250)

txt2 = tk.Entry(window,width=15  ,bg=bgcolor
,fg=fgcolor,font=('times', 15, ' bold ')  )
txt2.place(x=300, y=250)
```

```
lbl3 = tk.Label(window, text="CP",width=15  ,height=1  ,fg=fgcolor
,bg=bgcolor ,font=('times', 15, ' bold ') )
lbl3.place(x=50, y=300)

txt3 = tk.Entry(window,width=15,bg=bgcolor ,fg=fgcolor,font=('times',
15, ' bold '))
txt3.place(x=300, y=300)

lbl4 = tk.Label(window, text="tresp",width=15  ,fg=fgcolor
,bg=bgcolor    ,height=1 ,font=('times', 15, ' bold '))
lbl4.place(x=50, y=350)

txt4 = tk.Entry(window,width=15  ,bg=bgcolor
,fg=fgcolor,font=('times', 15, ' bold ')  )
txt4.place(x=300, y=350)

lbl5 = tk.Label(window, text="chol",width=15  ,height=1  ,fg=fgcolor
,bg=bgcolor ,font=('times', 15, ' bold ') )
lbl5.place(x=50, y=400)

txt5 = tk.Entry(window,width=15,bg=bgcolor ,fg=fgcolor,font=('times',
15, ' bold '))
txt5.place(x=300, y=400)

lbl6 = tk.Label(window, text="fbs",width=15  ,fg=fgcolor  ,bg=bgcolor
,height=1 ,font=('times', 15, ' bold '))
lbl6.place(x=50, y=450)

txt6 = tk.Entry(window,width=15  ,bg=bgcolor
,fg=fgcolor,font=('times', 15, ' bold ')  )
txt6.place(x=300, y=450)


lbl7 = tk.Label(window, text="restecg",width=15  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
lbl7.place(x=50, y=500)

txt7 = tk.Entry(window,width=15,bg=bgcolor ,fg=fgcolor,font=('times',
15, ' bold '))
txt7.place(x=300, y=500)


lbl8 = tk.Label(window, text="thalach",width=15  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
lbl8.place(x=600, y=200)

txt8 = tk.Entry(window,width=15,bg=bgcolor ,fg=fgcolor,font=('times',
15, ' bold '))
txt8.place(x=850, y=200)

lbl9 = tk.Label(window, text="exang",width=15  ,height=1  ,fg=fgcolor
,bg=bgcolor ,font=('times', 15, ' bold ') )
```

```
lbl9.place(x=600, y=250)

txt9 = tk.Entry(window,width=15,bg=bgcolor ,fg=fgcolor,font=('times',
15, ' bold '))
txt9.place(x=850, y=250)

lbl10 = tk.Label(window, text="oldpeak",width=15  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
lbl10.place(x=600, y=300)

txt10 = tk.Entry(window,width=15,bg=bgcolor ,fg=fgcolor,font=('times',
15, ' bold '))
txt10.place(x=850, y=300)


lbl11 = tk.Label(window, text="slope",width=15  ,height=1  ,fg=fgcolor
,bg=bgcolor ,font=('times', 15, ' bold ') )
lbl11.place(x=600, y=350)

txt11 = tk.Entry(window,width=15,bg=bgcolor ,fg=fgcolor,font=('times',
15, ' bold '))
txt11.place(x=850, y=350)

lbl12 = tk.Label(window, text="ca",width=15  ,height=1  ,fg=fgcolor
,bg=bgcolor ,font=('times', 15, ' bold ') )
lbl12.place(x=600, y=400)

txt12 = tk.Entry(window,width=15,bg=bgcolor ,fg=fgcolor,font=('times',
15, ' bold '))
txt12.place(x=850, y=400)

lbl13 = tk.Label(window, text="thal",width=15  ,height=1  ,fg=fgcolor
,bg=bgcolor ,font=('times', 15, ' bold ') )
lbl13.place(x=600, y=450)

txt13 = tk.Entry(window,width=15,bg=bgcolor ,fg=fgcolor,font=('times',
15, ' bold '))
txt13.place(x=850, y=450)

lbl14 = tk.Label(window, text="Predicted Value",width=15  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
lbl14.place(x=600, y=500)

txt14 = tk.Entry(window,width=25,bg=bgcolor ,fg=fgcolor,font=('times',
15, ' bold '))
txt14.place(x=850, y=500)


elbl1 = tk.Label(window, text="Ex:60",width=7  ,height=1  ,fg=fgcolor
,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl1.place(x=500, y=200)
```

```
elbl2 = tk.Label(window, text="0-M/1-F",width=7  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl2.place(x=500, y=250)

elbl3 = tk.Label(window, text="Ex:1-4",width=7  ,height=1  ,fg=fgcolor
,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl3.place(x=500, y=300)

elbl4 = tk.Label(window, text="Ex:90-180",width=7  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl4.place(x=500, y=350)

elbl5 = tk.Label(window, text="Ex:180-320",width=7  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl5.place(x=500, y=400)


elbl6 = tk.Label(window, text="Ex:0/1",width=7  ,height=1  ,fg=fgcolor
,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl6.place(x=500, y=450)


elbl7 = tk.Label(window, text="Ex:0/2",width=7  ,height=1  ,fg=fgcolor
,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl7.place(x=500, y=500)


elbl8 = tk.Label(window, text="Ex:90-200",width=7  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl8.place(x=1020, y=200)

elbl9 = tk.Label(window, text="Ex:0/1",width=7  ,height=1  ,fg=fgcolor
,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl9.place(x=1020, y=250)

elbl10 = tk.Label(window, text="Ex:0.0-4.0",width=7  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl10.place(x=1020, y=300)

elbl11 = tk.Label(window, text="Ex:1-3",width=7  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl11.place(x=1020, y=350)

elbl12 = tk.Label(window, text="Ex:0-3",width=7  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl12.place(x=1020, y=400)

elbl13 = tk.Label(window, text="Ex:3/6/7",width=7  ,height=1
,fg=fgcolor  ,bg=bgcolor ,font=('times', 15, ' bold ') )
elbl13.place(x=1020, y=450)
```

```
def browse():
     path=filedialog.askopenfilename()
     print(path)
     txt.delete(0, 'end')
     txt.insert('end',path)
     if path !="":
          print(path)
     else:
          tm.showinfo("Input error", "Select Dataset")


def preprocess():
     path=txt.get()
     if path != "" :
          print("preprocess")
          # read synthetic cleveland dataset from full cleveland.data
          df_main = pd.read_table(path, sep=',')

          # Neural Net transfer function likes to work with floats
          df_main.astype(float)

          # Normalize values to range [0:1]
          df_main /= df_main.max()

          # split data into independent and dependent variables
          y_all = df_main['num']
          X_all = df_main.drop(columns = 'num')

          fig, axs = plt.subplots(nrows=1, ncols=1, sharey=False,
figsize=(10,5))

          # plot histograms
          #axs.set_xlabel('No Heart Disease
Heart Disease')
          axs.set_title('DataSet')
          axs.grid()
          axs.hist(y_all)
          axs.get_children()[0].set_color('g')
          axs.get_children()[2].set_color('c')
          axs.get_children()[5].set_color('b')
          axs.get_children()[7].set_color('y')
          axs.get_children()[9].set_color('r')
          fig.savefig('results/Preprocess.png')
          plt.pause(5)
          plt.show(block=False)
          plt.close()
          tm.showinfo("Input error", "Preprocess Successfully
Finished")
     else:
          tm.showinfo("Input error", "Select Dataset")

def RFprocess():
```

```
        sym=txt.get()
        if sym != "":
              RF.process(sym)
              tm.showinfo("Input", "RandomForest Successfully Finished")
        else:
              tm.showinfo("Input error", "Select Dataset")

def DTprocess():
        sym=txt.get()
        if sym != "":
              DT.process(sym)
              print("DT")
              tm.showinfo("Input", "DT Successfully Finished")
        else:
              tm.showinfo("Input error", "Select Dataset")

def hybridmodel():
        sym=txt.get()
        if sym != "":
              hy.process(sym)
              tm.showinfo("Input", "Hybrid Successfully Finished")
        else:
              tm.showinfo("Input error", "Select Dataset")

def predictprocess():
        print("predict")
        txt14.delete(0, 'end')
        #txt1.insert('end', "60")
        a1=txt1.get()
        a2=txt2.get()
        a3=txt3.get()
        a4=txt4.get()
        a5=txt5.get()
        a6=txt6.get()
        a7=txt7.get()
        a8=txt8.get()
        a9=txt9.get()
        a10=txt10.get()
        a11=txt11.get()
        a12=txt12.get()
        a13=txt13.get()

        if a1 == "":
              tm.showinfo("Insert error", "Enter Age")
        elif a2 == "":
              tm.showinfo("Insert error", "Enter Sex")
        elif a3 == "":
              tm.showinfo("Insert error", "Enter Cp")
        elif a4 == "":
              tm.showinfo("Insert error", "Enter tresp")
        elif a5 == "":
              tm.showinfo("Insert error", "Enter Chol")
```

```
      elif a6 == "":
            tm.showinfo("Insert error", "Enter fbs")
      elif a7 == "":
            tm.showinfo("Insert error", "Enter restecg")
      elif a8 == "":
            tm.showinfo("Insert error", "Enter thalach")
      elif a9 == "":
            tm.showinfo("Insert error", "Enter exang")
      elif a10=="":
            tm.showinfo("Insert error", "Enter oldpeak")
      elif a11 == "":
            tm.showinfo("Insert error", "Enter slope")
      elif a12 == "":
            tm.showinfo("Insert error", "Enter ca")
      elif a13 == "":
            tm.showinfo("Insert error", "Enter thal")
      else:

            #new_pred =
model.predict_classes(np.array([[58,1,3,112,230,0,2,165,0,2.5,2,1,7]])
)  #4
            #new_pred =
model.predict_classes(np.array([[a1,a2,a3,a4,a5,a6,a7,a8,a9,a10,a11,a1
2,a13]]))

      new_pred=pr.process([a1,a2,a3,a4,a5,a6,a7,a8,a9,a10,a11,a12,a13])
            res=int(new_pred[0])
            print(res)
            if res == 0:
                  print("sdsd")
                  txt14.insert('end', "No Heart Disease")
            else:
                  print("sds no")
                  txt14.insert('end', "Heart Disease Possible")

      #nn=np.array([[a1,a2,a3,a4,a5,a6,a7,a8,a9,a10,a11,a12,a13]])
            #print(nn)
            #pred.process(nn)




br = tk.Button(window, text="Browse", command=browse  ,fg=fgcolor
,bg=bgcolor1   ,width=10  ,height=1, activebackground = "Red"
,font=('times', 15, ' bold '))
br.place(x=800, y=140)

clearButton = tk.Button(window, text="Clear", command=clear
,fg=fgcolor  ,bg=bgcolor1  ,width=10  ,height=1 ,activebackground =
"Red" ,font=('times', 15, ' bold '))
clearButton.place(x=950, y=140)
```

```
process = tk.Button(window, text="Preprocess", command=preprocess
,fg=fgcolor  ,bg=bgcolor1   ,width=17  ,height=2, activebackground =
"Red" ,font=('times', 15, ' bold '))
process.place(x=50, y=600)

DTbutton = tk.Button(window, text="Decision Tree", command=DTprocess
,fg=fgcolor  ,bg=bgcolor1  ,width=17  ,height=2, activebackground =
"Red" ,font=('times', 15, ' bold '))
DTbutton.place(x=250, y=600)

RFbutton = tk.Button(window, text="Random Forest", command=RFprocess
,fg=fgcolor  ,bg=bgcolor1   ,width=17  ,height=2, activebackground =
"Red" ,font=('times', 15, ' bold '))
RFbutton.place(x=460, y=600)


HYbutton = tk.Button(window, text="Hybrid", command=hybridmodel
,fg=fgcolor  ,bg=bgcolor1   ,width=17  ,height=2, activebackground =
"Red" ,font=('times', 15, ' bold '))
HYbutton.place(x=650, y=600)


predict = tk.Button(window, text="Predict", command=predictprocess
,fg=fgcolor  ,bg=bgcolor1   ,width=17  ,height=2, activebackground =
"Red" ,font=('times', 15, ' bold '))
predict.place(x=850, y=600)



quitWindow = tk.Button(window, text="Quit", command=window.destroy
,fg=fgcolor  ,bg=bgcolor1   ,width=17  ,height=2, activebackground =
"Red" ,font=('times', 15, ' bold '))
quitWindow.place(x=1060, y=600)


window.mainloop()
```

RFALG.py

```
import pandas as pd
import matplotlib as plt
import numpy as np
from sklearn import linear_model
#from sklearn.model_selection cross_validation
from scipy.stats import norm

from sklearn.svm import SVC
from sklearn import svm
```

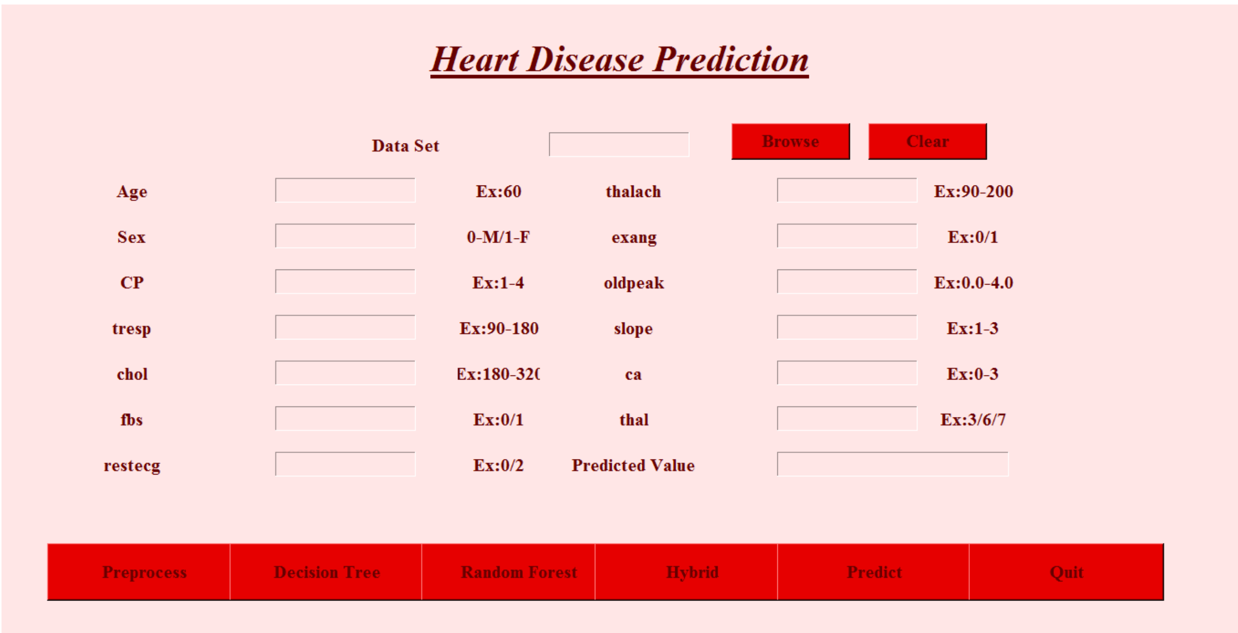
```
from sklearn.svm import LinearSVC
from sklearn.model_selection import train_test_split

from sklearn.metrics import accuracy_score
from random import seed
from random import randrange
from csv import reader
import csv
import numpy as np
import pandas as pd
from pandas import read_csv
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import r2_score
from sklearn.ensemble import RandomForestClassifier



def process(path):
	dataset = pd.read_csv(path)
	X = dataset.iloc[:, 0:13].values
	y = dataset.iloc[:, 13].values

	X_train, X_test, y_train, y_test = train_test_split(X, y)

	model2=RandomForestClassifier()
	model2.fit(X_train, y_train)
	y_pred = model2.predict(X_test)
	print("predicted")
	print(y_pred)
	print("test")
	print(y_test)

	result2=open("results/resultRF.csv","w")
	result2.write("ID,Predicted Value" + "\n")
	for j in range(len(y_pred)):
	    result2.write(str(j+1) + "," + str(y_pred[j]) + "\n")
	result2.close()

	mse=mean_squared_error(y_test, y_pred)
	mae=mean_absolute_error(y_test, y_pred)
	r2=r2_score(y_test, y_pred)


	print("---------------------------------------------------------
")
	print("MSE VALUE FOR RandomForest IS %f "  % mse)
	print("MAE VALUE FOR RandomForest IS %f "  % mae)
	print("R-SQUARED VALUE FOR RandomForest IS %f "  % r2)
	rms = np.sqrt(mean_squared_error(y_test, y_pred))
	print("RMSE VALUE FOR RandomForest IS %f "  % rms)
```

```
    ac=accuracy_score(y_test,y_pred)
    print ("ACCURACY VALUE RandomForest IS %f" % ac)
    print("-------------------------------------------------------------
")


    result2=open('results/RFMetrics.csv', 'w')
    result2.write("Parameter,Value" + "\n")
    result2.write("MSE" + "," +str(mse) + "\n")
    result2.write("MAE" + "," +str(mae) + "\n")
    result2.write("R-SQUARED" + "," +str(r2) + "\n")
    result2.write("RMSE" + "," +str(rms) + "\n")
    result2.write("ACCURACY" + "," +str(ac) + "\n")
    result2.close()


    df =  pd.read_csv('results/RFMetrics.csv')
    acc = df["Value"]
    alc = df["Parameter"]
    colors = ["#1f77b4", "#ff7f0e", "#2ca02c", "#d62728", "#8c564b"]
    explode = (0.1, 0, 0, 0, 0)

    fig = plt.figure()
    plt.bar(alc, acc,color=colors)
    plt.xlabel('Parameter')
    plt.ylabel('Value')
    plt.title(' Random Forest Metrics Value')
    fig.savefig('results/RFMetricsValue.png')
    plt.pause(5)
    plt.show(block=False)
    plt.close()
```

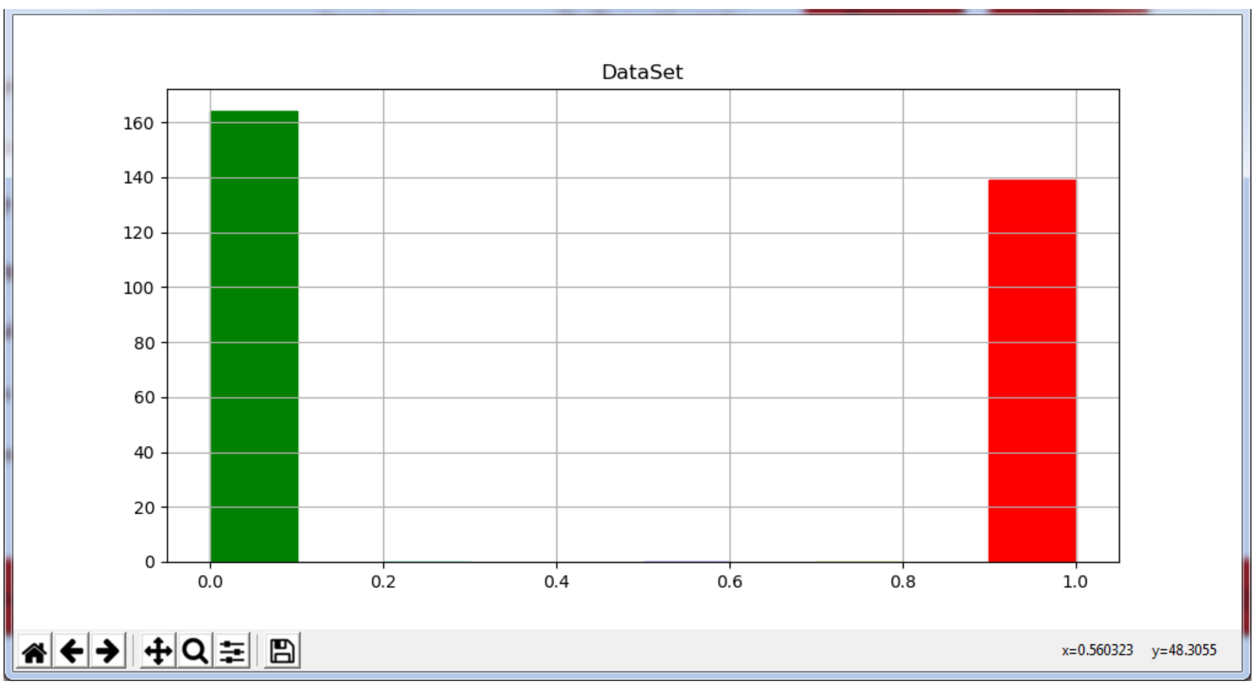
CHAPTER 11

SCREEN SHOTS

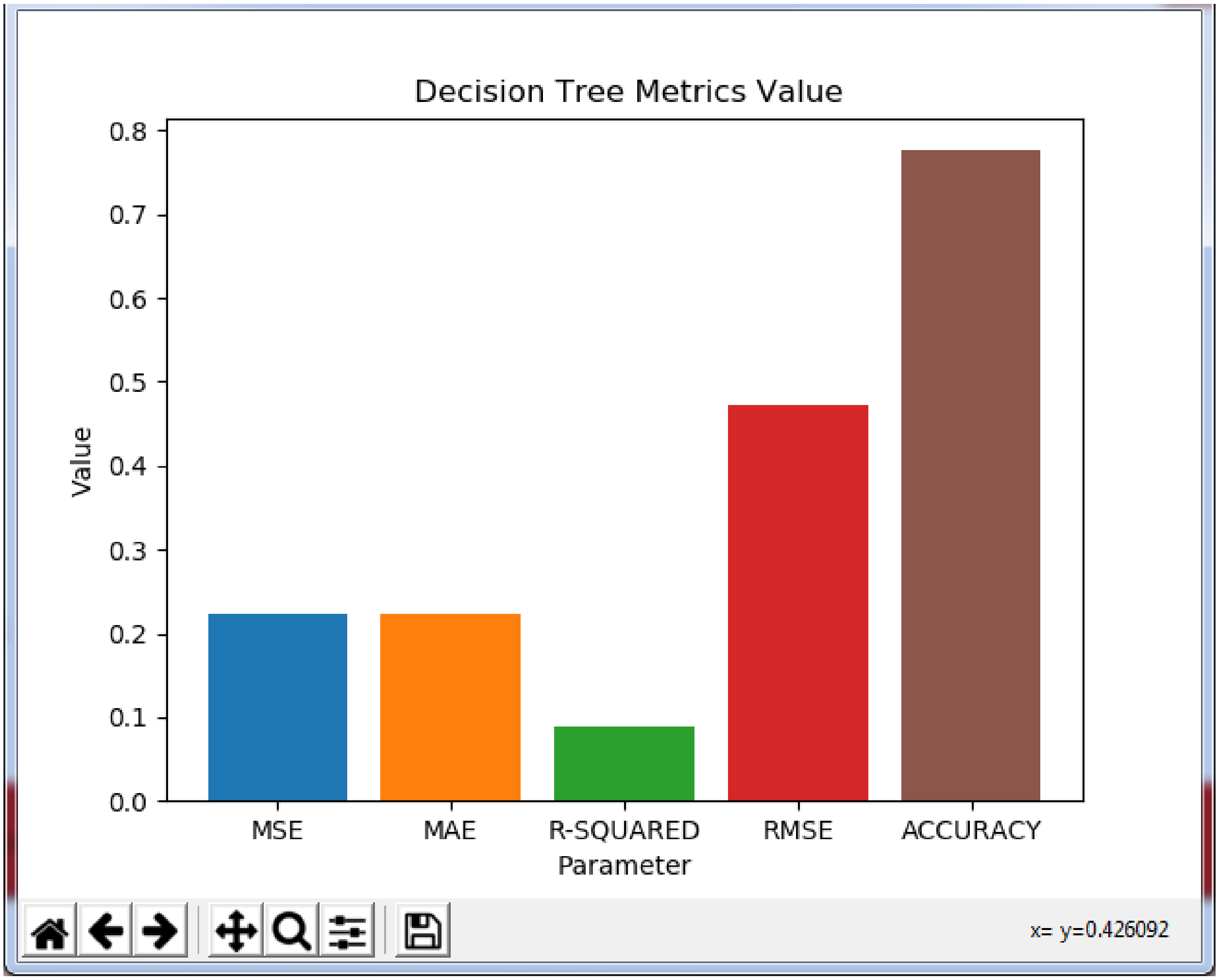
The following screen shows the application home page



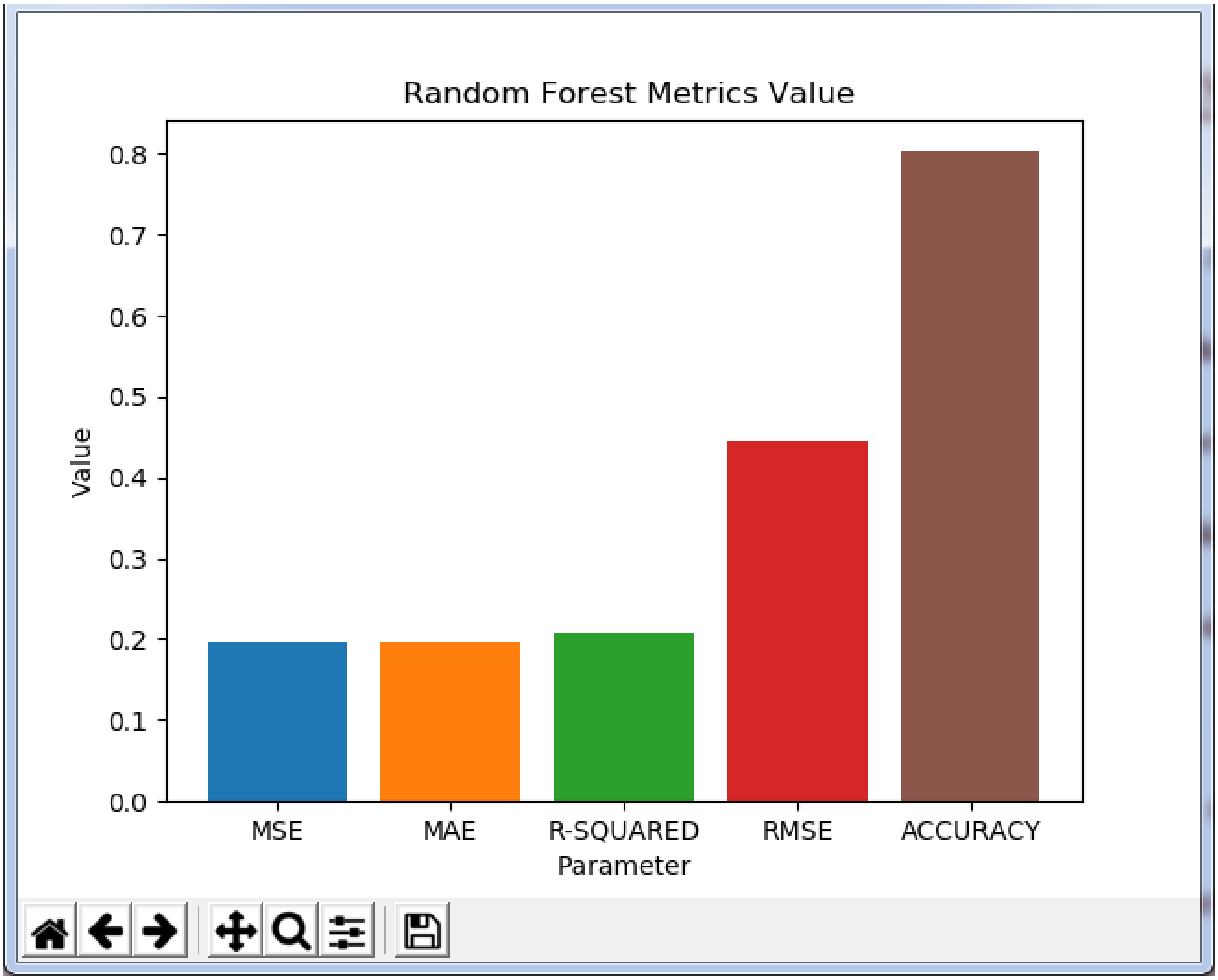
The following screen shows the result of pre-process

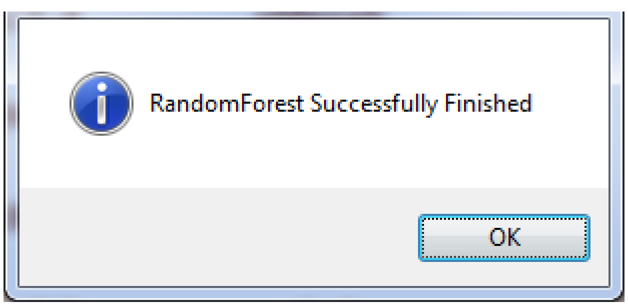


The following screen shows the result of decision tree algorithm

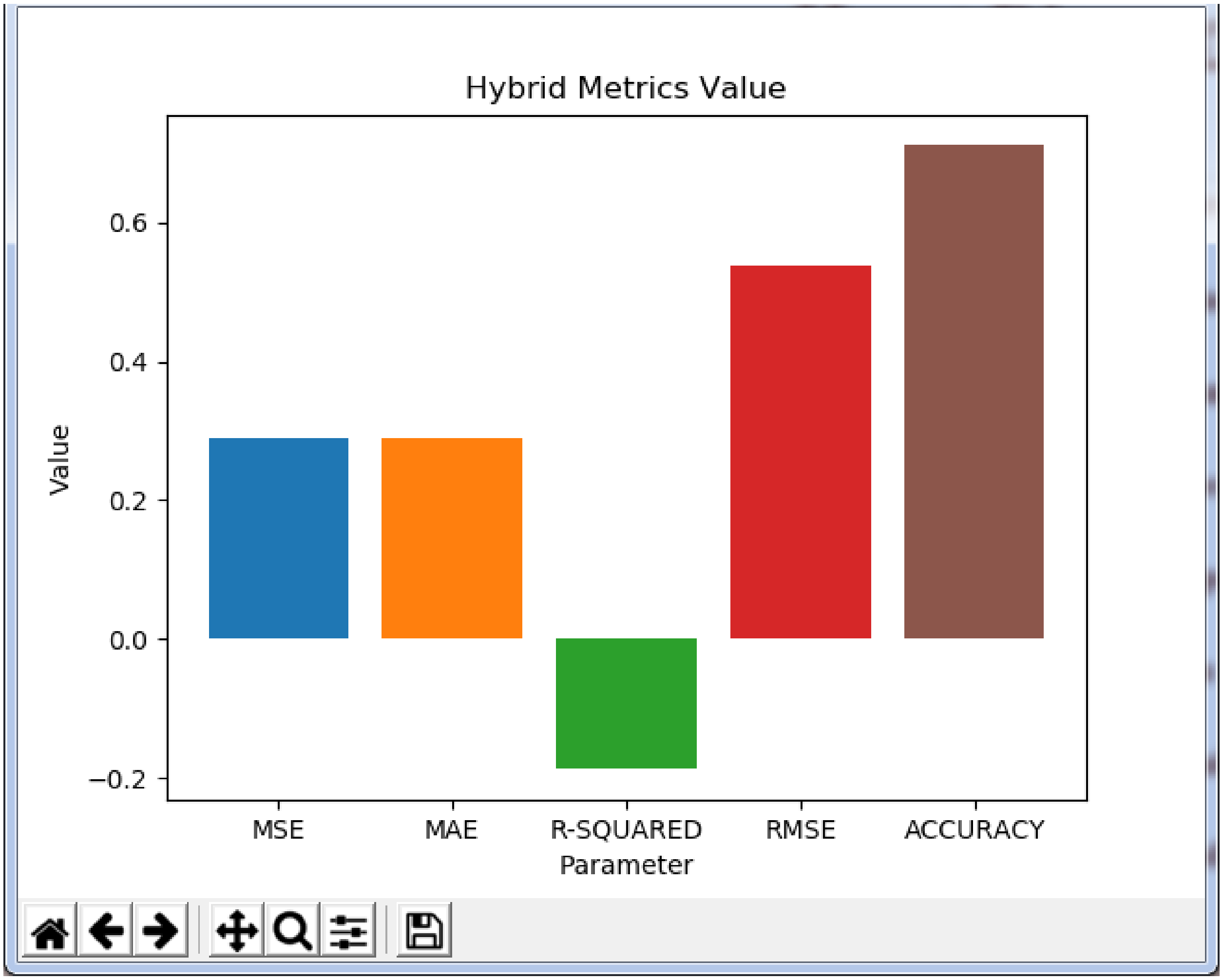


The following screen shows the results of random forest algorithm





The following screen shows the results of Hybrid model



CHAPTER 12

PLAGIARISM REPORT

ORIGINALITY REPORT

16%	7%	11%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	G Renugadevi, G Asha Priya, B Dhivyaa Sankari, R Gowthamani. "Predicting heart disease using hybrid machine learning model", Journal of Physics: Conference Series, 2021 Publication	4%
2	Submitted to B.S.Abdur Rahman Crescent Institute of Science & Technology Student Paper	2%
3	M. Kavitha, G. Gnaneswar, R. Dinesh, Y. Rohith Sai, R. Sai Suraj. "Heart Disease Prediction using Hybrid machine Learning Model", 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021 Publication	2%
4	Tulay Karayilan, Ozkan Kilic. "Prediction of heart disease using neural network", 2017 International Conference on Computer Science and Engineering (UBMK), 2017 Publication	1%
5	Submitted to SASTRA University Student Paper	

		1%
6	www.irjet.net Internet Source	1%
7	www.rjset.com Internet Source	<1%
8	www.ijert.org Internet Source	<1%
9	www.semanticscholar.org Internet Source	<1%
10	Submitted to Middlesex University Student Paper	<1%
11	Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava. "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE Access, 2019 Publication	<1%
12	www.igi-global.com Internet Source	<1%
13	research.ijcaonline.org Internet Source	<1%
14	"Mining Intelligence and Knowledge Exploration", Springer Science and Business Media LLC, 2013 Publication	<1%

15	medium.com Internet Source	<1 %
16	www.slideshare.net Internet Source	<1 %
17	www.ijitee.org Internet Source	<1 %
18	www.mirlabs.net Internet Source	<1 %
19	Elham Nasarian, Moloud Abdar, Mohammad Amin Fahami, Roohallah Alizadehsani et al. "Association between work-related features and coronary artery disease: A heterogeneous hybrid feature selection integrated with balancing approach", Pattern Recognition Letters, 2020 Publication	<1 %
20	Submitted to SUNY, Binghamton Student Paper	<1 %
21	www.ukessays.com Internet Source	<1 %
22	"Proceedings of the Second International Conference on Computational Intelligence and Informatics", Springer Science and Business Media LLC, 2018 Publication	<1 %

23	A. Kondababu, V. Siddhartha, BHK. Bhagath Kumar, Bujjibabu Penumutchi. "A comparative study on machine learning based heart disease prediction", Materials Today: Proceedings, 2021 Publication	<1 %
24	Roohallah Alizadehsani, Mohammad Javad Hosseini, Zahra Alizadeh Sani, Asma Ghandeharioun, Reihane Boghrati. "Diagnosis of Coronary Artery Disease Using Cost-Sensitive Algorithms", 2012 IEEE 12th International Conference on Data Mining Workshops, 2012 Publication	<1 %
25	Submitted to SRM University Student Paper	<1 %
26	inpressco.com Internet Source	<1 %
27	studentsrepo.um.edu.my Internet Source	<1 %
28	www.ijeat.org Internet Source	<1 %
29	Das, R.. "Effective diagnosis of heart disease through neural networks ensembles", Expert Systems With Applications, 200905 Publication	<1 %

CHAPTER 13

PROOF OF PUBLICATION

rohitshrivastava RA1711003010511   Sat, Apr 3, 8:32 PM (6 days ago)

to me

---------- Forwarded message ---------
From: **prof prince** <conferenceiccet@gmail.com>
Date: Sat, 3 Apr, 2021, 8:27 pm
Subject: Heart Disease Prediction Analysis Through Ensemble Learning Model
To: rohitshrivastava RA1711003010511 <rr5921@srmist.edu.in>

Dear Author/s,

We are happy to inform you that your paper, submitted for the conference ICCET 2021 has been **Accepted** based on the recommendations provided by the Technical Review Committee. By this mail you are requested to proceed with Registration for the Conference. Most notable is that the Conference must be registered on or before **April 8, 2021** from the date of acceptance.

www.iccet.in

Kindly fill the registration form, declaration form which is attached with the mail and it should reach us on above mentioned days.

9th INTERNATIONAL CONFERENCE ON CONTEMPORARY ENGINEERING AND TECHNOLOGY 2021
ORGANIZED BY
ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING AND TECHNOLOGY (OSIET), CHENNAI, INDIA.
IN ASSOCIATION WITH
PRINCE SHRI VENKATESHWARA PADMAVATHY ENGINEERING COLLEGE
PRINCE DR. K. VASUDEVAN COLLEGE OF ENGINEERING & TECHNOLOGY
(Approved By All India Council For Technical Education, Affiliated To Anna University)
Medavakkam - Mambakkam Main Road, Ponmar, Chennai - 600 127.

Certificate of Presentation

This is to certify that Mr/Mrs/Dr P. Vigneshwaran *from* SRM Institute of Science and Technology *has presented a paper titled* Heart Disease Prediction Analysis Through Ensemble Learning Model *in the "9th International Conference on Contemporary Engineering and Technology 2021" held on 10th & 11th April 2021.*

Dr. P. Aravinthan, M.E., Ph.D., Co-Convener - ICCET
Dr. A. Krishnamoorthy, M.E., Ph.D, Convener - ICCET
K.Janani, M.Tech., CEO, OSIET

9th INTERNATIONAL CONFERENCE ON CONTEMPORARY ENGINEERING AND TECHNOLOGY 2021
ORGANIZED BY
ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING AND TECHNOLOGY (OSIET), CHENNAI, INDIA.
IN ASSOCIATION WITH
PRINCE SHRI VENKATESHWARA PADMAVATHY ENGINEERING COLLEGE
PRINCE DR. K. VASUDEVAN COLLEGE OF ENGINEERING & TECHNOLOGY
(Approved By All India Council For Technical Education, Affiliated To Anna University)
Medavakkam - Mambakkam Main Road, Ponmar, Chennai - 600 127.

Certificate of Presentation

This is to certify that Mr/Mrs/Dr Anand Kumar Soni *from* SRM Institute of Science and Technology *has presented a paper titled* Heart Disease Prediction Analysis Through Ensemble Learning Model *in the "9th International Conference on Contemporary Engineering and Technology 2021" held on 10th & 11th April 2021.*

Dr. P. Aravinthan, M.E., Ph.D., Co-Convener - ICCET
Dr. A. Krishnamoorthy, M.E., Ph.D, Convener - ICCET
K.Janani, M.Tech., CEO, OSIET

From: **prof prince** <conferenceiccet@gmail.com>
Date: Sat, 15 May, 2021, 12:03 pm
Subject: Re: ICCET21539
To: rohitshrivastava RA1711003010511 <rr5921@srmist.edu.in>

We have sent the attached paper.

Regards,
ICCET