A MINI PROJECT REPORT

ON

Heart Disease Prediction Using Machine Learning

Submitted in partial fulfillment of the requirement for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

BY

B. GNANIKA - 21P61A0519
CH. SHIVANI - 21P61A0552
D.AMULYA - 21P61A0564

Under the esteemed guidance of

Dr. Dara Raju
B. Tech (CSE), M. Tech., Ph. D.
Professor & Head of Department

Counselling Code : VBIT

VIGNANA BHARATHI
Institute of Technology

(A UGC Autonomous Institution, Approved by AICTE, Accredited by NBA & NAAC-A Grade, Affiliated to JNTUH)

(An Autonomous institution, Accredited by NBA and NAAC, New Delhi)
AUSHAPUR (V), GHATKESAR (M), MEDCHAL.DIST-501 301

AUSHAPUR (V), GHATKESAR (M), MEDCHAL.DIST-501 301

Department of Computer Science & Engineering

CERTIFICATE

This is to certify that the mini project titled "Heart Disease Prediction Using Machine Learning Submitted by 21P61A0519 B. Gnanika, 21P61A0552 Ch. Shivani,21P61A0564 D. Amulya B. tech IV-I semester Computer Science &Engineering is a record of the Bonafide work carried out by them

The Design embodied in this report have not been submitted to any other University for the award of any degree.

Internal Guide	**Head of the Department**
Dr. Dara Raju	**Dr. Dara Raju**
Professor & HOD	**Professor& HOD**

EXTERNAL EXAMINER

AUSHAPUR (V), GHATKESAR (MEDCHAL.DIST-501301)

Department of Computer Science

DECLARATION

We, B. GNANIKA bearing hall ticket number 21P61A0519, CH. SHIVANI bearing hall ticket number 21P61A0552 and D. AMULYA bearing hall ticket number 21P61A0564 here by declare that the project report entitled “heart disease prediction using machine learning “ under the guidance of Mr. Dara Raju , Department of Information Technology, VBIT, Hyderabad, is submitted in partial fulfilment of the requirement for the award of the degree of Bachelor of Technology in Information Technology.

This is a record of Bonafide work carried out by me and the results embodied in this project has not been reproduced or copied from any source. The results embodied in this project report have not been submitted to any other university or institute for the award of any other degree or diploma.

B.GNANIKA - 21P61A0519
CH. SHIVANI - 21P61A0552
D.AMULYA - 21P61A0564

ACKNOWLEDGEMENT

We are extremely thankful to our beloved Chairman, Dr.N. Goutham Rao and Secretary, Dr.G. Manohar Reddy who took keen interest to provide us the infrastructural facilities for carrying out the project work for her encouragement and support throughout the project.

We whole-heartedly thank Dr.P.V.S. Srinivas Professor and Principal, and Dr. Dara Raju s, Professor & Head of the Department, Computer Science and Engineering for their encouragement and support and guidance in carrying out the major project phase II.

We would like to express our in debtedness to the Overall Project Coordinator, Dr. T. Praveen Associate Professor, and Section coordinators, Mr. G. Arun, Dr. N. Swapna, Department of CSE for their valuable guidance during the course of project work.

We thank our Project Guide, Dr Dara Raju, Associate Professor, for providing us with an excellent project and guiding us in completing our major project phase II successfully

We would like to express our sincere thanks to all the staff of Computer Science and Engineering, VBIT, for their kind cooperation and timely help during the course of our project.

Finally, we would like to thank our parents and friends who have always stood by us whenever we were in need of them

ABSTRACT

Heart disease is the major cause of deaths worldwide. To give treatment for heart disease, a lot of advanced technologies are used. In medical center it is the most common problem that many of medical persons do not have equal knowledge and expertise to treat their patient so they deduce their own decision and as a result it show poor outcome and sometime leads to death. To overcome these problems predictions of heart disease using machine learning algorithms and data mining techniques, it become easy to automatic diagnosis in hospitals as they are playing vital role in this regard. Heart disease can be predicted by performing analysis on patient's different health parameters. There are different algorithm to predict heart disease like naïve Bayes, k Nearest Neighbor (KNN), Decision tree, Artificial Neural Network (ANN). We have used different parameters to predict heart disease. Those parameters are Age, Gender, Cerebral palsey (CP), Nayab Akhtar, Nayabf52@gmail.com Department of Software Engineering Fatima Jinnah Women University, The Mall, Rawalpindi, Pakistan Gender, Cerebral palsey (CP), Blood Pressure (bp), Fasting blood sugar test (fbs) etc. In our research paper, we used built in dataset. we have implement the five different techniques with same dataset to predict heart disease These implemented algorithm are Naive Bayes, k Nearest Neighbor (KNN), Decision tree, Artificial Neural Network (ANN), Random Forest. This paper investigates that which technique gives more accuracy in predicting heart disease based on health parameters. Experiment show that Naïve Bayes has the highest accuracy of 88%.

Keyword: Naive Bayes, k Nearest Neighbor (KNN), Decision tree, Artificial Neural Network (ANN), Random Forest, Heart Disease

VISION

To become, a Center for Excellence in Computer Science and Engineering with a focused Research, Innovation through Skill Development and Social Responsibility.

MISSION

DM-1: Provide a rigorous theoretical and practical framework across State-of-the-art infrastructure with an emphasis on software development.

DM-2: Impact the skills necessary to amplify the pedagogy to grow technically and to meet interdisciplinary needs with collaborations.

DM-3: Inculcate the habit of attaining the professional knowledge, firm ethical values, innovative research abilities and societal needs.

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

PEO-01: Domain Knowledge: Synthesize mathematics, science, engineering fundamentals, pragmatic programming concepts to formulate and solve engineering problems using prevalent and prominent software.

PEO-02: Professional Employment: Succeed at entry- level engineering positions in the software industries and government agencies.

PEO-03: Higher Degree: Succeed in the pursuit of higher degree in engineering or other by applying mathematics, science, and engineering fundamentals.

PEO-04: Engineering Citizenship: Communicate and work effectively on team-based engineering projects and practice the ethics of the profession, consistent with a sense of social responsibility.

PEO-05: Lifelong Learning: Recognize the significance of independent learning to become experts in chosen fields and broaden professional knowledge.

PROGRAM SPECIFIC OUTCOMES (PSOs)

PSO-01: Ability to explore emerging technologies in the field of computer science and engineering.

PSO-02: Ability to apply different algorithms indifferent domains to create innovative products.

PSO-03: Ability to gain knowledge to work on various platforms to develop useful and secured applications to the society.

PSO-04: Ability to apply the intelligence of system architecture and organization in designing the new era of computing environment.

PROGRAM OUTCOMES (POs)

Engineering graduates will be able to:

PO-01: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO-02: Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO-03: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and cultural, societal, and environmental considerations.

PO-04: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO-05: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.

PO-06: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO-07: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO-08: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO-09: Individual and team work: Function effectively as an individual, and as a member or Leader in diverse teams, and in multidisciplinary settings.

PO-10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO-11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO-12: Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

TABLE OF CONTENTS

CHAPTER 7:

CHAPTER 8:

LIST OF FIGURES

CHAPTER – 1: Introduction

1.1: INTRODUCTION TO THE SYSTEM

Heart disease is the major cause of deaths globally. More people die annually from CVDs than from any other cause, an estimated 12 million people died from heart disease every year. Heart disease kills one person every 34 seconds in the United States. Heart attacks are often a tragic event and are the result of blocking blood flow to the heart or brain. People at risk of heart disease may show elevated blood pressure, glucose and lipid levels as well as stress. All of these parameters can be easily measured at home by basic health facilities.

Medical diagnosis is considered as crucial but difficult task to be done efficiently and effectively The automation of this task is very helpful. Un fortunately all physicians are not experts in any subject specialists and beyond the scarcity of resources there some places. Data mining can be used to find hidden patterns and knowledge that may contribute to successful decision making. This plays a key role for healthcare professionals in making accurate decisions and providing quality services to the public.

The approach provided by the health care organization to professionals who do not have more knowledge and skills is also very important. One of the main limitations of existing methods is the ability to draw accurate conclusions as needed. In our approach, we are using different data mining techniques and machine learning algorithms, Naïve Bayes, k Nearest Neighbor (KNN), Decision tree, Artificial Neural Network (ANN), Random Forest to predict the heart disease based on some health parameters.

1.2: PROBLEM STATEMENT

Maintaining good heart health is crucial because the heart is one of the body's largest and most important organs. Since the majority of diseases have a natural cause, heart disease prediction is vital. Because even a small mistake could lead to fatigue or even death, heart-related disorder diagnosis and prognosis require greater precision, perfection, and correctness. Heart disease is a leading cause of death, with the number of cases rising steadily every year. An illness prediction system is a critical component of the crisis management plan. Training with current data allows machine learning to produce the most accurate predictions.

1.3: OBJECTIVE

The main objective of this research is to develop a heart prediction system that can use heart data collection to find and extract hidden information related to diseases.

- This system aims to predict cardiac disease by applying machine learning techniques to a medical dataset.
- Medical tests should be less expensive.
- To assist in avoiding human prejudices.

1.4: AIM OF THE PROJECT

The aim of heart disease prediction using machine learning is to develop models that can accurately identify individuals at risk of developing heart disease based on various health parameters and demographic factors.

CHAPTER – 2: Literature Survey

2.LITERATURE SURVEY

In order to demonstrate the importance of heart disease as a major public health issue that causes considerable morbidity and mortality, we have set the stage for the literature review here. Effective prevention and treatment of heart disease depend on early and accurate heart disease prediction. It is necessary for us to discuss machine learning. In this context, our project is discussed as a useful tool for improving predictive accuracy. Based on research publications released between 2018 and 2022, this review of the literature examines the development of machine learning applications in the prediction of heart disease.

P.S. HIREMANTH AND S.N. PATIL's "A Hybrid Machine Learning Approach for Automated Heart Disease Risk Prediction" (2018) has caught my attention. In 2018, this research paper on expert systems with applications was published. I am somewhat informed; for example, I am aware of the potential of deep learning for risk factor assessment, which is essential for predicting heart disease.

Age, gender, diabetes, hypertension, cholesterol, and family history are common risk factors forheart disease. It also underlines how important it is to train and test machine learning models using high-quality datasets as information sources. In order to effectively predict heart disease, these datasets frequently include data from wearable devices, medical imaging, and electronic health records.

The authors of "Heart disease prediction using deep learning and traditional machine learning algorithms" are M.K. Hassan and colleagues. The publication appeared in the King Saud University Computer and Information Science Journal in 2019. In the same way, A.L. FONSECA et al.'s 2019 publication "Predicting cardiovascular disease risk factor using deep neural networks" was published in Computer in Biology and Machine. According to what I've studied, these strategies include ensemble methods, deep learning techniques like convolution neural networks, and supervised learning algorithms like logistic regression and random forests. This study outlines the advantages and disadvantages of using these machine learning methods for tasks predicting heart disease. Mostly the Cleveland Heart Disease and Framingham Heart Study datasets, which are widely used to predict heart disease. The important phases in data preprocessing, feature engineering, and feature selection that go into preparing data for machine learning models are also covered. For example, finding pertinent input variables that lead to more accurate predictions requires feature engineering. By looking through research papers from the previous five years, we can learn about the difficulties and constraints associated with using machine learning to predict heart disease. These difficulties could involve problems with data quality, class imbalance in datasets, and the requirement for better model interpretability. Certain research papers examine possible avenues for future investigation, including the integration of multi-modal data from diverse sources (e.g., imaging, genomics, and EHRs), the application of personalized medicine techniques, and the creation of real-time monitoring systems to improve patient care and heart disease prediction.

AUTHOR	TITLE	PURPOSE
Mr. santhana Krishna. J Dr. Geetha. S	**Prediction of heart disease using machine learning algorithm** **Year – 2016**	This method makes use ofa data set on cardiac illness. This system's primary goal is to forecast the patients' chances of developing heart disease as a percentage. It is performed through the use of data mining and categorization techniques.
M. Marimuthu, S.Deivarani, Gayathri .R	**Analysis of heart disease prediction using various machine learning techniques** **Year – 2018**	To increase the system's efficiency and accuracy to forecast the likelihood of a heart attack
Sanchayita Dhar, Pritha Datta, Ankur Biswas, Tanusree Dey, Krishan Roy	**A hybrid machine learning approach for prediction of heart disease** **Year – 2020**	This research aims to construct a prediction system that can anticipate heart illnesses based on measurements taken from 209 test cases from The ERIC laboratory.

Rajesh N, T Maneesha, Shaik Hafeez, Hari Krishna	**Prediction of heart disease using machinelearning** **Year - 2018**	In this work, we process patient datasets and patient data for which we must forecast the likelihood that a heart illness may manifest.

2.1: EXISTING SYSTEM

The existing systems for heart disease prediction using machine learning primarily focus on leveraging historical data and advanced algorithms to predict the likelihood of heart disease. These systems use various supervised and unsupervised learning techniques, coupled with clinical data, to make predictions. Here's an overview of the current state of such systems:

1. Input Data Sources
2. Machine Learning Algorithms Used
3. Technology Stack
4. Existing Frameworks and Models
5. Challenges in Existing Systems

2.2: PROPOSED SYSTEM

The first stage of the system's operation involves gathering data and choosing the most crucial attributes. Algorithms are employed, and the training data is this phase. The necessary format is then created by preprocessing the important data. The data is then separated into training and testing data. This approach was used to train the model. Test data is used to determine the accuracy of the system. The modules below are used to implement this system.

✓ Compiling the dataset

✓ The choice of characteristics

✓ Preprocessing Data

✓ Data Balancing

✓ Illness Prognosis

2.3: SCOPE OF THE PROJECT

The scope of a heart disease prediction project using machine learning outlines the boundaries, potential applications, and future possibilities of the system. It includes various aspects of healthcare, technology, and research, aiming to enhance the early detection and management of heart disease. Here's a detailed overview:

1. Objectives of the Project
2. Technical Scope
3. Functional Scope
4. Applications

Healthcare Providers:

5. Geographical and Demographic Scope
6. Future Expansion
7. Expected Outcomes

CHAPTER-3: Analysis

3: Analysis

Analysis of Heart Disease Prediction Using Machine Learning

Heart disease prediction using machine learning (ML) involves examining its feasibility, effectiveness, and limitations in real-world scenarios. This analysis evaluates the technical, healthcare, and societal aspects of implementing machine learning for predicting heart disease.

3.1: Technical Analysis

a. Data Requirements

Types of Data: Clinical (e.g., cholesterol levels, blood pressure), demographic (e.g., age, gender), and lifestyle data (e.g., smoking habits).

Volume and Quality: A large dataset is essential for model training and testing.

Missing or inconsistent data can lead to reduced model performance.

Feature Selection: Key features influencing predictions include age, BMI, cholesterol, glucose levels, and ECG results.

b. Algorithms Used

Supervised Learning: Logistic Regression: Simple, interpretable, but may underperform for complex patterns.

Random Forest and XG Boost : Effective for handling nonlinear relationships and feature importance.

Support Vector Machines: Useful for small datasets with high dimensionality.

Deep Learning: Neural Networks: Effective for large datasets, particularly for imaging data (e.g., echocardiograms).

Hybrid Models: Ensemble methods like stacking or boosting can improve prediction accuracy.

c. Performance Metrics

Accuracy, precision, recall, and F1-score are commonly used.

The AUC-ROC curve is essential for assessing model performance in distinguishing between heart disease presence and absence.

3.2: Healthcare Impact Analysis

a. Benefits

Early Diagnosis: Enables detection of heart disease before severe symptoms develop, improving patient outcomes.

Personalized Care: Helps create tailored healthcare plans based on individual risk factors.

Enhanced Efficiency: Automates risk assessment, reducing the workload on healthcare providers.

b. Applications

Clinical Settings: Aids in decision-making by providing probabilistic predictions.

Remote Monitoring: Integration with wearable devices enables real-time monitoring of risk factors.

c. Challenges

Data Privacy: Ensuring patient data confidentiality is critical (e.g., complying with HIPAA, GDPR).

Bias and Fairness: Models may reflect biases in training data, leading to unfair predictions for certain groups.

Integration with Healthcare Systems: Seamless integration with existing electronic health records (EHRs) is complex.

3.3: Societal and Economic Analysis

a. Accessibility

Developed Regions: Advanced systems can integrate with existing healthcare infrastructure.

Low-Resource Settings: The lack of comprehensive datasets and technology may hinder implementation.

b. Cost Effectiveness

Reducing the burden of late-stage heart disease treatment through preventive care saves healthcare costs.

Initial development and deployment costs may be high but can be offset by long-term benefits.

3.4: Comparative Analysis

Traditional Methods: Machine learning provides faster and more accurate predictions than conventional statistical methods.

Human Expertise: ML models act as decision support tools, augmenting but not replacing clinical judgment.

3.5: Limitations

Data Quality Issues: Missing or inaccurate data can lead to unreliable predictions. Imbalanced datasets may cause models to favor the majority class.

Interpretability: Complex models like neural networks are less interpretable, posing challenges in clinical use.

Generalizability: Models trained on specific populations may not perform well across diverse demographics.

3.6: Future Opportunities

Explainable AI: Enhancing model interpretability to gain trust among healthcare professionals.
Real-Time Monitoring: Integration with IoT devices for continuous health tracking.
Global Implementation: Developing systems tailored for underserved regions using open-source tools and datasets.

CHAPTER –4: Hardware and Software

4.HARDWARE AND SOFTWARE REQUIREMENTS

4.1: Hardware Requirements

Operating System: Windows 11

Database:

- **Relational Database:** MySQL 8.0 or higher
- **Non-Relational Database (NoSQL):** MongoDB for large-scale unstructured data

Programming Language: Python

Libraries: Scikit-learn, Pandas, NumPy, Matplotlib

4.2: Software Requirements

Processor: Intel CORE i5 or higher, Processor Speed

Memory: 16GB RAM minimum; 32GB recommended

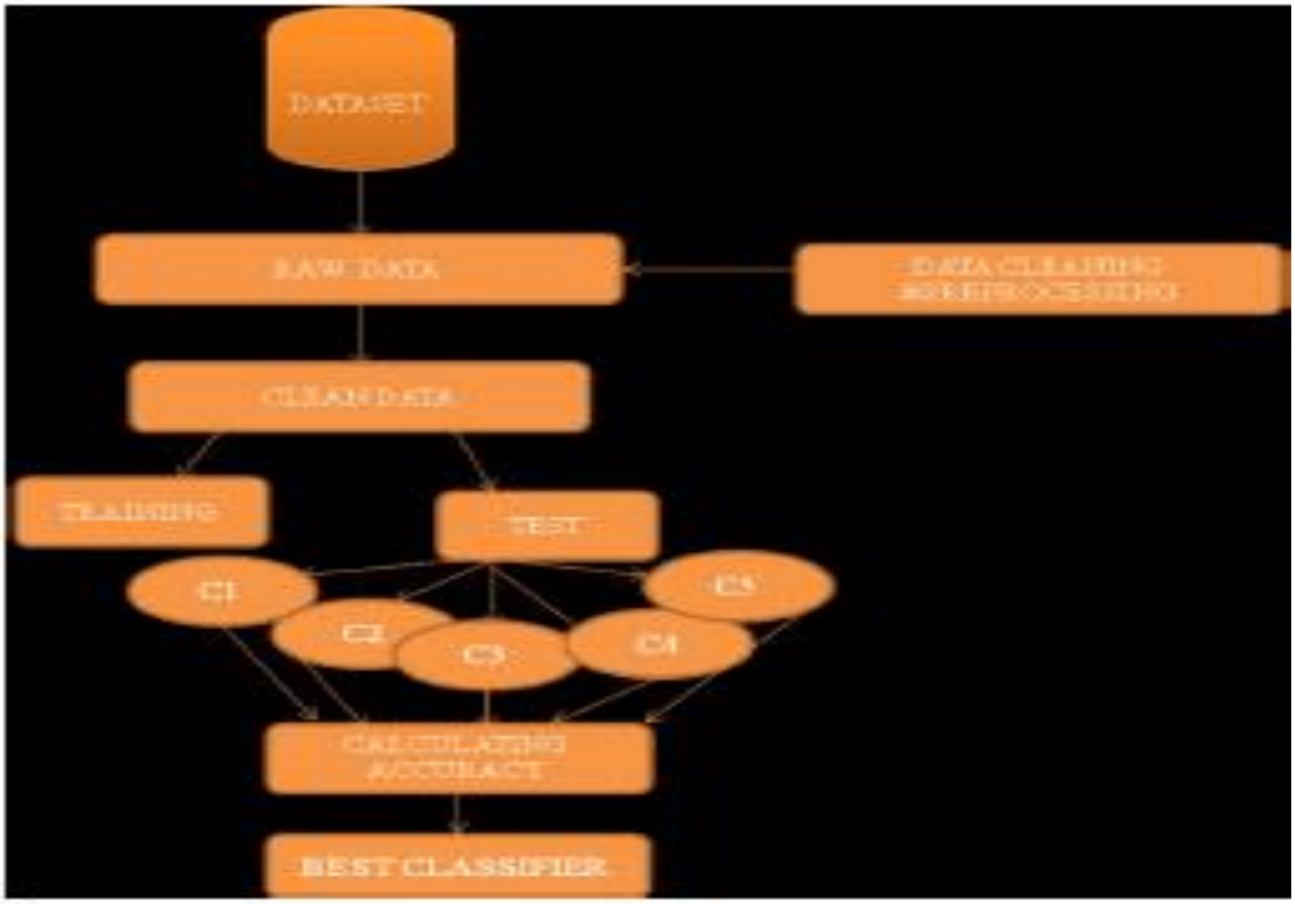
Storage:

- **Primary:** 500GB SSD for OS and software
- **Data Storage:** 1TB HDD or SSD, expandable

CHAPTER – 5: System Design

5.SYSTEM DESIGN

This system is an effective heart attack prediction system that we are developing with the Naïve Bayes algorithm. An input CSV file or manual entry can be fed into the system. Naïve Bayes is applied after the data is received. The operation is completed and a suitable heart attack level is produced after gaining access to the data set. After discussing the proposed system with experts in medicine and medical professionals, the weight, age, and priority levels of the additional criteria significant to heart attacks will be added. The heart attack prediction system was developed to help identify different heart attack risk levels, such as standard, low, and high, and to provide drugs information regarding the expected result. In this system, we're using the Naïve Bayes algorithm to build an effective heart attack prediction system. Either a CSV file or manual input can be fed into the system. The Naïve Baye algorithm is used as soon as the data is received. An appropriate heart attack level is generated once the operation is completed and the data set is accessed. After consulting with experts in medicine and skilled doctors, the weight, age, and priority levels of the additional criteria significant to heart attacks will be added to the proposed system. The heart attack prediction system was developed to help identify different heart attack risk levels, such as standard, low, and high, and to provide medication information regarding the expected result.



5.1: SOFTWARE DESIGN

Designing software for heart disease prediction using machine learning involves several steps, including defining objectives, selecting appropriate algorithms, and developing a system architecture. Below is a high-level design plan:

1. Problem Definition
2. System Architecture
3. Workflow
4. Security and Privacy
5. Challenges and Considerations
6. Tools and Libraries

5.2: INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

➢What data should be given as input?

➢How the data should be arranged or coded?

➢The dialog to guide the operating personnel in providing input.

➢Methods for preparing input validations and steps to follow when error occur.

5.3: OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2.Select methods for presenting information.

3.Create document, report, or other formats that contain information produced by the system.

5.4: UML DIAGRAMS

5.4.1 USECASE DIAGRAM

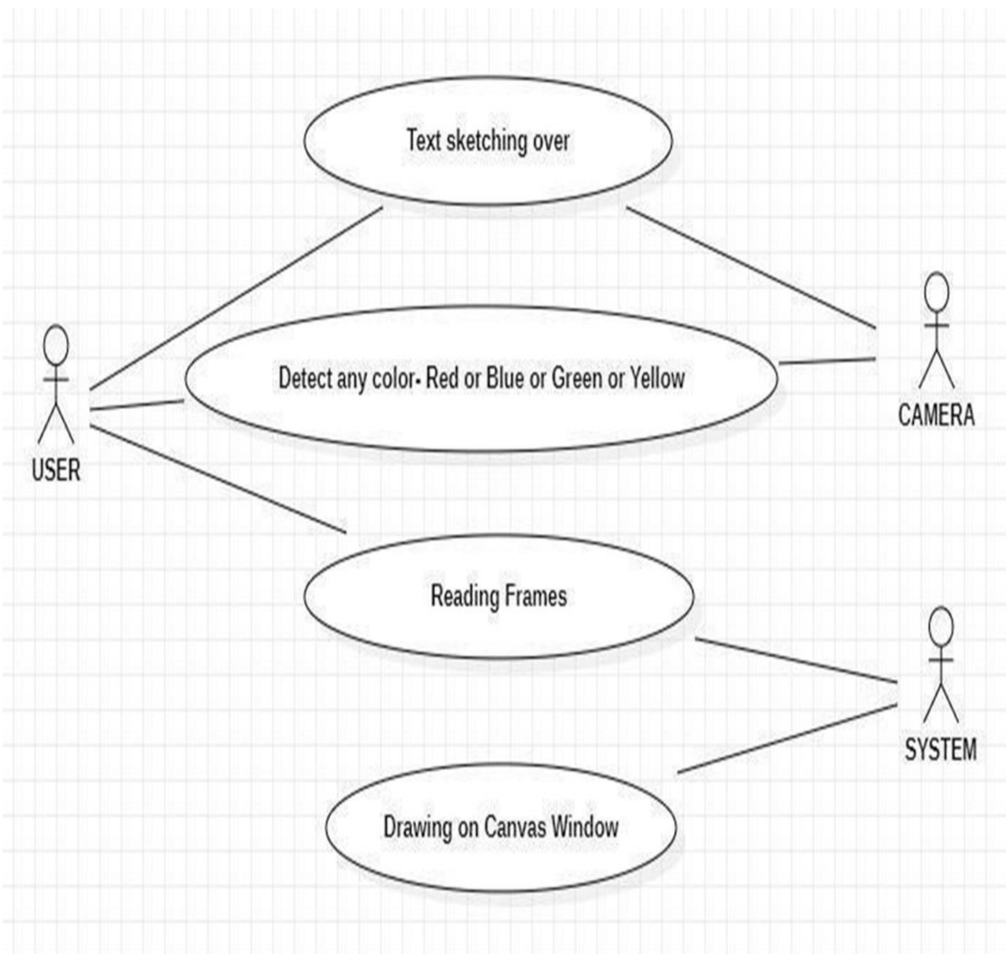


Fig:5.4.1

5.4.2 SEQUENCE DIAGRAM

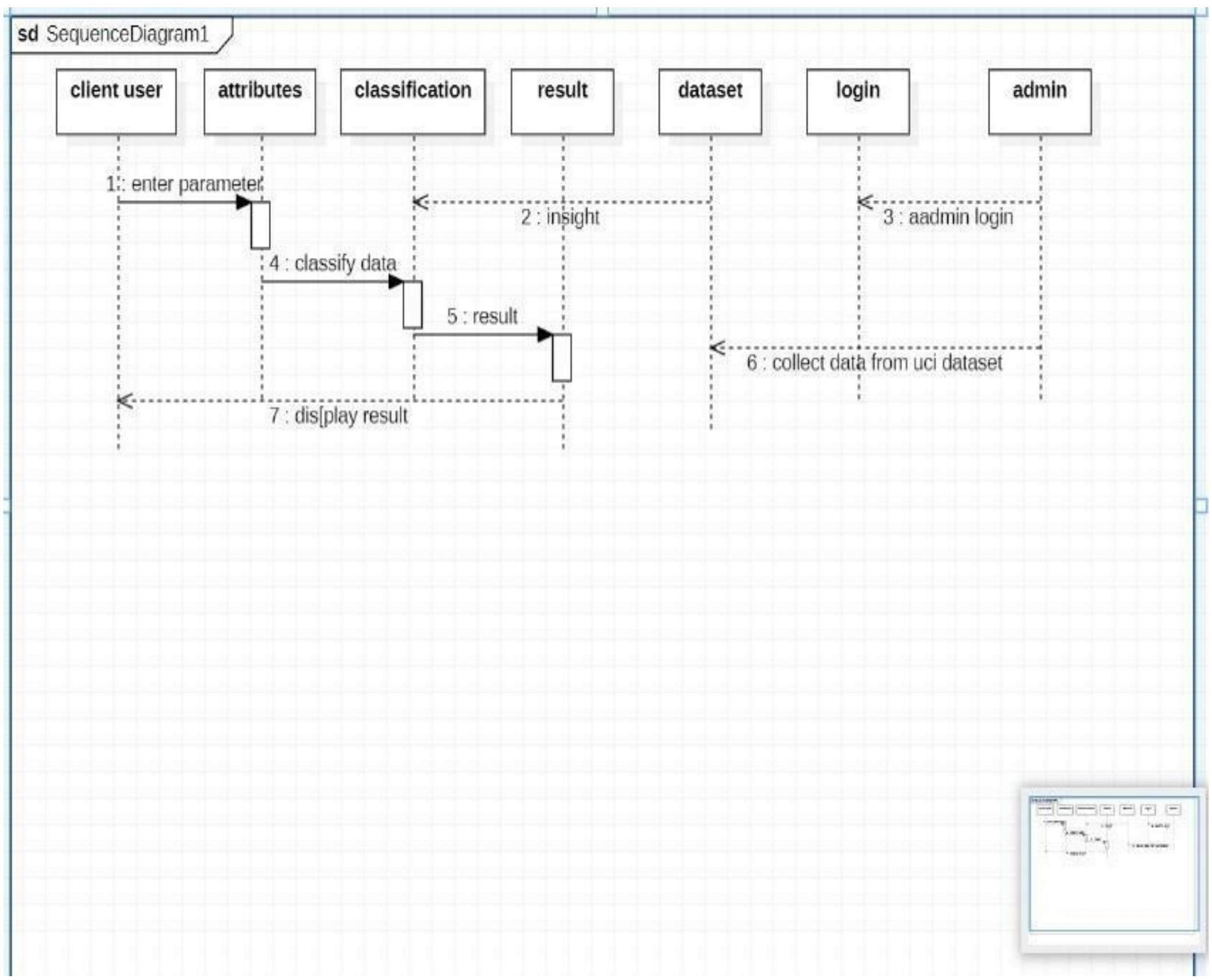


Fig:5.4.2

5.4.2 ACTIVITY DIAGRAM

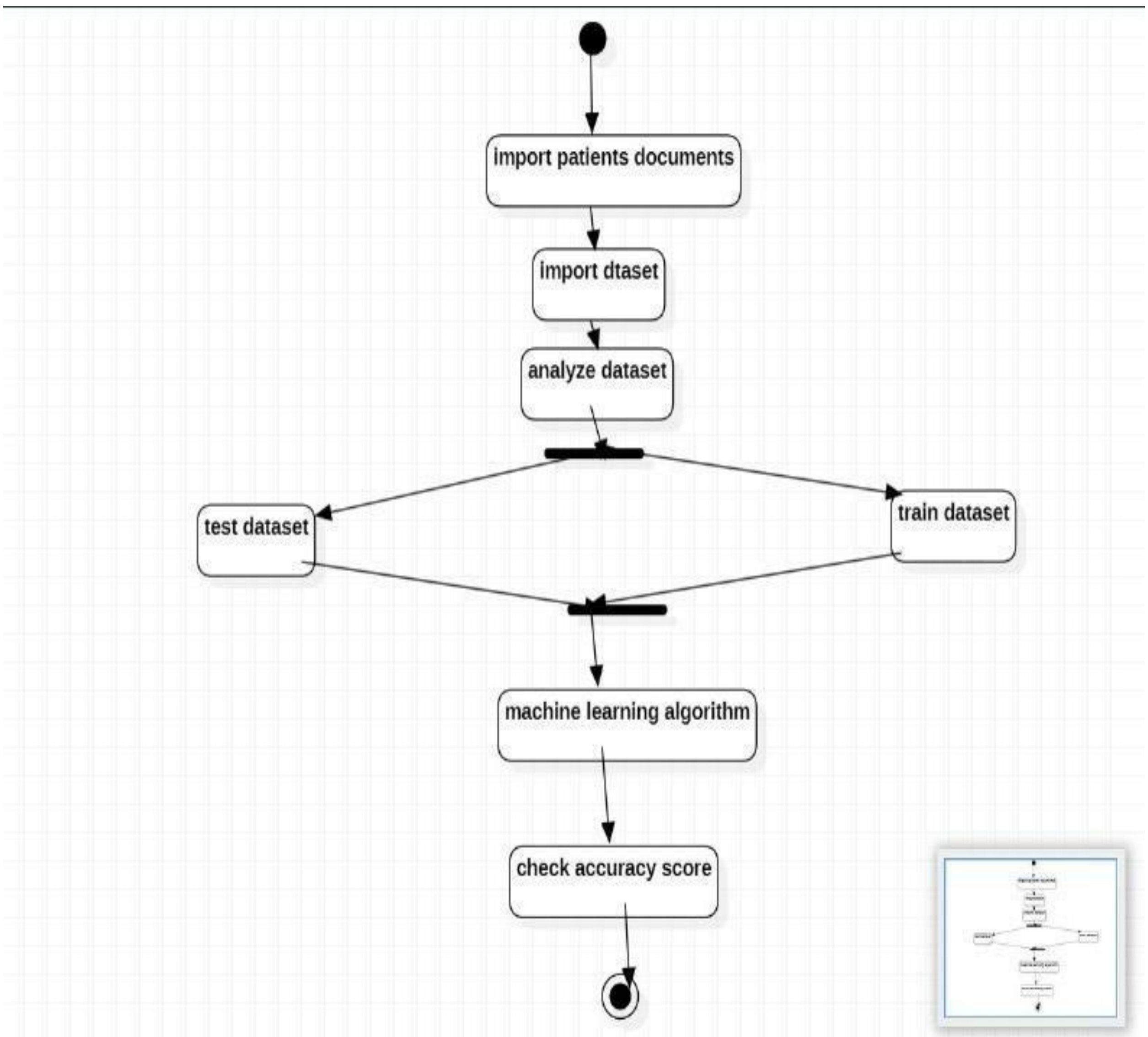


Fig:5.4.3

CHAPTER-6: Result and performance Analysis

6.RESULTS AND PERFORMANCE ANALYSIS

BACK END CODE

RANDOM FOREST & KNN NEIGHBOUR:

```
import numpy as np import pandas as pd
import matplotlib.pyplot as plt from matplotlib import rcParams from matplotlib.cm import rainbow
%matplotlib inline import warnings
warnings.filterwarnings('ignore')
from sklearn.neighbors import KNeighborsClassifier from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier df = pd.read_csv('dataset.csv')
df.info() df.describe()
import seaborn as sns
top_corr_features = corrmat.index
plt.figure(figsize=(20,20)) #plot heat map
g=sns.heatmap(df[top_corr_features].corr(),annot=True,cmap="RdYlGn") df.hist()
sns.set_style('whitegrid') sns.countplot(x='target',data=df,palette='RdBu_r')
dataset = pd.get_dummies(df, columns = ['sex', 'cp', 'fbs', 'restecg', 'exang', 'slope', 'ca', 'thal']) from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler standardScaler = StandardScaler()
columns_to_scale = ['age', 'trestbps', 'chol', 'thalach', 'oldpeak'] dataset[columns_to_scale] = standardScaler.fit_transform(dataset[columns_to_scale]) dataset.head()
y = dataset['target']
X = dataset.drop(['target'], axis = 1)
from sklearn.model_selection import cross_val_score knn_scores = []
for k in range(1,21):
knn_classifier=KNeighborsClassifier(n_neighbors = k) score=cross_val_score(knn_classifier,X,y,cv=10)
knn_scores.append(score.mean())
plt.plot([k for k in range(1, 21)], knn_scores, color = 'red')
for i in range(1,21):
plt.text(i, knn_scores[i-1], (i, knn_scores[i-1])) plt.xticks([i for i in range(1, 21)]) plt.xlabel('Number of
```

Neighbors (K)') plt.ylabel('Scores')

plt.title('K Neighbors Classifier scores for different K values') knn_classifier = KNeighborsClassifier(n_neighbors = 12) score=cross_val_score(knn_classifier,X,y,cv=10) score.mean()

from sklearn.ensemble import RandomForestClassifier randomforest_classifier= RandomForestClassifier(n_estimators=10)

score=cross_val_score(randomforest_classifier,X,y,cv=10) score.mean()

LOGISTIC REGRESSION

import numpy as np import pandas as pd

from sklearn.model_selection import train_test_split from sklearn.linear_model import LogisticRegression from sklearn.metrics import accuracy_score

loading the csv data to a Pandas DataFrame heart_data = pd.read_csv('heart_data.csv')

print first 5 rows of the dataset

heart_data.head()

print last 5 rows of the dataset heart_data.tail()

number of rows and columns in the dataset heart_data.shape

getting some info about the data heart_data.info()

checking for missing values heart_data.isnull().sum()

statistical measures about the data heart_data.describe()

checking the distribution of Target Variable heart_data['target'].value_counts()

X = heart_data.drop(columns='target', axis=1) Y = heart_data['target']

print(X) print(Y)

X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=2) print(X.shape, X_train.shape, X_test.shape)

model = LogisticRegression()

X_train_prediction = model.predict(X_train)

training_data_accuracy = accuracy_score(X_train_prediction, Y_train) print('Accuracy on Training data : ', training_data_accuracy)

X_test_prediction = model.predict(X_test)

test_data_accuracy = accuracy_score(X_test_prediction, Y_test) print('Accuracy on Test data : ', test_data_accuracy)

input_data = (60,1,0,130,253,0,1,144,1,1.4,2,1,3)

```
prediction = model.predict(input_data_reshaped) print(prediction)
if (prediction[0]== 0):
print('The Person does not have a Heart Disease') else:
print('The Person has Heart Disease')
```

OUTPUT

LOGISTIC REGRESSION

```
Accuracy on Training data :  0.8512396694214877
```

```
Accuracy on Test data :  0.819672131147541
```

```
[0]
The Person does not have a Heart Disease
```

RANDOM FOREST

```
score.mean()
```

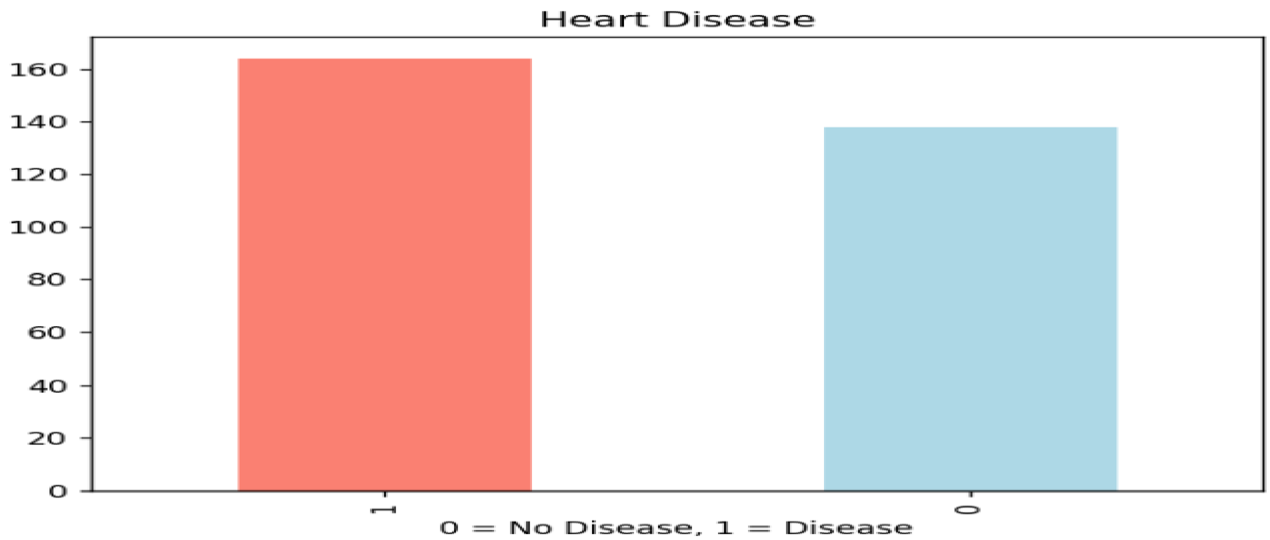
```
0.8019354838709678
```

K-NEIGHBOUR CLASSIFIER

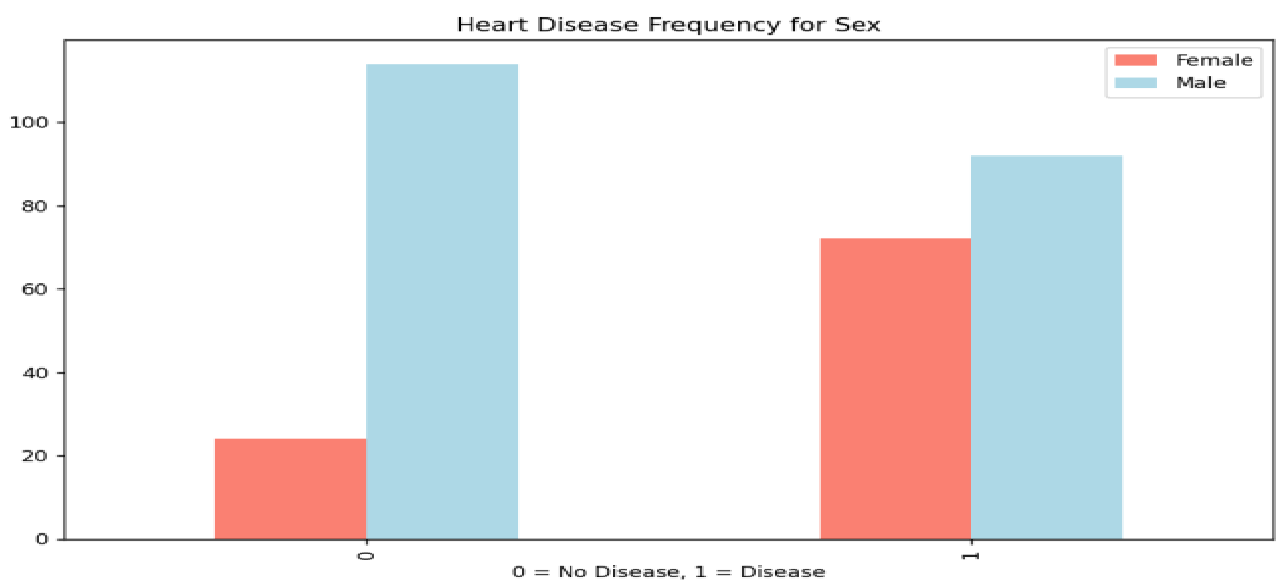
```
score.mean()
```

```
0.8448387096774195
```

A comparison of testing and training accuracy:



Loss from Testing vs. Training:



CHAPTER-7: Conclusion

CONCLUSION

7.1: PRESENT WORK

Expandable, scalable, dependable, GUI-based, and intuitive is how the that it can help reduce treatment costs by providing timely early diagnoses. Medical professionals, cardiologists, and medical students alike will be ableto use the model as a soft diagnostic tool and training aid. This method is useful for initial cardiac patient diagnosis for general practitioners. To improve this prediction system's accuracy and scalability, a number of improvements could be looked at. Now that a universal design has been produced, we can use it going forward to examine other sets of data. Another productive direction for future research is to manage multiple class labelsin the prediction process, which can greatly improve the performance of the health diagnosis. The dimensionality of the heart database in DM warehouses makes it likely that future research will have difficulty identifying and selecting critical variables for a more accurate diagnosis of heart disease.

7.2: FUTURE SCOPE

Future developments in data science and technology will accelerate the prediction of cardiac disease, which is very promising. Personalized health and cutting-edge teaching methods are two of the main areas of development.

- Clinical trials and research
- global health
- early detection
- integration with healthcare systems
- public health initiatives

REFERENCE

1. Kelly, B. B., &Fuster, V. (Eds.). (2010). Promoting cardiovascular health in the developing world: a critical challenge to achieve global health. National Academies Press.

2. Poirier, Paul, et al. "Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism." Circulation 113.6 (2006): 898-918.

3. Bhatnagar, Prachi, et al. "Trends in the epidemiology of cardiovascular disease in the UK." Heart 102.24 (2016): 1945-1952.

4. Beunza, Juan-Jose, et al. "Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease)." Journal of biomedical informatics 97 (2019): 103257.

5. Zhao, Lina, et al. "Enhancing Detection Accuracy for Clinical Heart Failure Utilizing Pulse Transit Time Variability and Machine Learning." IEEE Access 7 (2019): 17716-17724.

6. Borkar, Sneha, and M. N. Annadate. "Supervised Machine Learning Algorithm for Detection of Cardiac Disorders." 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE, 2018.

7. Omar Boursalie, Reza Samavi, Thomas E. Doyle. “M4CVD: Mobile Machine Learning Model for Monitoring Cardiovascular Disease.” Procedia Computer Science, 63 (2015): 384-391.

8. Chen, Rui, et al. "Using Machine Learning to Predict One-year Cardiovascular Events in Patients with Severe Dilated Cardiomyopathy." European Journal of Radiology (2019).

9. Dhar, Sanchayita, et al. "A Hybrid Machine Learning Approach for Prediction of Heart Diseases." 2018 4th International Conference on Computing Communication and Automation (ICCCA). IEEE, 2018.

10. Dinesh, Kumar G., et al. "Prediction of Cardiovascular Disease Using Machine Learning Algorithms." 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT). IEEE, 2018.

11. Mezzatesta, Sabrina, et al. "A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis." Computer Methods and Programs in

Biomedicine 177 (2019): 9-15.
12. Terrada, Oumaima, et al. "Classification and Prediction of atherosclerosis diseases using machine learning algorithms." 2019 5th International Conference on Optimization and Applications (ICOA). IEEE, 2019.
13. Alić, Berina, Lejla Gurbeta, and Almir Badnjevic. "Machine learning techniques for classification of diabetes and cardiovascular diseases." 2017 6th Mediterranean Conference on Embedded Computing (MECO). IEEE, 2017.
14. Awan, Shahid Mehmood, Muhammad Usama Riaz, and Abdul Ghaffar Khan. "Prediction of heart disease using artificial neural network." VFAST Transactions on Software Engineering 13.3 (2018): 102-112.