HEART DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS

A PROJECT REPORT

Submitted by

LOGESH RAJ R (2019108026)

MITHUN RAJ S (2019108030)

MOHAMED AFZAL M (2019108032)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHNOLOGY

SETHU INSTITUTE OF TECHNOLOGY

An Autonomous Institution | Accredited with 'A' Grade by NAAC

PULLOOR, KARIAPATTI-626 115.

ANNA UNIVERSITY: CHENNAI 600 025

MAY 2023

SETHU INSTITUTE OF TECHNOLOGY

An Autonomous Institution | Accredited with 'A' Grade by NAAC

PULLOOR, KARIAPATTI-626 115

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"HEART DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS"** is the bonafide work of **"Logesh Raj R (2019108026), Mithun Raj S (2019108030), Mohamed Afzal M (2019108032)"** who carried out the project work under my supervision.

SIGNATURE	**SIGNATURE**
Dr.M.Poomani @ Punitha B.Tech.,M.E.,Ph.D.,	**Mrs.G.Rajalakshmi**
Head of the Department	**Assistant Professor**
Department of Information Technology	Department of Information Technology
Sethu Institute of technology	Sethu Institute of technology
Pullor, Kariapatti-626 115.	Pullor, Kariapatti-626 115.

Submitted for the project viva-vice conducted on_____________

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

First, we would like to thank **GOD** the Almighty for giving us the talent and opportunity to complete our project. We wish to express our earned great fullness to our Honorable Founder and Chairman, **Mr. S. MOHAMED JALEEL B.Sc., B.L.,** for his encouragement extended us to undertake this project work.

We wish to thank and express our gratitude to our Chief Executive Officer **Mr. S. M.SEENI MOHIDEEN M.B.A.,** and Joint chief executive officer, **Mr. S. M. SEENI MOHAMED ALIAR MARAIKKAYAR B.E.,MBA.,** for their support in this project.

We would like to thank and express our gratitude to our respected Principal, **Dr. A. SENTHIL KUMAR B.E., M.E.,Ph.D.,** for providing all necessary for the completion of the project. We wish to thank and express our gratitude to our Vice Principal, **Dr. G. D. SIVA KUMAR B.E., M.E.,Ph.D.,** for his support to our project work.

We wish to express our profound gratitude to our Dean & Professor **Dr.S.SIVA RANJANI B.E.,M.E.,Ph.D.,** and the Head of the Department **DR.M.POOMANI@PUNITHA B.Tech.,M.E.,Ph.D.,** for granting us the necessary permission to proceed with our project and who motivated and encouraged us to do this outrival project for this academic year.

Our profound, delightful and sincere thanks to our project guide **[GUIDE NAME]** whose support was inevitable during the entire period of our work.

We thank our teaching staff for sharing their knowledge and view to enhance our project. We thank our parents for giving us such a wonderful life and our friends for their friendly encouragement throughout the project. Finally, we thank the almighty for giving the full health to finish the project successfully.

ABSTRACT

Heart Disease is the most dangerous life-threatening chronic disease globally. The objective of the work is to predicts the occurrence of heart disease of a patient using Random Forest Algorithm. The dataset was accessed from Kaggle site. The dataset contains 1025 samples and 14 attributes are taken for features of the dataset. Then it was processed using python open access software in jupyter notebook. The datasets are classified and processed using machine learning algorithm Random Forest. The outcomes of the dataset are expressed in terms of accuracy, sensitivity, and specificity in percentage. Using random forest algorithm, we obtained accuracy of 96% for prediction of heart disease with sensitivity value 93% and specificity value 88%. The random forest algorithm has proven to be the most efficient algorithm for classification of heart disease and therefore it is used in the proposed system.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	**iv**
	TABLE OF CONTENTS	**v**
	LIST OF FIGURES	**vii**
	LIST OF SYMBOLS	**viii**
	LIST OF ABBREVIATIONS	**ix**
1	**INTRODUCTION**	**1**
	1.1 OVERVIEW	1
	1.2 GENERAL	1
	1.3 PROBLEM STATEMENT	2
2	**LITERATURE REVIEW**	**3**
3	**SYSTEM STUDY**	**8**
	3.1 FEASIBILITY STUDY	8
	3.2 ECONOMIC FEASIBILITY	8
	3.3 TECHNICAL FEASIBILITY	9
	3.4 BEHAVIOURAL FEASIBILITY	9
4	**PROBLEM DEFINITION**	**11**
	4.1 EXISTING SYSTEM	11
	4.2 PROPOSED SYSTEM	11
5	**SYSTEM REQUIREMENTS**	**12**
	5.1 HARDWARE REQUIREMENT	12
	5.2 SOFTWARE REQUIREMENT	12
	5.3 DESCRIPTION OF TECHNOLOGY	12
6	**SYSTEM DESIGN**	**13**
	6.1 ARCHITECTURE DIAGRAM	13
	6.2 FLOW DIAGRAM	14
7	**SYSTEM IMPLEMENTATION**	**15**

	7.1MODULES DESCRIPTION	**15**
8	**SYSTEM TESTING**	**19**
	8.1 UNIT TESTING	19
	8.2 INTEGRATION TESTING	19
	8.3 WHITE BOX TESTING	20
	8.4 BLACK BOX TESTING	20
	8.5 ACCEPTANCE TESTING	21
9	**RESULTS AND DISCUSSION**	**22**
10	**CONCLUSION AND FUTURE ENHANCEMENT**	**23**
11	**APPENDIX**	**24**
	11.1 SAMPLE CODE	24
	11.2 SCREEN SHOTS	32
	REFERENCES	**34**
	COs & POs MAPPING	
	LIST OF PUBLICATIONS	

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
6.1	ARCHITECTURE DIAGRAM	13
6.2	FLOW DIAGRAM	14
7.1.1	COLLECTION OF DATA	15
7.1.2	CORRELATION MATRIX	16
7.1.3	DATA PRE-PROCESSING	17
7.1.4	DATA BALANCING	18
7.1.5	PREDICTION OF DISEASE	18
11.2	INPUT/OUTPUT SCREEN SHOTS	32

LIST OF SYMBOLS

TP	Number of people with heart diseases.
TN	Number of people with heart diseases and no heart diseases.
FP	Number of people with no heart diseases.
FN	Number of people with no heart diseases and with heart diseases.
f(x)	Output between the 0 and 1 value.
e	Base of the natural logarithm.
x	Input to the function.

LIST OF ABBREVIATIONS

ACRONYMS	ABBREVIATIONS
ICD	INTERNATION CLASSIFICATION OF DISEASE
CVD	CARDIOVASULAR DISEASE
IHD	ISCHEMIC HEART DISEASE
WHO	WORLD HEALTH ORGANIZATION
ROI	RETURN OF INVESTMENT
ORM	OBJECT RELATIONAL MAPPING

CHAPTER 1
INTRODUCTION

1.1. OVERVIEW

According to the World Health Organization, every year 12 million deaths occur worldwide due to heart disease. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of data analysis. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications.

Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aims to predict future heart disease by analysing data of patients which classifies whether they have heart disease or not using machine-learning algorithm. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can say that this technique can be very well adapted to do the prediction of heart disease.

1.2. GENERAL

The main motivation of doing this research is to present a heart disease prediction model for the prediction of occurrence of heart disease. Further, this research work is aimed towards identifying the best classification algorithm for identifying the possibility of heart disease in a patient. This work is justified by performing a comparative study and analysis using three classification algorithms namely Naïve Bayes, Decision Tree, and Random Forest are used at different levels of evaluations. Although these are commonly used machine learning algorithms, the heart disease prediction is a vital task involving highest possible accuracy. Hence, the three algorithms are evaluated at numerous levels and types of evaluation strategies. This will provide researchers and medical practitioners to establish a better.

1.3. PROBLEM STATEMENT

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either it is expensive or are not efficient to calculate chance of heart disease in human. Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients every day in all cases accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time, and expertise. Since we have a good amount of data in today's world, we can use various machine learning algorithms to analyse the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data.

CHAPTER 2

LITERATURE REVIEW

2.1. Deaths: leading causes for 2010

Objectives: This report presents final 2010 data on the 10 leading causes of death in the United States by age, sex, race, and Hispanic origin. Leading causes of infant, neonatal, and post neonatal death are also presented. This report supplements the Division of Vital Statistics' annual report of final mortality statistics.

Methods: Data in this report are based on information from all death certificates filed in the 50 states and the District of Columbia in 2010. Causes of death classified by the International Classification of Diseases, Tenth Revision (ICD-10) are ranked according to the number of deaths assigned to rank able causes. Cause-of-death statistics are based on the underlying cause of death.

Results: In 2010, the 10 leading causes of death were, in rank order: Diseases of heart; Malignant neoplasms; Chronic lower respiratory diseases; Cerebrovascular diseases; Accidents (unintentional injuries); Alzheimer's disease; Diabetes mellitus; Nephritis, nephrotic syndrome and nephrosis; Influenza and pneumonia; and Intentional self-harm (suicide). These 10 causes accounted for 75% of all deaths occurring in the United States. Differences in the rankings are evident by age, sex, race, and Hispanic origin. Leading causes of infant death for 2010 were, in rank order: Congenital malformations, deformations and chromosomal abnormalities; Disorders related to short gestation and low birth weight, not elsewhere classified; Sudden infant death syndrome; New-born affected by maternal complications of pregnancy; Accidents (unintentional injuries); Newburn affected by complications of placenta, cord and membranes; Bacterial sepsis of new born; Respiratory distress of Newburn; Diseases of the circulatory system; and Necrotizing enterocolitis of Newburn. Important variations in the leading causes of infant death are noted for the neonatal and post-neonatal periods.

2.2. Cardiovascular Diseases in India: Current Epidemiology and Future Directions

Cardiovascular diseases (CVDs) have now become the leading cause of mortality in India. A quarter of all mortality is attributable to CVD. Ischemic heart disease and stroke are the predominant causes and are responsible for >80% of CVD deaths. The Global Burden of Disease study estimate of age-standardized CVD death rate of 272 per 100 000 population in India is higher than the global average of 235 per 100 000 population. Some aspects of the CVD epidemic in India are causes of concern, including its accelerated build-up, the early age of disease onset in the population, and the high case fatality rate. In India, the epidemiological transition from predominantly infectious disease conditions to noncommunicable diseases has occurred over a rather brief period. Premature mortality in terms of years of life lost because of CVD in India increased by 59%, from 23.2 million (1990) to 37 million (2010). Despite wide heterogeneity in the prevalence of cardiovascular risk factors across different regions, CVD has emerged as the leading cause of death in all parts of India, including poorer states and rural areas. The progression of the epidemic is characterized by the reversal of socioeconomic gradients; tobacco use and low fruit and vegetable intake have become more prevalent among those from lower socioeconomic backgrounds. In addition, individuals from lower socioeconomic backgrounds frequently do not receive optimal therapy, leading to poorer outcomes. Countering the epidemic requires the development of strategies such as the formulation and effective implementation of evidence-based policy, reinforcement of health systems, and emphasis on prevention, early detection, and treatment with the use of both conventional and innovative techniques. Several ongoing community-based studies are testing these strategies.

2.3. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019:

Cardiovascular diseases (CVDs), principally ischemic heart disease (IHD) and stroke, are the leading cause of global mortality and a major contributor to disability. This paper reviews the magnitude of total CVD burden, including 13 underlying causes of cardiovascular death and 9 related risk factors, using estimates from the Global Burden of Disease (GBD) Study 2019. GBD, an ongoing multinational collaboration to provide comparable and consistent estimates of population health over time, used all available population-level data sources on incidence, prevalence, case fatality, mortality, and health risks to produce estimates for 204 countries and territories from 1990 to 2019. Prevalent cases of total CVD nearly doubled from 271 million (95% uncertainty interval [UI]: 257 to 285 million) in 1990 to 523 million (95% UI: 497 to 550 million) in 2019, and the number of CVD deaths steadily increased from 12.1 million (95% UI:11.4 to 12.6 million) in 1990, reaching 18.6 million (95% UI: 17.1 to 19.7 million) in 2019. The global trends for disability-adjusted life years (DALYs) and years of life lost also increased significantly, and years lived with disability doubled from 17.7 million (95% UI: 12.9 to 22.5 million) to 34.4 million (95% UI:24.9 to 43.6 million) over that period. The total number of DALYs due to IHD has risen steadily since 1990, reaching 182 million (95% UI: 170 to 194 million) DALYs, 9.14 million (95% UI: 8.40 to 9.74 million) deaths in the year 2019, and 197 million (95% UI: 178 to 220 million) prevalent cases of IHD in 2019. The total number of DALYs due to stroke has risen steadily since 1990, reaching 143 million (95% UI: 133 to 153 million) DALYs, 6.55 million (95% UI: 6.00 to 7.02 million) deaths in the year 2019, and 101 million (95% UI: 93.2 to 111 million) prevalent cases of stroke in 2019. Cardiovascular diseases remain the leading cause of disease burden in the world. CVD burden continues its decades-long rise for almost all countries outside high-income countries, and alarmingly, the age-standardized rate of CVD has begun to rise in some locations where it was previously declining in high-income countries. There is an urgent need to focus on implementing existing costeffective policies and interventions if the world is to meet the targets for Sustainable Development Goal 3 and achieve a 30% reduction in premature mortality due to noncommunicable diseases.

2.4. Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks

The World Health Organization (WHO) estimated one third of all global deaths reason as cardiovascular diseases in 2015. Some computational techniques were proposed for investigation of heart diseases. This study proposes a genetic algorithm (GA) based trained recurrent fuzzy neural networks (RFNN) to diagnosis of heart diseases. The University of California Irvine (UCI) Cleveland heart disease dataset is used in this study. Out of total 297 instances of patient data, 252 are used for training and 45 of them are chosen to be the testing. The results showed that 97.78% accuracy was obtained from testing set. In addition to the accuracy, root mean square error, the probability of the misclassification error, specificity, sensitivity, precision, and F-score are calculated. The results were found to be satisfying based on comparison.

2.5. Feature selection for medical diagnosis: Evaluation for cardiovascular diseases

Machine learning has emerged as an effective medical diagnostic support system. In a medical diagnosis problem, a set of features that are representative of all the variations of the disease are necessary. The objective of our work is to predict more accurately the presence of cardiovascular disease with reduced number of attributes. We investigate intelligent system to generate feature subset with improvement in diagnostic performance. Features ranked with distance measure are searched through forward inclusion, forward selection and backward elimination search techniques to find subset that gives improved classification result. We propose hybrid forward selection technique for cardiovascular disease diagnosis. Our experiment demonstrates that this approach finds smaller subsets and increases the accuracy of diagnosis compared to forward inclusion and backelimination techniques.

2.6. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques

Machine learning involves artificial intelligence, and it is used in solving many problems in data science. One common application of machine learning is the prediction of an outcome based upon existing data. The machine learns patterns from the existing dataset, and then applies them to an unknown dataset in order to predict the outcome. Classification is a powerful machine learning technique that is commonly used for prediction. Some classification algorithms predict with satisfactory accuracy, whereas others exhibit a limited accuracy. This paper investigates a method termed ensemble classification, which is used for improving the accuracy of weak algorithms by combining multiple classifiers. Experiments with this tool were performed using a heart disease dataset. A comparative analytical approach was done to determine how the ensemble technique can be applied for improving prediction accuracy in heart disease. The focus of this paper is not only on increasing the accuracy of weak classification algorithms, but also on the implementation of the algorithm with a medical dataset, to show its utility to predict disease at an early stage. The results of the study indicate that ensemble techniques, such as bagging and boosting, are effective in improving the prediction accuracy of weak classifiers, and exhibit satisfactory performance in identifying risk of heart disease. A maximum increase of 7% accuracy for weak classifiers was achieved with the help of ensemble classification. The performance of the process was further enhanced with a feature selection implementation, and the results showed significant improvement in prediction accuracy.

2.7. Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods

Objective: Machine learning (ML) and artificial intelligence (AI) are rapidly becoming the most talked about and controversial topics in radiology and medicine. Over the past few years, the numbers of ML- or AI-focused studies in the literature have increased almost exponentially, and ML has become a hot topic at academic and industry conferences. However, despite the increased awareness of ML as a tool, many medical professionals have a poor understanding of how ML works and how to critically appraise studies and tools that are presented to us. Thus, we present a brief overview of ML, explain the metrics used in ML and how to interpret them, and explain some of the technical jargon associated with the field so that readers with a medical background and basic knowledge of statistics can feel more comfortable when examining ML applications. Conclusion: Attention to sample size, overfitting, underfitting, cross validation, as well as a broad knowledge of the metrics of machine learning, can help those with little or no technical knowledge begin to assess machine learning studies. However, transparency in methods and sharing of algorithms is vital to allow clinicians to assess these tools themselves. Keywords: artificial intelligence; machine learning; medicine; supervised machine learning; unsupervised machine learning.

2.8. An Effective Heart Disease Prediction Model for a Clinical Decision Support System

Heart disease, one of the major causes of mortality worldwide, can be mitigated by early heart disease diagnosis. A clinical decision support system (CDSS) can be used to diagnose the subjects' heart disease status earlier. This study proposes an effective heart disease prediction model (HDPM) for a CDSS which consists of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to detect and eliminate the outliers, a hybrid Synthetic Minority Over-sampling Technique-Edited Nearest Neighbour (SMOTEENN) to balance the training data distribution and XGBoost to predict heart disease. Two publicly available datasets (Statlog and Cleveland) were used to build the model and compare the results with those of other models (naive bayes (NB), logistic regression (LR), multilayer perceptron (MLP), support vector machine (SVM), decision tree (DT), and random forest (RF)) and of previous study results. The results revealed that the proposed model outperformed other models and previous study results by achieving accuracies of 95.90% and 98.40% for Statlog and Cleveland datasets, respectively. In addition, we designed and developed the prototype of the Heart Disease CDSS (HDCDSS) to help doctors/clinicians diagnose the patients'/subjects' heart disease status based on their current condition. Therefore, early treatment could be conducted to prevent the deaths caused by late heart disease diagnosis.

2.9. Heart Disease Prediction Framework Based on Hybrid Classifiers and Genetic Algorithm

Supervised machine learning algorithms are powerful classification techniques commonly used to build prediction models that help diagnose the disease early. However, some challenges like overfitting and underfitting need to be overcome while building the model. This paper introduces hybrid classifiers using the ensembled model with a majority voting technique to improve prediction accuracy. Furthermore, a proposed pre-processing technique and features selection based on a genetic algorithm is suggested to enhance prediction performance and overall time consumption. In addition, the 10- folds cross-validation technique is used to overcome the overfitting problem. Experiments were performed on a dataset for cardiovascular patients from the UCI Machine Learning Repository. Through a comparative analytical approach, the study results indicated that the proposed ensemble classifier model achieved a classification accuracy of 98.18% higher than the rest of the relevant developments in the study.

2.10. A Case Study with Cardiovascular Disease Risk Prediction

Machine learning is often perceived as a sophisticated technology accessible only by highly trained experts. This prevents many physicians and biologists from using this tool in their research. The goal of this paper is to eliminate this out-dated perception. We argue that the recent development of auto machine learning techniques enables biomedical researchers to quickly build competitive machine learning classifiers without requiring in-depth knowledge about the underlying algorithms. We study the case of predicting the risk of cardiovascular diseases. To support our claim, we compare auto machine learning techniques against a graduate student using several important metrics, including the total amounts of time required for building machine learning models and the final classification accuracies on unseen test datasets. In particular, the graduate student manually builds multiple machine learning classifiers and tunes their parameters for one month using scikit-learn library, which is a popular machine learning library to obtain ones that perform best on two given, publicly available datasets. We run an auto machine learning library called auto-sklearn on the same datasets. Our experiments find that automatic machine learning takes 1 h to produce classifiers that perform better than the ones built by the graduate student in one month. More importantly, building this classifier only requires a few lines of standard code. Our findings are expected to change the way physicians see machine learning and encourage wide adoption of Artificial Intelligence (AI) techniques in clinical domains.

CHAPTER 3

SYSTEM STUDY

3.1. FEASIBILITY STUDY

- Availability of data: The availability of a large dataset of high-quality data related to heart disease is crucial for the development of a robust machine learning model. The feasibility study should assess the availability and accessibility of such data.
- Data pre-processing: The feasibility study should consider the complexity and feasibility of data pre-processing tasks such as cleaning, handling missing values, and feature selection.
- Model selection: The feasibility study should consider the availability of appropriate machine learning algorithms that can be used for heart disease prediction. The study should also evaluate the computational resources required for model training and optimization.
- Accuracy and Performance: The feasibility study should evaluate the accuracy and performance of the machine learning model in predicting heart disease. The study should compare the performance of the machine learning model with existing methods used in clinical practice.
- User acceptance: The feasibility study should assess the potential user acceptance and willingness to use the heart disease prediction system based on machine learning. It is essential to understand the user's perspective and evaluate the usability and user interface of the system.
- Regulatory and ethical considerations: The feasibility study should also consider the regulatory and ethical requirements for developing and deploying a heart disease prediction system based on machine learning. This includes compliance with data protection laws and patient confidentiality requirements.
- Cost-effectiveness: The feasibility study should assess the cost-effectiveness of developing and deploying a heart disease prediction system based on machine learning. This includes the cost of data acquisition, pre-processing, software development, and maintenance.

3.2. ECONOMIC FEASIBILITY

- Cost of data collection and pre-processing: The cost of data collection and pre-processing can be significant as it involves accessing and compiling large amounts of data from multiple sources. The feasibility study should consider the cost of acquiring and cleaning the data required for the development of the machine learning model.
- Cost of software development: The cost of software development includes the cost of hiring software developers, purchasing hardware and software tools, and developing the necessary infrastructure for deploying and maintaining the system.
- Cost of model training and optimization: The cost of model training and optimization includes the computational resources required for training the machine learning model, including hardware and software tools, and cloud-based services.
- Cost of deployment and maintenance: The cost of deploying the system includes hosting the system on a server, ensuring that the system is secure and scalable, and

maintaining the system over time. The feasibility study should consider the cost of maintaining and updating the system.

- Revenue potential: The revenue potential for a heart disease prediction system based on machine learning can come from multiple sources, such as licensing the technology to healthcare providers, partnering with insurers or selling the system directly to consumers. The feasibility study should evaluate the potential revenue sources and estimate the potential revenue generated by the system.
- Return on Investment (ROI): The feasibility study should consider the potential ROI for the heart disease prediction system. The ROI is the ratio of the net benefits to the costs of the project, and it should be calculated over the expected lifetime of the system.

3.3. TECHNICAL FEASIBILITY

- Availability and quality of data: The availability and quality of data related to heart disease is critical for the development of an accurate machine learning model. The feasibility study should consider the availability, accessibility, and quality of data sources, and evaluate if they are sufficient to build a reliable model.
- Data pre-processing: Data pre-processing involves cleaning, transforming, and selecting features from the data before feeding it to the machine learning algorithm. The feasibility study should consider the complexity of pre-processing tasks and determine if the available data can be pre-processed effectively to generate features that can be used to train a machine learning model.
- Model selection: The feasibility study should evaluate the suitability of different machine learning algorithms for heart disease prediction. The study should consider the complexity of the algorithms, their performance on similar datasets, and their suitability for real-time prediction.
- Hardware and software requirements: The feasibility study should evaluate the hardware and software requirements to run the machine learning model. This includes computing resources such as processing power, memory, and storage, as well as software tools required for data preprocessing and model development.
- Model training and optimization: The feasibility study should consider the computational resources and time required for training and optimizing the machine learning model. This includes assessing the scalability of the system, the time required to train the model, and the performance of the model as the dataset size increases.
- Performance evaluation: The feasibility study should evaluate the performance of the machine learning model in terms of accuracy, precision, recall, and other relevant metrics. The study should compare the performance of the machine learning model with existing methods used in clinical practice.

3.4. BEHAVIOURAL FEASIBILITY

- User acceptance: The feasibility study should assess the willingness of potential users to adopt and use the heart disease prediction system. This includes evaluating their attitudes towards using technology for healthcare purposes, their perceived usefulness of the system, and their willingness to share personal health data.
- User interface design: The user interface design should be intuitive and easy to use to encourage user adoption. The feasibility study should consider the usability of the system, including the layout, visual design, and ease of use.

- Privacy and security: The feasibility study should evaluate the privacy and security concerns of potential users. The study should assess the system's compliance with relevant data protection regulations and ensure that the user's personal health information is protected.
- Health literacy: The feasibility study should consider the health literacy level of potential users. The system should be designed with clear and concise language and avoid using medical jargon to ensure that users can understand the information provided.
- User training: The feasibility study should evaluate the need for user training to ensure that users can use the system effectively. This includes assessing the level of technical knowledge required to use the system and identifying potential barriers to adoption.
- Provider buy-in: The feasibility study should evaluate the willingness of healthcare providers to use the heart disease prediction system. Providers should be willing to integrate the system into their clinical workflows and should be willing to use the system to make informed decisions about patient care.

CHAPTER 4
PROBLEM DEFINITION

4.1. EXISTING SYSTEM

Heart disease is even being highlighted as a silent killer which leads to the death of a person without obvious symptoms. The nature of the disease is the cause of growing anxiety about the disease & its consequences. Hence continued efforts are being done to predict the possibility of this deadly disease in prior. So that various tools & techniques are regularly being experimented with to suit the present-day health needs. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can conclude. This technique can be very well adapted to the do the prediction of heart disease. As the well-known quote says "Prevention is better than cure", early prediction & its control can be helpful to prevent & decrease the death rates due to heart disease.

4.2. PROPOSED SYSTEM

The working of the system starts with the collection of data and selecting the important attributes. Then the required data is pre-processed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data. The heart disease prediction can be carried out using various algorithm such as decision tree and random forest algorithm. It can perform both regression and classification tasks. It produces good predictions that can be understood easily. It can handle large datasets efficiently.

Random Forest is a supervised learning algorithm. It is an extension of machine learning classifiers which include the bagging to improve the performance of Decision Tree. It combines tree predictors, and trees are dependent on a random vector which is independently sampled. The distribution of all trees are the same. Random Forests splits nodes using the best among of a predictor subset that are randomly chosen from the node itself, instead of splitting nodes based on the variables. The time complexity of the worst case of learning with Random Forests is $O(M(dn\log n))$, where M is the number of growing trees, n is the number of instances, and d is the data dimension

CHAPTER 5

SYSTEM REQUIREMENT

5.1. HARDWARE REQUIREMENT

Processer	:	Any Update Processer
RAM	:	Min 4GB
Hard Disk	:	Min 100GB

5.2. SOFTWARE REQUIREMENT

Operating System	:	Windows family
Technology	:	Python3.9, Django
IDE	:	Jupyter notebook
Code Editor	:	VS code

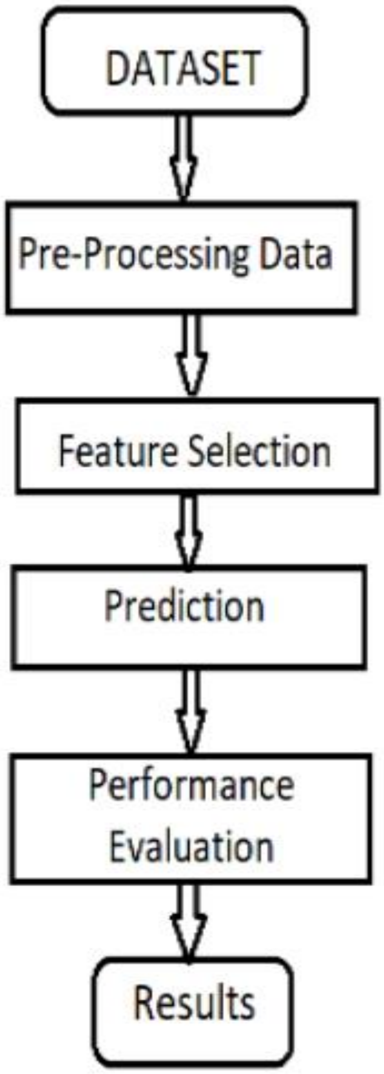
5.3. DESCRIPTION OF TECHNOLOGY

Django:

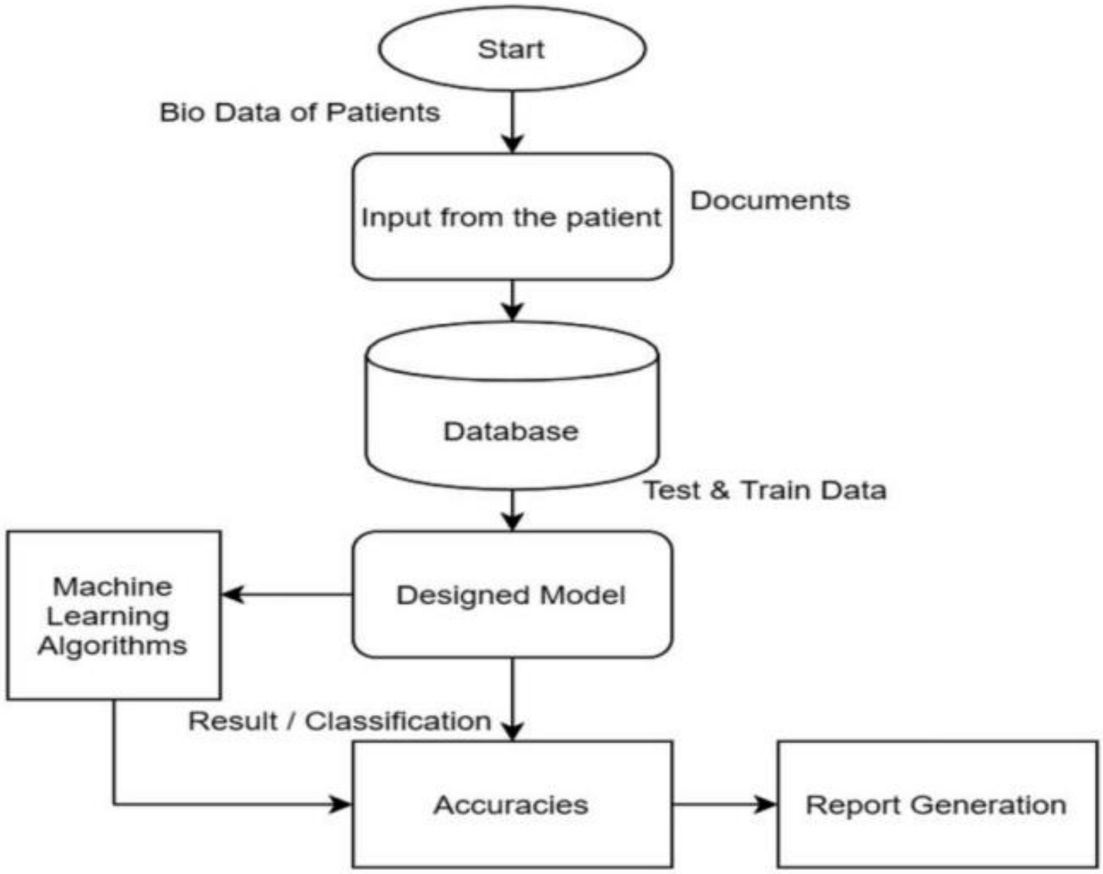
Django is a high-level, open-source Python web framework that follows the Model-View-Controller (MVC) architectural pattern. It was created to help developers build web applications quickly and efficiently by providing a lot of built-in functionality and a consistent structure for organizing code. Django includes many features out of the box, such as an ORM (Object-Relational Mapping) for database access, URL routing, template rendering, form handling, and user authentication. It also has a robust administrative interface that makes it easy to manage data and user accounts. Django's modularity and extensibility make it highly customizable, and it integrates well with other Python libraries and frameworks. It also has a large and active community that provides support, documentation, and third-party packages.

CHAPTER 6
SYSTEM DESIGN

6.1. ARCHITECTURE DIAGRAM



6.2. FLOW DIAGRAM



CHAPTER 7

SYSTEM IMPLEMENTATION

7.1. MODULES DESCRIPTION

The working of the system starts with the collection of data and selecting the important attributes. Then the required data is pre-processed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data. This system is implemented using the following modules.

1) Collection of Dataset
2) Selection of attributes
3) Data Pre-Processing
4) Balancing of Data
5) Disease Prediction

7.1.1. Collection of dataset

Initially, we collect a dataset for our heart disease prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model. For this project, 70% of training data is used and 30% of data is used for testing. The dataset used for this project is Heart Disease UCI. The dataset consists of 76 attributes; out of which, 14 attributes are used for the system.

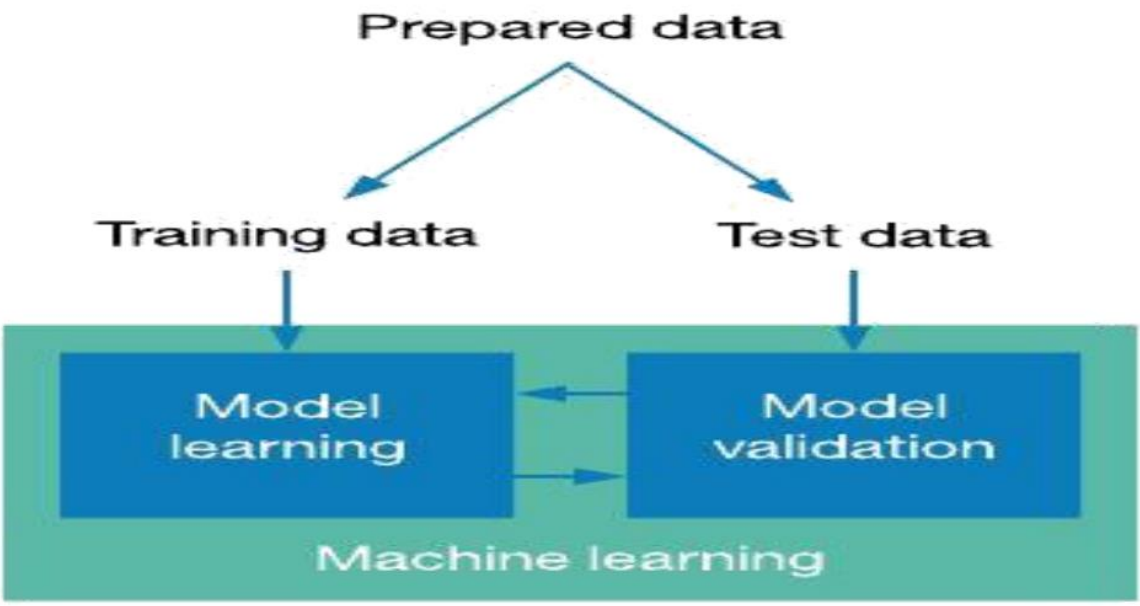


Figure: Collection of Data

7.1.2 Selection of attributes

Attribute or Feature selection includes the selection of appropriate attributes for the prediction system. This is used to increase the efficiency of the system. Various attributes of the patient like gender, chest pain type, fasting blood pressure, serum cholesterol, exang, etc are selected for the prediction. The Correlation matrix is used for attribute selection for this model.

Figure: Correlation matrix

7.1.3 Pre-processing of Data

Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, and missing values of the dataset. Data pre-processing has the activities like importing datasets, splitting datasets, attribute scaling, etc. Preprocessing of data is required for improving the accuracy of the model.

Figure: Data Pre-processing

7.1.4 Balancing of Data

Imbalanced datasets can be balanced in two ways. They are Under Sampling and Over Sampling

(a) Under Sampling: In Under Sampling, dataset balance is done by the reduction of the size of the sample class. This process is considered when the amount of data is adequate.

(b) Over Sampling: In Over Sampling, dataset balance is done by increasing the size of the scarce samples. This process is considered when the amount of data is inadequate.

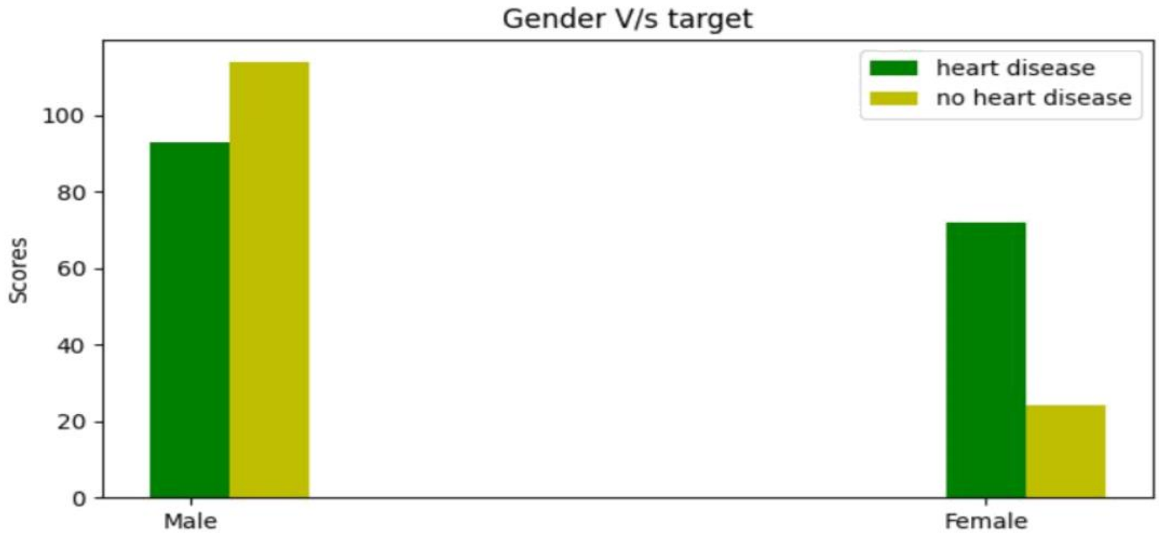


Figure: Data Balancing

7.1.5 Prediction of Disease

Various machine learning Random Tree used for classification. Comparative analysis is performed among algorithms and the algorithm that gives the highest accuracy is used for heart disease prediction.

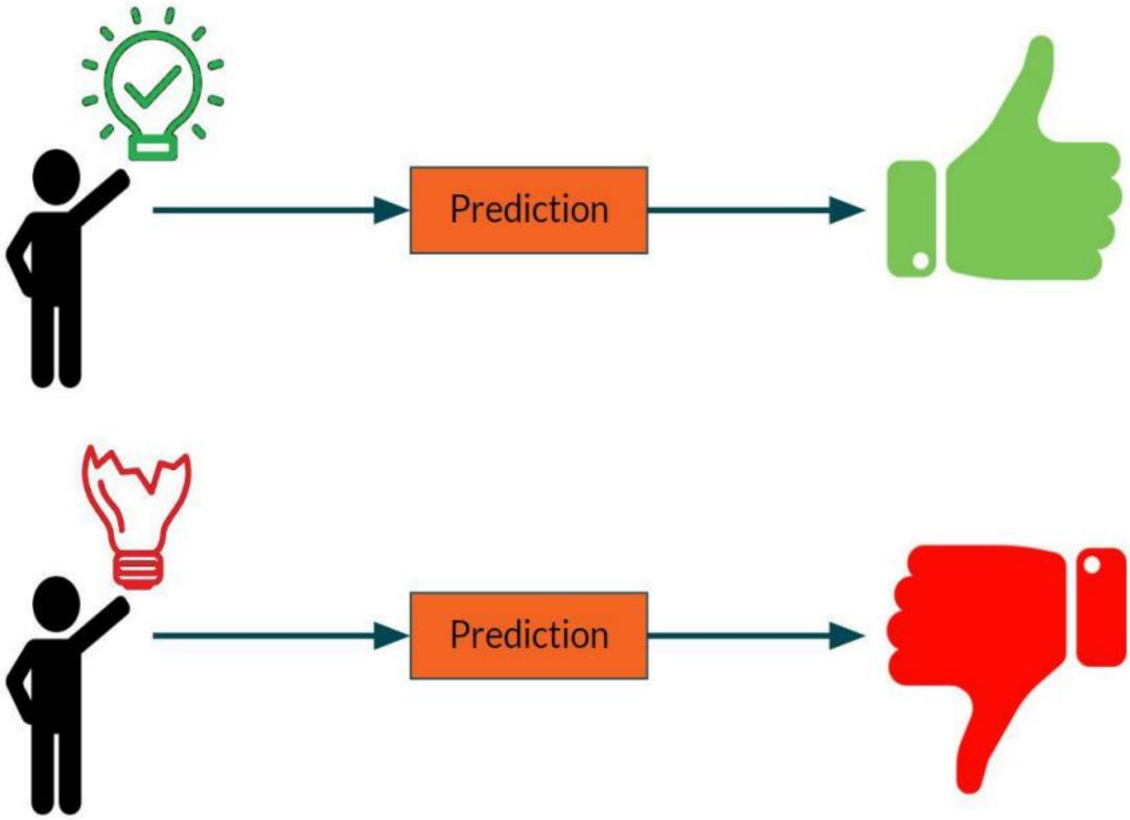


Figure: Prediction of Disease

CHAPTER 8

SYSTEM TESTING

8.1. UNIT TESTING

Unit testing for heart disease prediction using a random forest algorithm can involve several steps to ensure that individual components of the system are functioning as expected. Here are some general steps that could be taken:

1. Test Data Selection: Selecting appropriate test data is crucial to ensure that the unit test evaluates the functionality of the individual components of the system.
2. Test Setup: Set up the environment for testing, including the necessary software packages, libraries, and dependencies, and ensure that the testing environment is like the production environment.
3. Unit Test Design: Design unit tests for individual components of the system, including functions that perform data pre-processing, feature engineering, model loading, and model predictions.
4. Test Execution: Execute the unit tests on the selected test data, and observe the results.
5. Evaluation Metrics: Evaluate the test results using appropriate evaluation metrics, such as accuracy, precision, recall, and F1 score. These metrics will help to determine how well the individual components of the system are performing.
6. Debugging and Error Handling: Debug any issues that arise during unit testing and implement error handling procedures to ensure that the system can handle unexpected errors.
7. Documentation: Document the unit testing process, including the test data, environment setup, evaluation metrics, and any issues encountered during testing.

8.2. INTEGRATION TESTING

Integration testing for heart disease prediction using a random forest algorithm can involve several steps to ensure that the system's components work together as expected. Here are some general steps that could be taken:

1. Test Data Selection: Selecting appropriate test data is crucial to ensure that the integration test evaluates the functionality of the system's components working together.
2. Test Setup: Set up the environment for testing, including the necessary software packages, libraries, and dependencies, and ensure that the testing environment is like the production environment.
3. Integration Test Design: Design integration tests for the entire system, including the process of data pre-processing, feature engineering, model loading, and model predictions.
4. Test Execution: Execute the integration tests on the selected test data and observe the results.
5. Evaluation Metrics: Evaluate the test results using appropriate evaluation metrics, such as accuracy, precision, recall, and F1 score. These metrics will help to determine how well the system is performing.

6. Debugging and Error Handling: Debug any issues that arise during integration testing and implement error handling procedures to ensure that the system can handle unexpected errors.
7. Documentation: Document the integration testing process, including the test data, environment setup, evaluation metrics, and any issues encountered during testing.

8.3. WHITE BOX TESTING

White box testing for heart disease prediction using a random forest algorithm involves testing the internal workings of the system's code. Here are some general steps that could be taken:

1. Code Analysis: Perform a thorough analysis of the system's source code to understand its internal workings, including the data pre-processing, feature engineering, model training, and prediction steps.
2. Test Case Design: Design test cases that target specific parts of the code and ensure that each component is tested thoroughly.
3. Test Execution: Execute the test cases on the system's code, and observe the results.
4. Evaluation Metrics: Evaluate the test results using appropriate evaluation metrics, such as accuracy, precision, recall, and F1 score. These metrics will help to determine how well the individual components of the system are performing.
5. Debugging and Error Handling: Debug any issues that arise during white box testing and implement error handling procedures to ensure that the system can handle unexpected errors.
6. Code Coverage Analysis: Analyse the code coverage to ensure that all parts of the code are tested.
7. Documentation: Document the white box testing process, including the test cases, evaluation metrics, and any issues encountered during testing.

8.4. BLACK BOX TESTING

Black box testing for heart disease prediction using a random forest algorithm involves testing the system without any knowledge of its internal workings. Here are some general steps that could be taken:

1. Test Data Selection: Selecting appropriate test data is crucial to ensure that the black box test evaluates the functionality of the entire system.
2. Test Setup: Set up the environment for testing, including the necessary software packages, libraries, and dependencies, and ensure that the testing environment is like the production environment.
3. Test Case Design: Design test cases that cover a range of input scenarios and expected outputs, ensuring that the system is tested under different conditions.
4. Test Execution: Execute the test cases on the system, and observe the results.
5. Evaluation Metrics: Evaluate the test results using appropriate evaluation metrics, such as accuracy, precision, recall, and F1 score. These metrics will help to determine how well the system is performing as a whole.
6. Debugging and Error Handling: Debug any issues that arise during black box testing and implement error handling procedures to ensure that the system can handle unexpected errors.

7. Documentation: Document the black box testing process, including the test data, environment setup, evaluation metrics, and any issues encountered during testing.

8.5. ACCEPTANCE TESTING

Acceptance testing for heart disease prediction using a random forest algorithm involves testing the system's functionality in a real-world scenario. Here are some general steps that could be taken:

1. Test Data Selection: Selecting appropriate test data is crucial to ensure that the acceptance test evaluates the functionality of the system in a real-world scenario.
2. Test Setup: Set up the environment for testing, including the necessary software packages, libraries, and dependencies, and ensure that the testing environment is like the production environment.
3. User Acceptance Test Design: Design acceptance tests that are specific to the user's needs and requirements. These tests should be designed to ensure that the system meets the user's expectations and requirements.
4. Test Execution: Execute the acceptance tests on the system and observe the results.
5. Evaluation Metrics: Evaluate the test results using appropriate evaluation metrics, such as accuracy, precision, recall, and F1 score. These metrics will help to determine how well the system is performing in a real-world scenario.
6. Debugging and Error Handling: Debug any issues that arise during acceptance testing and implement error handling procedures to ensure that the system can handle unexpected errors.
7. Documentation: Document the acceptance testing process, including the test data, environment setup, evaluation metrics, and any issues encountered during testing.

CHAPTER 9

RESULT AND DISCUSSION

Result

Accuracy is the number of correctly predicted instances divided by the total number of instances. Precision is the number of true positive instances divided by the total number of instances predicted as positive. Recall is the number of true positive instances divided by the total number of actual positive instances. The F1 score is the harmonic mean of precision and recall.

A high accuracy, precision, recall, and F1 score indicate that the random forest algorithm is performing well for heart disease prediction. However, the specific performance of the algorithm can vary depending on the dataset used, the pre-processing and feature engineering techniques employed, the hyperparameters chosen, and the evaluation metrics used. It is important to conduct rigorous testing and validation to ensure that the algorithm is performing optimally for the specific use case.

After performing the machine learning approach for training and testing we find that accuracy of the Random Forest is better compared to other algorithms. Accuracy is calculated with the support of the confusion matrix of each algorithm, using the equation of accuracy, value has been calculated and it is concluded that extreme gradient boosting is best with 96% accuracy.

CHAPTER 10

CONCLUSION AND FUTURE ENHANCEMENT

Conclusion

Heart diseases are a major killer in India and throughout the world, application of promising technology like machine learning to the initial prediction of heart diseases will have a profound impact on society. The early prognosis of heart disease can aid in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine. The number of people facing heart diseases is on a raise each year. This prompts for its early diagnosis and treatment. The utilization of suitable technology support in this regard can prove to be highly beneficial to the medical fraternity and patients. In this paper, the machine learning algorithms used to measure the performance is Random Forest applied on the dataset.

The expected attributes leading to heart disease in patients are available in the dataset which contains 76 features and 14 important features that are useful to evaluate the system are selected among them. If all the features taken into the consideration, then the efficiency of the system the author gets is less. To increase efficiency, attribute selection is done. In this n features must be selected for evaluating the model which gives more accuracy. The correlation of some features in the dataset is almost equal and so they are removed. If all the attributes present in the dataset are considered, then the efficiency decreases considerably.

The machine learning methods accuracy are compared based on which one prediction model is generated. Hence, the aim is to use various evaluation metrics like confusion matrix, accuracy, precision, recall, and f1-score which predicts the disease efficiently. Random Forest algorithm gives the highest accuracy of 96%.

CHAPTER 11

APPENDIX

11.1. Sample Code

BASIC:

```
#Basic
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import os

#Model
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.tree import export_graphviz #plot tree

#Model Evaluation
from sklearn.metrics import roc_curve, auc
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix

import eli5 #for attribute importance
from eli5.sklearn import PermutationImportance
```

PREPROCESSING:

```
#Name columns
data.columns = ['age', 'sex', 'chest_pain_type', 'resting_blood_pressure',
'cholesterol', 'fasting_blood_sugar', 'rest_ecg', 'max_heart_rate_achieved',
       'exercise_induced_angina', 'st_depression', 'st_slope',
'num_major_vessels', 'thalassemia', 'target']
#Conversion of data type
data['sex'][data['sex'] == 0] = 'female'
data['sex'][data['sex'] == 1] = 'male'

data['chest_pain_type'][data['chest_pain_type'] == 1] = 'typical angina'
data['chest_pain_type'][data['chest_pain_type'] == 2] = 'atypical angina'
data['chest_pain_type'][data['chest_pain_type'] == 3] = 'non-anginal pain'
data['chest_pain_type'][data['chest_pain_type'] == 4] = 'asymptomatic'

data['fasting_blood_sugar'][data['fasting_blood_sugar'] == 0] = 'lower than
120mg/ml'
```

```python
data['fasting_blood_sugar'][data['fasting_blood_sugar'] == 1] = 'greater than 120mg/ml'

data['rest_ecg'][data['rest_ecg'] == 0] = 'normal'
data['rest_ecg'][data['rest_ecg'] == 1] = 'ST-T wave abnormality'
data['rest_ecg'][data['rest_ecg'] == 2] = 'left ventricular hypertrophy'

data['exercise_induced_angina'][data['exercise_induced_angina'] == 0] = 'no'
data['exercise_induced_angina'][data['exercise_induced_angina'] == 1] = 'yes'

data['st_slope'][data['st_slope'] == 1] = 'upsloping'
data['st_slope'][data['st_slope'] == 2] = 'flat'
data['st_slope'][data['st_slope'] == 3] = 'downsloping'

data['thalassemia'][data['thalassemia'] == 1] = 'normal'
data['thalassemia'][data['thalassemia'] == 2] = 'fixed defect'
data['thalassemia'][data['thalassemia'] == 3] = 'reversable defect'
data['sex'] = data['sex'].astype('object')
data['chest_pain_type'] = data['chest_pain_type'].astype('object')
data['fasting_blood_sugar'] = data['fasting_blood_sugar'].astype('object')
data['rest_ecg'] = data['rest_ecg'].astype('object')
data['exercise_induced_angina'] =
data['exercise_induced_angina'].astype('object')
data['st_slope'] = data['st_slope'].astype('object')
data['thalassemia'] = data['thalassemia'].astype('object')
data.head()
#Creation of labels, etc.
labels = data['target']

data = data.drop('target', axis = 1)
#Using Dummy variables to convert categorical to numeric
data = pd.get_dummies(data, drop_first=True)
data.head()
```

MODEL:

```python
#Split Data
x_train, x_test, y_train, y_test = train_test_split(data, labels, test_size =
0.2, random_state = 56)

print("Shape of x_train :", x_train.shape)
print("Shape of x_test :", x_test.shape)
print("Shape of y_train :", y_train.shape)
print("Shape of y_test :", y_test.shape)
#Run Model
model = RandomForestClassifier(max_depth = 5)
model.fit(x_train, y_train)
y_predict = model.predict(x_test)
```

```
y_pred_quant = model.predict_proba(x_test)[:, 1] #store predicted probabilites
y_pred = model.predict(x_test)
```

EVALUATION:

```
#Accuracy
print("Training Accuracy :", model.score(x_train, y_train))
print("Testing Accuracy :", model.score(x_test, y_test))
#Confusion Matrix
cm = confusion_matrix(y_test, y_pred)
plt.rcParams['figure.figsize'] = (5, 5)
sns.heatmap(cm, annot = True, annot_kws = {'size':15}, cmap = 'PuBu')
plt.title('Confusion Matrix for Decision Tree Model', y = 1.1)
plt.ylabel('Actual Label')
plt.xlabel('Predicted Label')
plt.show()
#Sensitivity and Specificity
total=sum(sum(cm))

sensitivity = cm[0,0]/(cm[0,0]+cm[1,0])
print('Sensitivity : ', sensitivity )

specificity = cm[1,1]/(cm[1,1]+cm[0,1])
print('Specificity : ', specificity)
#Receiver Operating Characteristic Curve
falsepr, truepr, thresholds = roc_curve(y_test, y_pred_quant) #Create true and
false positive rates

fig, ax = plt.subplots()
ax.plot(falsepr, truepr)
ax.plot([0, 1], [0, 1], transform=ax.transAxes, ls="-", c=".3")
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.0])

plt.rcParams['figure.figsize'] = (15, 5)
plt.title('ROC curve for diabetes classifier', fontweight = 30)
plt.xlabel('False Positive Rate (1 - Specificity)')
plt.ylabel('True Positive Rate (Sensitivity)')
plt.show()

#Area Under Curve
auc = auc(falsepr, truepr)
print("AUC Score :", auc)
```

HOME.HTML:

```
{% load static %}
<style>
    * {
        margin: 0;
    }
    .header {
        background-image: url("static/bg.jpg");
        height: 100%;
        background-position: center;
        background-repeat: no-repeat;
        background-size: cover;
        background-attachment: fixed;
    }

    body {
        background-color: #a0c0e8;
    }

    .section h1{
        text-align: center;
        font-size: 50px;
        padding: 50px;
    }

    .section p{
        text-align: center;
        padding: 50px;
        font-size: 20px;
    }

    .item form{
        margin: 100px;
        background-color: aliceblue;
        padding: 30px;
        border-radius: 50px;
        text-align: center;
    }

    h3 {
        padding: 20px;
    }

    input[type=number]{
        width: 10%;
        height: 5%;
        font-size: 15px;
        text-align: center;
```

```css
    border-radius: 10px;
}

select{
    width: 10%;
    height: 5%;
    font-size: 15px;
    text-align: center;
    border-radius: 10px;
}

button{
    background-color: lightblue;
    color: black;
    border: 2px solid black;
    padding: 16px 60px;
    text-align: center;
    text-decoration: none;
    display: inline-block;
    font-size: 20px;
    transition-duration: 0.3s;
    cursor: pointer;
    border-radius: 10px;
    margin: 3%;
}

button:hover{
    background-color: black;
    color: white;
}

.footer{
    width: 100%;
    height: 31.1%;
    background-color:darkblue;
    color: white;
    text-align: center;
}

.footer div{
    padding: 65px;
    font-family: Arial, Helvetica, sans-serif;
}

.footer div h5{
    padding-top: 30px;
    padding-bottom: 10px;
}
```

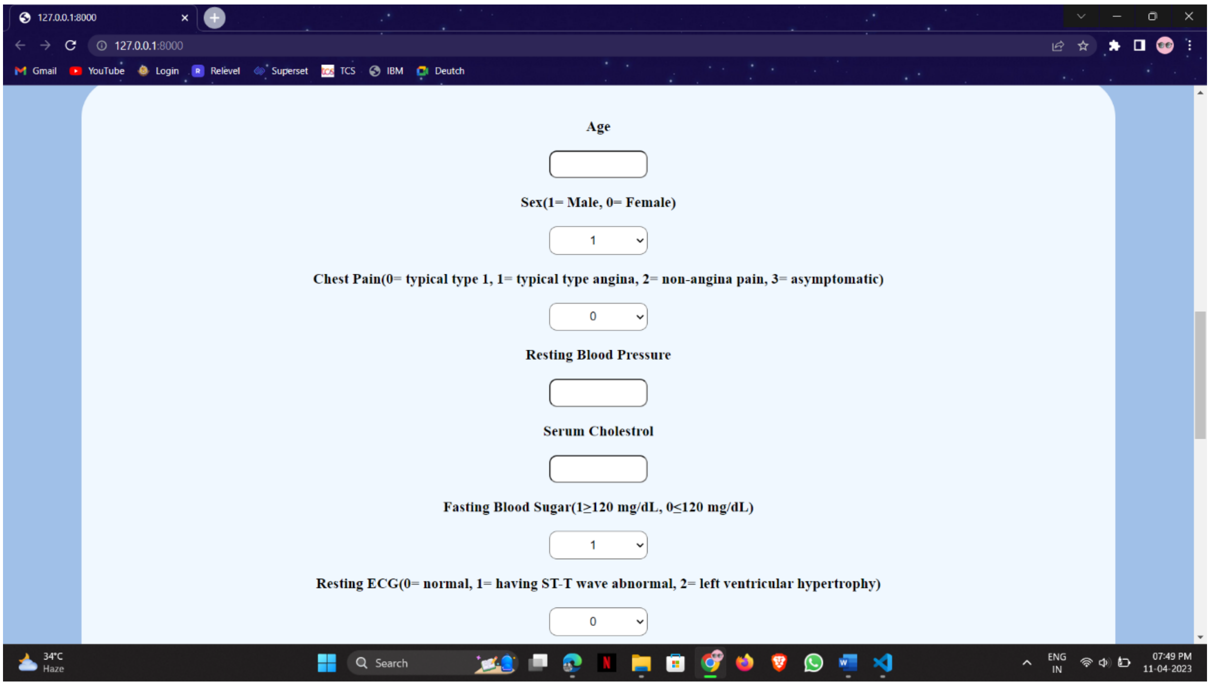
```
</style>
<body>
    <div class="header"></div>
    <div class="section">
        <h1>Predict your chance of having a heart disease because prevention
is better than cure!</h1>
        <center>_____________</center>
        <p>Enter the appropriate values of symptoms you face. Get the chances
of you contracting heart disease based on those values.</p>
        <center>_____________</center>
        <div class="item">
            <form method="POST">
                {% csrf_token %}
                <h3>Age</h3>
                <input name="age" type="number"/>
                <h3>Sex(1= Male, 0= Female)</h3>
                <select name="sex">
                    <option value="1">1</option>
                    <option value="0">0</option>
                </select>
                <h3>Chest Pain(0= typical type 1, 1= typical type angina, 2=
non-angina pain, 3= asymptomatic)</h3>
                <select name="cp">
                    <option value="0">0</option>
                    <option value="1">1</option>
                    <option value="2">2</option>
                    <option value="3">3</option>
                </select>
                <h3>Resting Blood Pressure</h3>
                <input name="trestbps" type="number"/>
                <h3>Serum Cholestrol</h3>
                <input name="chol" type="number"/>
                <h3>Fasting Blood Sugar(1≥120 mg/dL, 0≤120 mg/dL)</h3>
                <select name="fbs">
                    <option value="1">1</option>
                    <option value="0">0</option>
                </select>
                <h3>Resting ECG(0= normal, 1= having ST-T wave abnormal, 2=
left ventricular hypertrophy)</h3>
                <select name="restecg">
                    <option value="0">0</option>
                    <option value="1">1</option>
                    <option value="2">2</option>
                </select>
                <h3>Max heart rate achieved</h3>
                <input name="thalach" type="number"/>
                <h3>Exercise induced angina(0= no, 1= yes)</h3>
                <select name="exang">
```

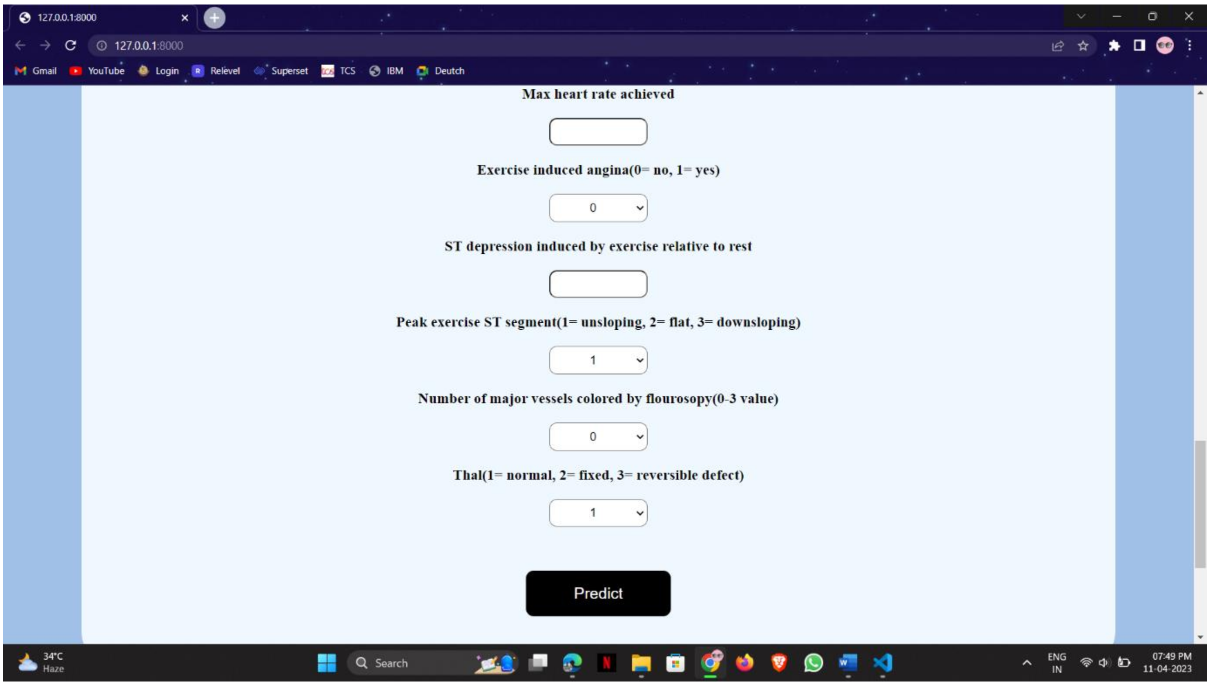
```
                <option value="0">0</option>
                <option value="1">1</option>
            </select>
            <h3>ST depression induced by exercise relative to rest</h3>
            <input name="oldpeak" type="number"/>
            <h3>Peak exercise ST segment(1= unsloping, 2= flat, 3=
downsloping)</h3>
            <select name="slope">
                <option value="1">1</option>
                <option value="2">2</option>
                <option value="3">3</option>
            </select>
            <h3>Number of major vessels colored by flourosopy(0-3
value)</h3>
            <select name="ca">
                <option value="0">0</option>
                <option value="1">1</option>
                <option value="2">2</option>
                <option value="3">3</option>
            </select>
            <h3>Thal(1= normal, 2= fixed, 3= reversible defect)</h3>
            <select name="thal">
                <option value="1">1</option>
                <option value="2">2</option>
                <option value="3">3</option>
            </select>
            <br><br>
            <button type="submit">Predict</button>
        </form>
    </div>
    <div class="footer">
        <div>
            <h4>&copy; 2023 All Rights Reserved.</h4>
            <h5>Designed & Developed by <h4>Mithun Raj S, Mohamed Afzal M,
Logesh Raj R</h4></h5>
        </div>
    </div>
</body>
```

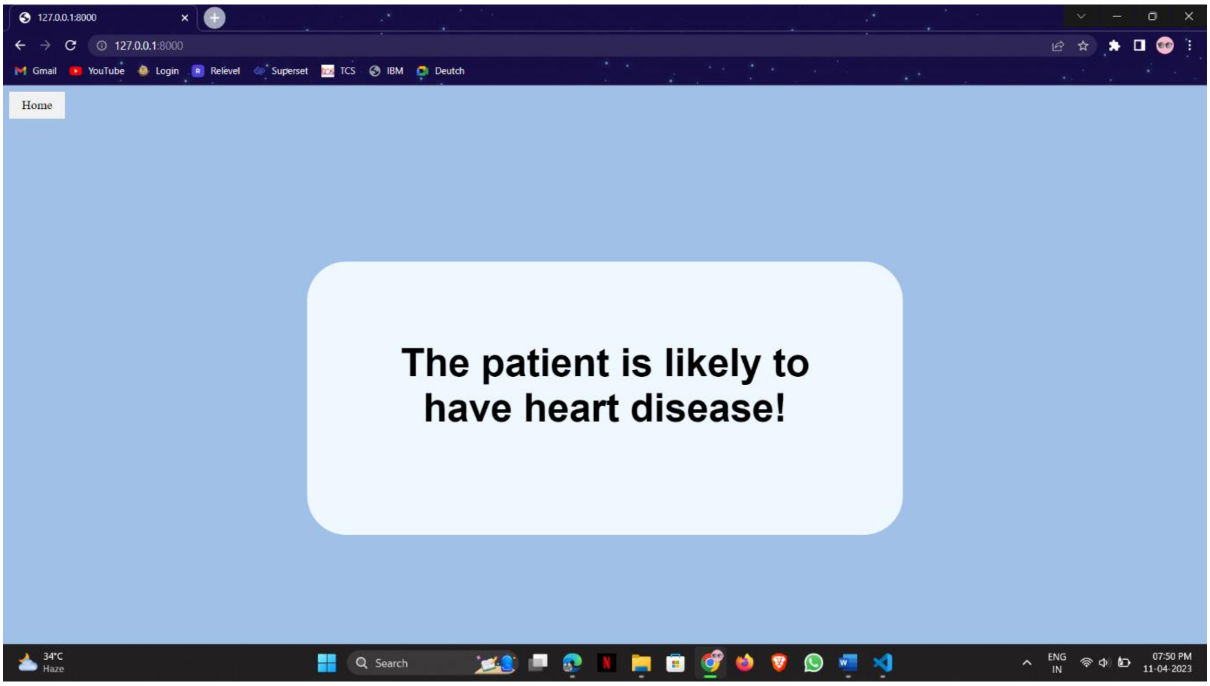
RESULT.HTML:

```html
<style>
    body {
        background-color: #a0c0e8;
    }

    .result{
        width: 50%;
        height: 50%;
        text-align: center;
        background-color: aliceblue;
        margin: auto;
        margin-top: 12%;
        border-radius: 50px;
    }

    .result h1{
        font-size: 50px;
        font-family: Arial, Helvetica, sans-serif;
        padding: 100px;
    }

    a {
        text-decoration: none;
        display: inline-block;
        padding: 8px 16px;
    }

    a:hover {
        background-color: rgb(106, 122, 207);
        color: white;
    }

    .home {
        background-color: #f1f1f1;
        color: black;
    }
</style>
<a href="{% url 'home' %}" class="home">Home</a>
<div class="result">
    {% if res == 1 %}
    <h1>The patient is likely to have heart disease!</h1>
    {% else %}
    <h1>The patient is not likely to have heart disease!</h1>
    {% endif %}
</div>
```

11.2. Screen Shots




Max heart rate achieved
Exercise induced angina(0= no, 1= yes)
ST depression induced by exercise relative to rest
Peak exercise ST segment(1= unsloping, 2= flat, 3= downsloping)
Number of major vessels colored by flourosopy(0-3 value)
Thal(1= normal, 2= fixed, 3= reversible defect)
Predict


Home
The patient is likely to have heart disease!

REFERENCE

Reference:

[1] M. Heron, "Deaths: Leading causes for 2010," Natl. Vital Stat. Rep., vol. 62, no. 6, pp. 1–96, 2013.

[2] Cardiovascular Diseases (CVDs). Accessed: May 9, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovasculardiseases-(cvds)

[3] S. Chauhan and B. T. Aeri, "The rising incidence of cardiovascular diseases in India: Assessing its economic impact," J. Preventive Cardiol., vol. 4, no. 5, pp. 735–740, 2015.

[4] G. A. Roth, G. A. Mensah, and C. O. Johnson, "Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study," J. Amer. College Cardiol., vol. 76, no. 25, pp. 2982–3021, Dec. 2020.

[5] K. Uyar and A. Ilhan, "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks," Proc. Comput. Sci., vol. 120, pp. 588–593, Jan. 2017.

[6] S. Shilaskar and A. Ghatol, "Feature selection for medical diagnosis: Evaluation for cardiovascular diseases," Expert Syst. Appl., vol. 40, no. 10, pp. 4146–4153, Aug. 2013.

[7] S. Ismaeel, A. Miri, and D. Chourishi, "Using the extreme learning machine (ELM) technique for heart disease diagnosis," in Proc. IEEE Canada Int. Humanitarian Technol. Conf. (IHTC), May 2015, pp. 1–3.

[8] C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," Informat. Med. Unlocked, vol. 16, Jan. 2019, Art. no. 100203.

[9] G. S. Handelman, H. K. Kok, R. V. Chandra, A. H. Razavi, S. Huang, M. Brooks, M. J. Lee, and H. Asadi, "Peering into the black box of artificial intelligence: Evaluation metrics of machine learning methods," Amer. J. Roentgenol., vol. 212, no. 1, pp. 38–43, Jan. 2019.

[10] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDPM: An effective heart disease prediction model for a clinical decision support system," IEEE Access, vol. 8, pp. 133034–133050, 2020.

[11] K. Saxena and U. Banodha, "A fuzzy logic based cardiovascular disease risk level prediction system in correlation to diabetes and smoking," in Data Management, Analytics and Innovation (Advances in Intelligent Systems and Computing), vol. 1042. Singapore: Springer, 2020, pp. 29–40.

[12] M. Padmanabhan, P. Yuan, G. Chada, and H. V. Nguyen, "Physicianfriendly machine learning: A case study with cardiovascular disease risk prediction," J. Clin. Med., vol. 8, no. 7, p. 1050, Jul. 2019.

[13] S. D. Desai, S. Giraddi, P. Narayankar, N. R. Pudakalakatti, and S. Sulegaon, "Back-propagation neural network versus logistic regression in heart disease classification," in Advanced Computing and Communication Technologies (Advances in Intelligent Systems and Computing), vol. 702. Singapore: Springer, 2019, pp. 133–144.

[14] S. Islam, N. Jahan, and M. E. Khatun, ‘‘Cardiovascular disease forecast using machine learning paradigms,’’ in Proc. 4th Int. Conf. Comput. Methodol. Commun. (ICCMC), Mar. 2020, pp. 487–490.

[15] F. Z. Abdeldjouad, M. Brahami, and N. Matta, ‘‘A hybrid approach for heart disease diagnosis and prediction using machine learning techniques,’’ in Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Lecture Notes in Computer Science), vol. 12157. Cham, Switzerland: Springer, 2020, pp. 299–306.

Cos & POs MAPPING

COURSE OUTCOMES

1. Design/Develop sustainable solutions for societal issues with environmental considerations applying the basic engineering knowledge. [K6-Create]
2. Analyse and review research literature to synthesize research methods including design of experiments to provide valid conclusion. [K4-Analyze]
3. Utilize the new tools, algorithms, techniques to provide valid conclusion following the norms of engineering practice. [K3-Apply]
4. Test and evaluate the performance of the developed solution using appropriate techniques and tools. [K5-Evaluate]
5. Apply management principles to function effectively in the project team for project execution. [A4-Organize]
6. Engage in learning for effective project implementation in the broadest context of technological change with consideration for public health, safety, cultural and societal needs. [A3-Value]
7. Write effective reports and make clear presentation to the engineering community and society. [A2-Respond]

MAPPING OF COS WITH POS

Course Outcomes	Program Outcomes (POs)												PSOs	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**
CO 1	3		3				3						3	3
CO 2		3		3									3	3
CO 3					3			3					3	3
CO 4		3			3								3	3
CO 5									3		3		3	3
CO 6						3	3					3	3	3
CO 7										3			3	3
	3	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**

LIST OF PUBLICATION

Logesh Raj, Mithun Raj, Mohamed Afzal," Heart Disease Prediction using Random Forest Algorithm", Journal of Computing and Information Science in Engineering, Volume 23, paper number JCISE-23-1184.