HEART DISEASE PREDICTION USING MACHINE LEARNING MODELS

A DESIGN PROJECT REPORT

Submitted by

JEEVA M

SHANKARANARAYANAN R

THINESH M

HARRISH GOKUL M

in partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENGE AND DATA SCIENCE

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

SAMAYAPURAM-621112

JUNE 2024

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(AUTONOMOUS)

SAMAYAPURAM–621112

BONAFIDE CERTIFICATE

Certified that this design project report titled **“HEART PREDICTION USING MACHINE LEARNING MODELS”** is the bonafide work of **JEEVA M** (**REG NO: 811721243019**), **SHANKARANARAYANAN R (REG NO: 811721243051), THINESH M (REG NO:811721243058), HARRISH GOKUL M (REG NO:811721243301)** who carried out the project under my supervision.

SIGNATURE
Dr. T. Avudaiappan ,M.E., Ph.D.,
HEAD OF THE DEPARTMENT
Associate Professor
Department of Artificial Intelligence
K. Ramakrishnan College of
Technology(Autonomous)
Samayapuram – 621 112

SIGNATURE
Mrs. R.Reetha Jeyarani, M.E.,
SUPERVISOR
Assistant Professor
Department of Artificial Intelligence
K. Ramakrishnan College of
Technology(Autonomous)
Samayapuram – 621 112

Submitted for the viva-voce examination held on ………………

INTERNAL EXAMINER **EXTERNAL EXAMINER**

DECLARATION

We jointly declare that the project report on **"HEART PREDICTION USING MACHINE LEARNING MODELS"** is the result of original work done by us and best of our knowledge, similar work has not been submitted to **"ANNA UNIVERSITY CHENNAI"** for the requirement of Degree of **BACHELOR OF TECHNOLOGY**. This design project report is submitted on the partial fulfilment of the requirement of the award of Degree of **BACHELOR OF TECHNOLOGY**.

SIGNATURE

____

JEEVA M

____

SHANKARANARAYANAN R

____

THINESH M

____

HARRISH GOKUL M

PLACE **:** SAMAYAPURAM

DATE **:**

ACKNOWLEDGEMENT

It is with great pride that we express our gratitude and in - debt to our institution "**K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY (AUTONOMOUS)**", for providing us with the opportunity to do this project.

We are glad to credit honorable chairman **Dr. K. RAMAKRISHNAN**, **B.E.,** for having provided for the facilities during the course of our study in college.

We would like to express our sincere thanks to our beloved Executive Director **Dr. S. KUPPUSAMY, MBA., Ph.D.,** for forwarding to our project and offering adequate duration in completing our project.

We would like to thank **Dr. N. VASUDEVAN, M.E., Ph.D.,** Principal, who gave opportunity to frame the project the full satisfaction.

We whole heartily thanks to **Dr. T. AVUDAIAPPAN**, **M.E., Ph.D.,** Head of The Department, **ARTIFICAL INTELLIGENCE** for providing his encourage pursuing this project.

I express my deep and sincere gratitude to my project guide **Mrs. R. REETHA JEYARANI, M.E.,** ASSISTANT PROFESSOR, **ARTIFICIAL INTELLIGENCE** for his incalculable suggestions, creativity, assistance and patience which motivated me to carry out the project successfully.

I render my sincere thanks to my Project Coordinator **Mrs. G. NALINA KEERTHANA, M.E.,**and other staff members for providing valuable information during the course. I wish to express my special thanks to the officials & Lab Technicians of our departments who rendered their help during the period of the work progress.

ABSTRACT

The cases of heart diseases are increasing at a rapid rate and it's very Important and concerning to predict any such diseases beforehand. The diagnosis is a difficult task i.e. it should be performed precisely and efficiently. Heart disease prediction system can be used to predict whether the patient is likely to be diagnosed with a heart disease or not using the medical history of the patient. Different algorithms of machine learning such as logistic regression and KNN to predict and classify the patient with heart disease. A quite Helpfulapproach was used to regulate how the model can be used to improve the accuracy of prediction of Heart Attack in any individual. The strength of the proposed model was quiet satisfying and was able to predict evidence of havinga heart disease in a particular individual by using KNN and Logistic Regressionwhich showed a good accuracy in comparison to the previously used classifier such as naive bayes etc. So a quiet significant amount of pressure has been lift off by using the given model in finding the probability of the classifier to correctly and accurately identify the heart disease. The Given heart disease prediction system enhances medical care and reduces the cost. Heart disease prediction system gives us significant knowledge that can help us predict the patients with heart disease It is implemented on the .pynb format.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
CNN	Convolutional Neural Networks
ECG	Electro Cardiogram
EHR	Electronic Health Records
KNN	K Nearest Neighbor
ML	Machine Learning

CHAPTER 1

INTRODUCTION

Machine Learning is a way of Manipulating and extraction of implicit, previously unknown/known and potential useful information about data” . Machine Learning isa very vast and diverse field and its scope and implementation is increasing day by day.Machine learning Incorporates various classifiers of Supervised, Unsupervised and Ensemble Learning which are used to predict and Find the Accuracy of the given dataset.We can use that knowledge in our project of HDPS as it will help a lot of people. Cardiovascular diseases are very common these days, they describe a range of conditionsthat could affect your heart. World health organization estimates that 17.9 million global deaths from (Cardiovascular diseases) CVDs .

1.1 BACKGROUND

Heart disease prediction using machine learning involves utilizing advanced computational techniques to analyze diverse health-related data for the purpose of forecasting the likelihood of an individual developing heart disease. This innovative approach harnesses the capabilities of machine learning algorithms to discern patterns, correlations, and trends within datasets, enabling more accurate predictions compared to traditional methods.

The process begins with the collection of relevant health data, encompassing demographic information, lifestyle factors, medical history, and clinical measurements like blood pressure and cholesterol levels. This data is then meticulously cleaned, preprocessed, and organized to create a format suitable for machine learning analysis. Feature selection becomes pivotal at this stage, involving the identification of the most pertinent predictive variables from the dataset.

Various machine learning algorithms are considered for heart disease prediction, including logistic regression, decision trees, random forests, support vector machines (SVM), and neural networks. The chosen model is trained on a subset of the dataset, allowing it to learn patterns and relationships between the features and the presence or absence of heart disease.

The trained model undergoes validation and testing on separate datasets to assess its generalizability and performance, using metrics such as accuracy, precision, recall, and F1 score. Hyperparameter tuning may be employed to optimize the model's performance, involving adjustments to parameters not learned during training.

Upon satisfactory performance, the model is deployed for real-world use, potentially integrating it into healthcare systems or applications for automated heart disease risk assessments. Continuous improvement is essential, with the model regularly updated as more data becomes available or as it encounters new cases. This iterative process ensures the model remains accurate and relevant over time.

Predictive modeling through machine learning holds promise in assisting healthcare professionals by identifying individuals at a higher risk of heart disease, enabling early intervention and personalized preventive measures. However, it's crucial to acknowledge that these models should complement rather than replace clinical judgment and expertise. Ethical considerations and patient privacy must be carefully addressed throughout the development and deployment of such models.

1.2 PROBLEM STATEMENT

It is the primary reason of deaths in adults. Our project can help predict the people who are likely to diagnose with a heart disease by help of their medical history. It recognizes who all are having any symptoms of heart disease such as chest pain or high blood pressure and can help in diagnosing disease with less medical tests and effective treatments, so that they can be cured accordingly. This project focuses on mainly three data mining techniques namely: (1) Logistic regression, (2) KNN and (3) Random Forest Classifier. The accuracy of our project is 87.5% for which is better than previous system where only one data mining technique is used. So, using more data mining techniques increased the HDPS accuracy and efficiency. Logistic regression falls under the category of supervised learning. Only discrete values are used in logistic regression.

1.3 OBJECTIVES

The aim of heart disease prediction is to leverage data and machine learning techniques to develop a model that can effectively predict the likelihood of an individual having heart disease. Identifying individuals who may be at risk at an early stage allows for timely intervention and preventive measures, potentially reducing the severity of the condition. Providing a quantitative assessment of an individual's risk factors based on various health-related parameters can assist healthcare professionals in tailoring personalized interventions and treatment plans. Additionally, identifying modifiable risk factors encourages and enables proactive healthcare measures.

The objective is to determine whether a patient is likely to be diagnosed with cardiovascular heart diseases based on their medical attributes, such as gender, age, chest pain, and fasting sugar level. A dataset selected from the UCI repository contains patients' medical history and attributes. Using this dataset, predictions are made about the likelihood of a patient developing heart disease.

By employing sophisticated machine learning models, this approach can analyze vast amounts of patient data to uncover patterns and correlations that may not be immediately apparent through traditional methods. This enhances the ability to predict heart disease with greater accuracy. The insights gained from such models can inform healthcare strategies and policies, ultimately aiming to reduce the incidence and impact of heart disease. By focusing on data-driven methodologies, this initiative seeks to contribute to the broader goal of improving public health outcomes through early detection and personalized care.

CHAPTER 2
LITERATURE SURVEY

2.1 RANDOM FOREST BASED HEART DISEASEPREDICTION

Author: Preethi Sondhi

Year Of Publication: 2021

Algorithms Used: Naïve bayes, Support Vector Machine, Random Forest

Abstract

This research investigates the application of a Random Forest-based approach for predicting heart disease, aiming to enhance the accuracy and reliability of predictive models. Health-related data, encompassing demographic details, lifestyle factors, and clinical measurements, is collected and processed to create a suitable dataset. Feature selection is employed to identify the most influential variables for predicting heart disease. The Random Forest model is chosen for its ensemble learning capabilities, leveraging multiple decision trees to collectively contribute to the prediction process. The model undergoes rigorous training on a representative dataset, learning intricate patterns and relationships within the data. Validation and testing procedures assess the model's generalizability and performance, employing established metrics such as accuracy, precision, recall, and F1 score.

Merit

By using random forest algorithm has been found to have 95.08% more precisioncompared to other algorithms.

Demerit

Forest is an ensemble of decision trees, and the combination of multiple trees can result in a relatively complex model. This complexity can lead to longer trainingtimes and increased computational overhead, especially when dealing with large datasets.

2.2 HEART DISEASE PREDICTION USING HYBRID MACHINELEARNING

Author: Tanmay Salunke, Pavan Kumar Jagade.

Year Of Publication: 2023

Algorithm Used: leverage ML techniques to identify significant features that can improve accuracy of heart disease prediction.

Abstract

Heart disease remains a significant global health concern, necessitating innovative approaches for early detection and risk assessment. This study explores the application of hybrid machine learning techniques for heart disease prediction, aiming to enhance predictive accuracy and robustness. Health-related data, including demographic information, lifestyle factors, and clinical measurements, are collected and meticulously prepared for analysis. Feature selection is employed to identify pertinent variables crucial for model development. A hybrid model is constructed by integrating various machine learning algorithms, such as traditional methods like logistic regression or decision trees with advanced techniques like neural networks or ensemble methods. The hybrid model undergoes thorough training on a representative dataset to learn patterns and relationships indicative of heart disease.

Merit

By using ml techniques achieving a superior level of performance with an accuracy rate of 88.7%for predicting heart disease.

Demerit

Integrating machine learning models with IoT devices and the associated data streams can be complex. Ensuring seamless communication and data flow between IoT sensors and machine learning algorithms requires careful engineering.

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

Existing heart disease prediction systems employ various machine learning algorithms, such as Logistic Regression, Decision Trees, Random Forest, SVM, and Neural Networks. These systems enhance prediction accuracy but face challenges like data quality, overfitting, and interpretability.

ALGORITHMS USED

Logistic Regression: Logistic regression is a simple and interpretable algorithm commonly used for binary classification tasks like predicting the presence or absence of heart disease**.**

Decision Trees: Decision trees are tree-like structures where each node represents a decision based on a particular feature. They can be useful for understanding feature importance and decision pathways.

Random Forest: Random Forest is an ensemble learning method that builds multiple decision trees and merges their predictions. It is known for its robustness and can handlenoisy data.

Support Vector Machines (SVM): SVM is a powerful algorithm for classification tasks. It works well in high- dimensional spaces and is effective when there is a clear margin of separation between classes.

K-Nearest Neighbors (KNN): KNN is a simple algorithm that classifies a data point based on the majority class of its k nearest neighbors. It is effective for smaller datasets.

Naive Bayes: Naive Bayes is a probabilistic algorithm based on Bayes' theorem. It assumes independence between features, making it computationally efficient and often used for text classification and medical diagnosis.

Neural Networks: Deep learning models, particularly neural networks, can be used for heart disease prediction. They are capable of learning complex relationships but mayrequire larger datasets and more computational resources.

Gradient Boosting Algorithms: Algorithms like XGBoost, LightGBM, and CatBoost are popular gradient boosting techniques that sequentially build weak learners to improve predictiveperformance.

Ensemble Methods: Combining multiple models using ensemble methods, such as bagging or boosting, can often lead to improved predictive accuracy.

3.1.1 DRAWBACKS

These systems face challenges such as data quality and reliability, model overfitting, limited generalization, complexity, resource intensiveness, and model interpretability. Addressing these issues is crucial for improving the effectiveness and reliability of heart disease prediction models.

3.2 PROPOSED SYSTEM

Logistic Regression

Advantages: Simple, interpretable, and works well for linearly separable data. Can provideprobabilities for predictions.

Considerations: May not capture complex relationships in the data.

Random Forest

Advantages: Ensemble method that combines multiple decision trees, providing high accuracy and

robustness. Handles non-linearity and interactions between features. Inherently providesfeature importance.

Considerations: Can be computationally expensive, especially with a large number of trees.

Gradient Boosting Algorithms (e.g., XGBoost, LightGBM)

Advantages: Powerful ensemble methods that sequentially build weak learners to improve predictive performance. Handle complex relationships and provide high accuracy.

Considerations: May require careful hyperparameter tuning, and training time can berelatively longer than simpler models.

Support Vector Machines (SVM)

Advantages: Effective in high-dimensional spaces, good for cases where there is a clear margin of separation between classes.

Considerations: May be sensitive to the choice of kernel and parameters, and interpretability can be limited.

K-Nearest Neighbors (KNN):

Advantages: Simple and intuitive, effective for smaller datasets and local patterns.

Considerations: Computationally expensive for large datasets, and may not perform wellwith high-dimensional data.

Naive Bayes:

Advantages: Simple, computationally efficient, and works well with small datasets. Assumes

independence between features.

Considerations: May not capture complex relationships in the data.

Neural Networks:

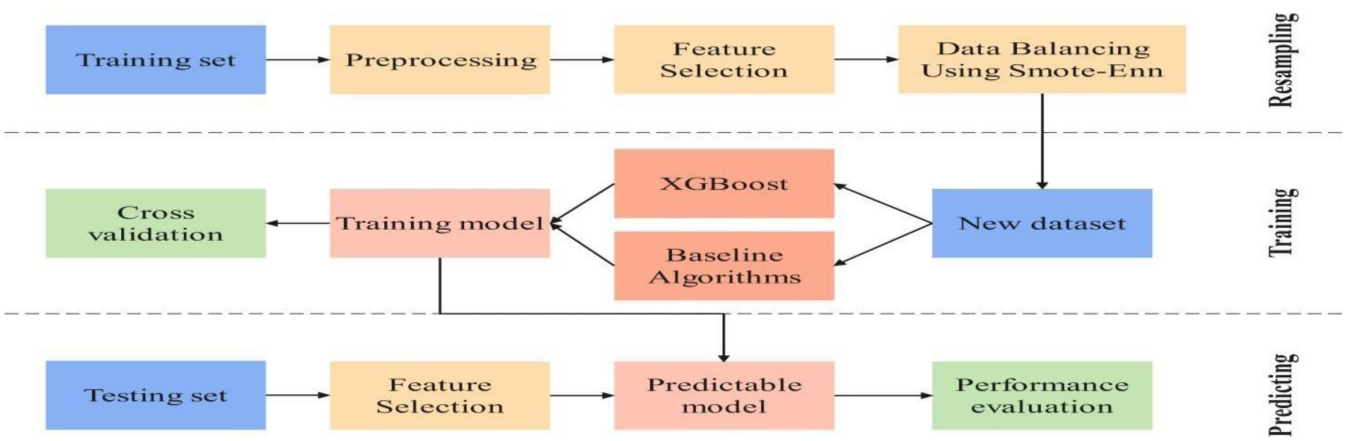
Advantages: Deep learning models can learn complex patterns and relationships. Suitable for large datasets and high-dimensional data.

Considerations: Requires a large amount of data and computational resources. May be considered overkill for smaller datasets.

Best Practices

Ensemble Methods: Random Forest and Gradient Boosting (e.g., XGBoost) are often recommended due to their high predictive performance.

Interpretability: Logistic Regression, Decision Trees, or simpler models with explainability techniques (e.g., SHAP, LIME) may be preferred in situations .



3.2.1 ADVANTAGES

The proposed heart disease prediction system benefits from flexibility and extensibility, a large community and open-source ecosystem, and the availability of pre-trained models. It leverages a rich set of machine learning libraries, robust community support for AI and ML, and ensures scalability and high performance, making it a powerful and adaptable solution for accurate heart disease prediction.

CHAPTER 4

SYSTEM SPECIFICATION

4.1 HARDWARE CONFIGURATION

- Processor (CPU)
- Graphics Processing Unit (GPU)
- Random Access Memory (RAM)
- Storage
- Motherboard
- Power Supply Unit (PSU)
- Cooling
- Case
- Networking

4.2 SOFTWARE CONFIGURATION

- **Python programming language** - Python 3.x installed on the computer/server
- **Operating system** - Windows, Linux, or macOS.
- **Python libraries** such as – opencv, numpy, pandas, tensorflow, Scikit-Learn.
- **Django or Flask (Python):** For developing the backend of a web-based traffic prediction dashboard.
- **HTML/CSS or JavaScript -** for UI design.
- **Integrated Development Environment (IDE):** Use a Python-friendly IDE fordevelopment. Recommended: PyCharm, Jupyter Notebooks, VSCode.

4.3 SOFTWARE DESCRIPTION

To implement a heart disease prediction system, a comprehensive software solution is required. Below is a detailed software solution that includes various components for data processing, machine learning model development, and potential deployment.

COMPONENTS

- **TensorFlow or PyTorch:** These are deep learning frameworks used for building and training neuralnetworks. TensorFlow and PyTorch are widely used for natural language understanding and other machine learning tasks.
- **Scikit-learn:** Scikit-learn is a versatile machine learning library that provides simple and efficient tools for data analysis and modeling. It includes various algorithms for classification, regression, clustering, and more.
- **Pandas:** Pandas is a data manipulation library that provides data structures for efficiently storing and manipulating large datasets. It is commonly used for data preprocessing.
- **NumPy:** NumPy is a fundamental library for scientific computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with mathematical functions to operate on these arrays.
- **Matplotlib and Seaborn:** These libraries are used for data visualization. They can be helpful in creating charts and graphs to present information in a user-friendly way.
- **Beautiful Soup:** Beautiful Soup is a library for web scraping. If your virtual assistant needs to extract information from websites, Beautiful Soup can be useful for parsing HTML and XML documents.

4.4 DEVELOPING ENVIRONMENT

To develop the Heart disease prediction system, you would typically set up the following environment:

- **Python:** Python is the primary programming language used for developing thesystem. Ensure that Python is installed on your system.
- **Integrated Development Environment (IDE):** Choose an IDE for Python development, such as PyCharm, Visual Studio Code, or Jupyter Notebook. These IDEs provide features like code editing, debugging, and project management, enhancing the development process.
- **Install Required Libraries:** Use the Python package manager, pip, to install the necessary libraries such as OpenCV, NumPy, TensorFlow or PyTorch, and Flask. You can install them using the command line interface or directly within your IDE.
- **MongoDB, Cassandra, or InfluxDB:** NoSQL databases suitable for storing and retrieving traffic-related data.
- **Docker**: Containerization allows for consistent deployment across different environments and simplifies dependencies management.
- **Git:** Utilize Git for version control to track changes, collaborate with a team, and manage code history.
- **SQL Alchemy:** A SQL toolkit and Object-Relational Mapping (ORM) libraryfor Python, useful for interacting with relational databases.

CHAPTER 5

ARCHITECTURAL DESIGN

5.1 SYSTEM DESIGN

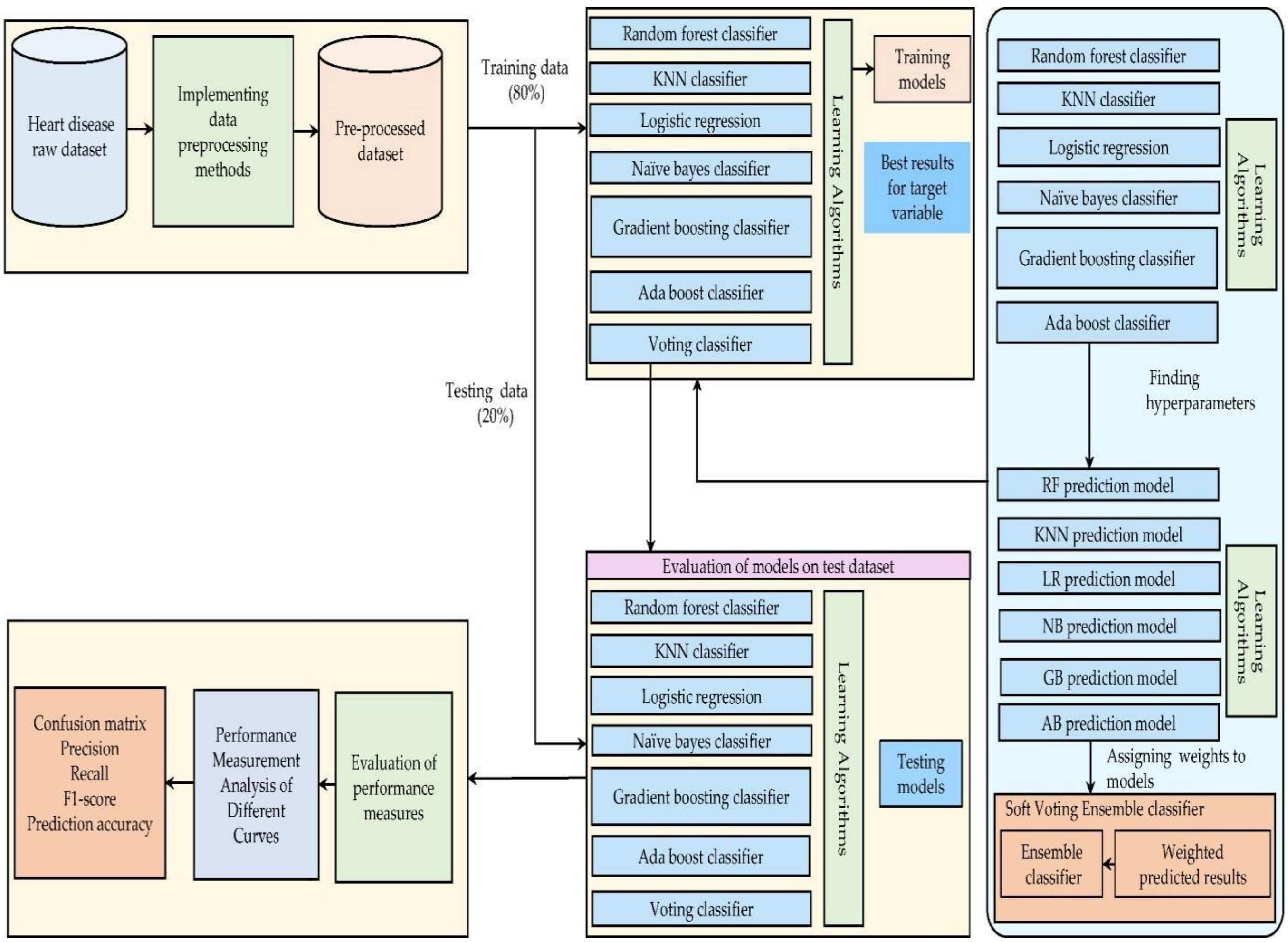


Figure No.5.1.Architectural Diagram

5.2 DATA FLOW DIAGRAM

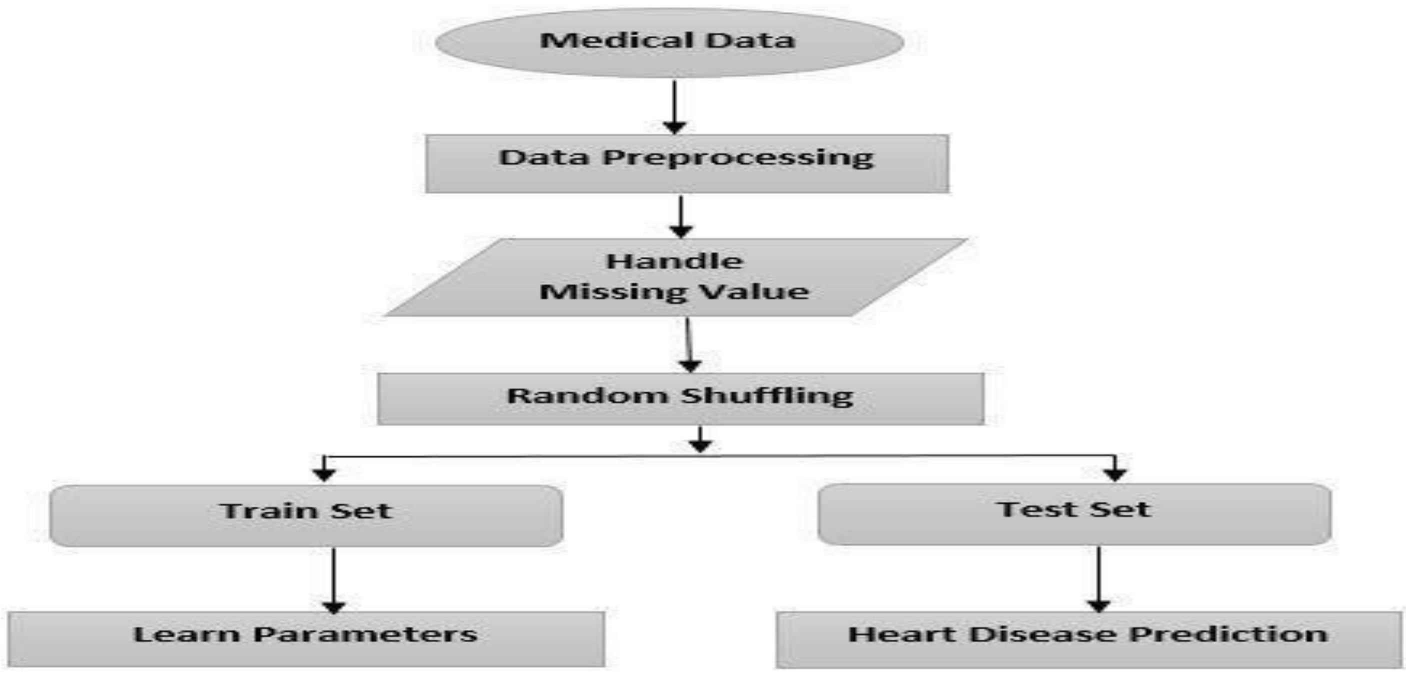


Figure No. 5.2. Data Flow Diagram

5.3 USE CASE DIAGRAM:

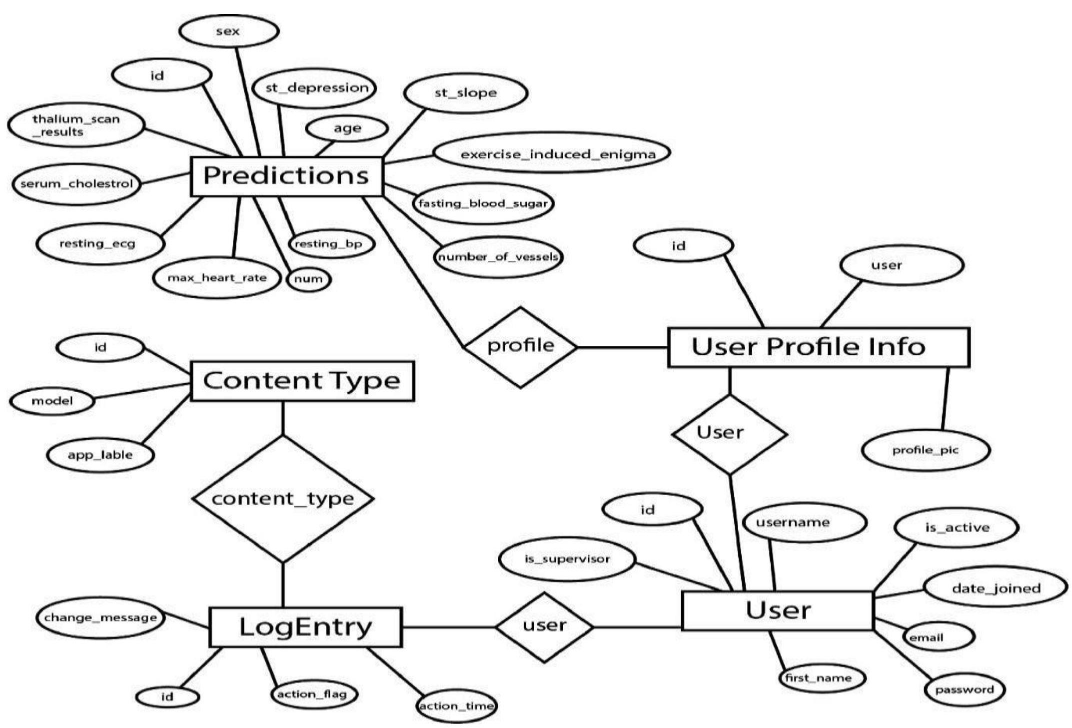


Figure No. 5.3. Use Case Diagram

5.4 ACTIVITY DIAGRAM:

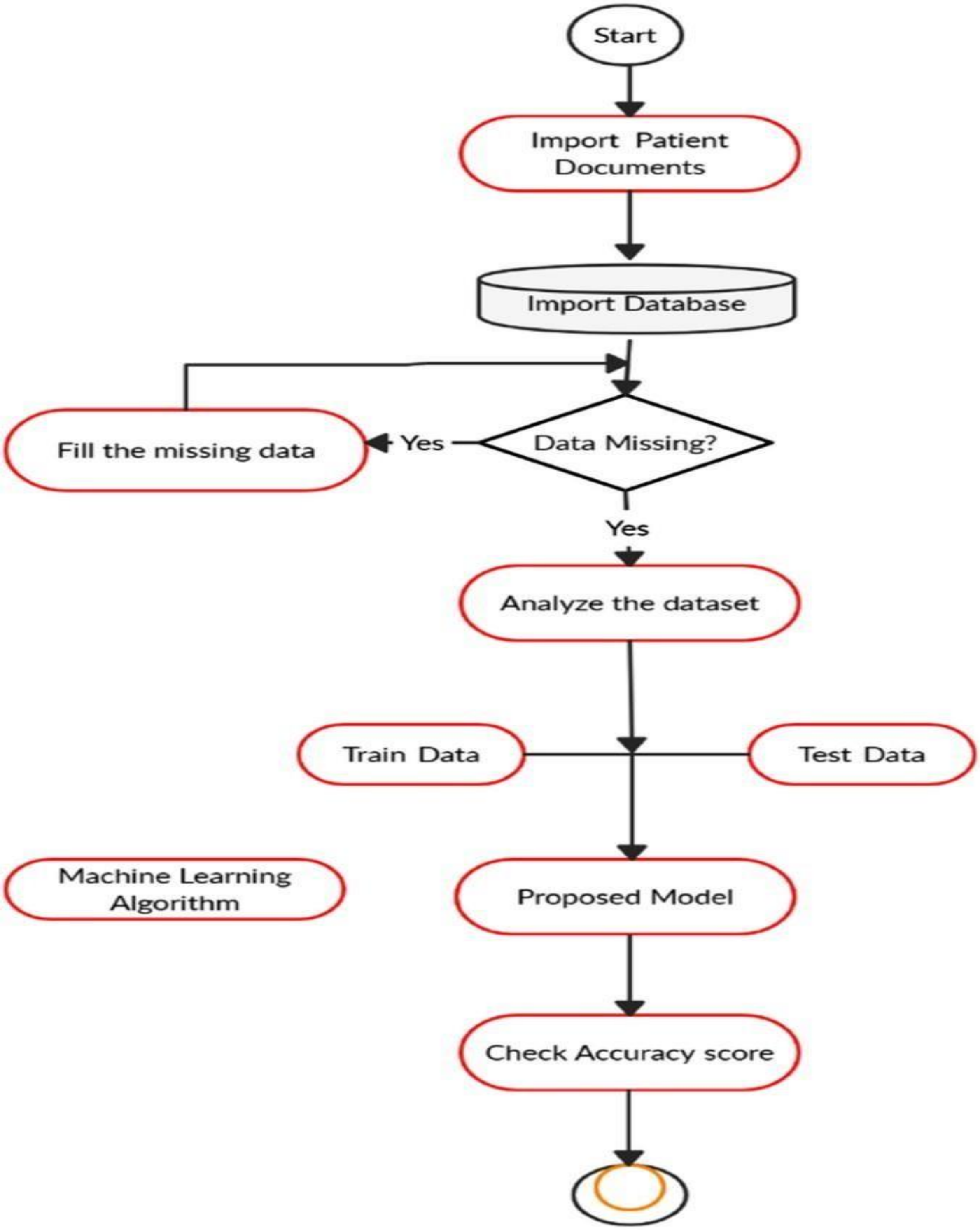


Figure No. 5.4. Activity Diagram

5.5 SEQUENCE DIAGRAM:

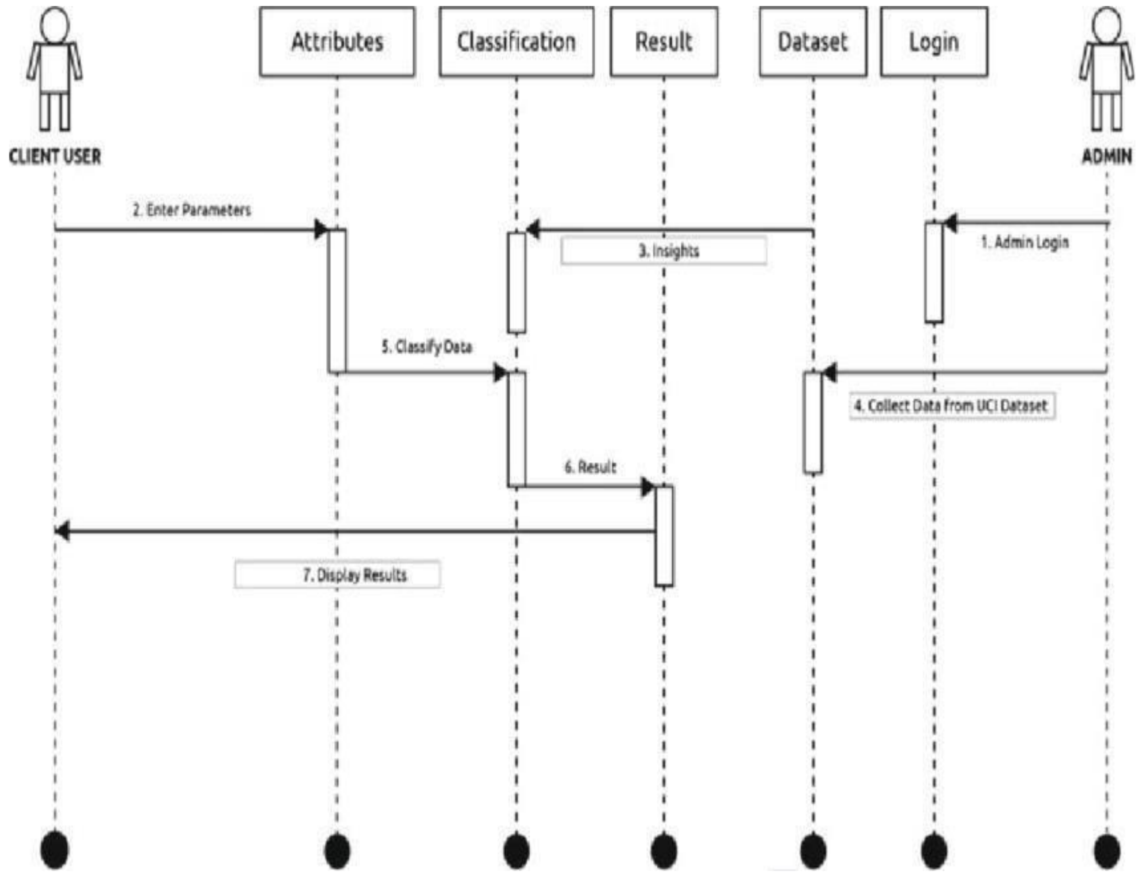


Figure No. 5.5. Sequence Diagram

CHAPTER 6

MODULE DESCRIPTION

6.1 MODULES

6.1.1 DATA RETREATING MODULE

Data Collection: Gather a dataset containing relevant features and target labels (presence or absenceof heart disease). Common datasets for heart disease prediction include the Framingham Heart Study,Cleveland Heart Disease dataset, etc.

Data preprocessing: Handle missing data: Impute or remove missing values. Encode categorical variables: Convert categorical variables into numericalrepresentations (e.g., one-hot encoding). Normalize/ Standardize numerical features: Scale numerical features to a standardrange.

Data Exploration: Perform exploratory data analysis (EDA) to understand the distribution of data,correlations, and potential patterns.

Feature Selection: Identify relevant features for prediction. Techniques like correlation analysis or feature importance from models can help.

6.1.2 WEB INTERFACE MODULE

When designing a web interface for heart disease prediction, We need several components to create a user-friendly and effective application. Here are the key components you should consider including:

Input Form: Create a form to collect user input for relevant features used by the machine learning model. Common features for heart disease prediction include age, sex, cholesterol levels, blood pressure, and more.

Validation: Implement client-side and server-side validation to ensure that the user inputs are accurate and within acceptable ranges.

Submit Button: Include a button that users can click to submit the input data for prediction.

Loading Indicator: Display a loading indicator to inform users that the prediction is being processed,especially if there is a delay.

Results Display: Show the prediction results to the user. This can include whether the individual is at risk of heart disease, the probability score, or any other relevant information.

Visualization: Consider using charts or graphs to visually represent the prediction results. This canenhance user understanding.

6.1.3 INFO PREDICTION MODULE

It seems like you're asking about an information prediction module in the context of aheart disease prediction system. If I understand correctly, you may be interested in a component that provides information or explanations about the prediction results. Here's a breakdown of what such a module could entail:

Explanation Generation: Develop a module that generates human-readable explanations for the model's predictions. This can help users understand why the model made a particular prediction.

Consider using techniques like LIME (Local Interpretable Model-agnostic Explanations) or SHAP (SHapley Additive exPlanations) to provide interpretability to complex models.

6.1.4 RELATIONAL MODULE

Patient Data Relationships: Establish relationships between different pieces of patient data. For example, connectage with cholesterol levels, blood pressure, and other relevant factors.

Data Integration: Ensure that the relational module integrates and processes data from various sources.

Correlation Analysis: Implement features that allow users to explore correlations between different variables.This can help users understand how changes in one variable might affect the prediction.

Temporal Relationships: If your dataset includes temporal data, consider incorporating a module that visualizeshow heart disease risk factors evolve over time for an individual.

User Profiles: Allow users to create profiles or accounts where their health data is stored. This canenable the system to establish relationships and patterns unique to each user.

CHAPTER 8

CONCLUSION AND FUTURE ENHANCEMENT

8.1 CONCLUSION

Developing a heart disease prediction model using machine learning holds significant promise in enhancing early diagnosis and intervention, thereby improving patient outcomes. The use of diverse algorithms, data preprocessing techniques, and model interpretability tools allows for the creation of robust and effective systems. However, it's crucial to address challenges such as data quality, interpretability, ethical considerations, and model generalization to diverse populations.

The implementation of machine learning models in healthcare requires a collaborativeapproach, involving domain experts, healthcare professionals, and data scientists. Continuous monitoring, model updates, and adherence to privacy and security standards are essential components for successful deployment in real-world clinical settings.

FUTURE ENHANCEMENT

Integration with Electronic Health Records (EHR): Future models can benefit from integrating with electronic health records to leveragea broader range of patient data. This could enhance the accuracy and reliability of predictions.

Advanced Explainability Techniques: Developing more advanced explainability techniques can further enhance the interpretability of models, fostering trust among healthcare professionals. Future models may move towards personalized medicine by considering individual patient characteristics, genetic factors, and lifestyle data to tailor predictions and treatment plans.

Ensemble of Models: Combining predictions from multiple models (ensemble learning) can potentially improve overall performance and robustness, especially when dealing with diverse patient populations.

Real-time Monitoring and Wearable Devices: Integrating machine learning models with wearable devices and real-time monitoringsystems can enable continuous health tracking, allowing for early detection and intervention.

Explainable AI in Healthcare Regulations: As the use of AI in healthcare becomes more prevalent, there is a need for regulations that address the ethical use, interpretability, and transparency of these models to ensure patient safety and privacy.

Population-specific Models: Developing models tailored to specific populations or demographic groups can improve the generalization of predictions across diverse communities.

Longitudinal Data Analysis: Incorporating longitudinal data analysis could provide insights into the progression of heart disease over time, enabling more accurate predictions and proactive interventions.

Integration with Clinical Decision Support Systems: Seamless integration with clinical decision support systems can empower healthcareprofessionals by providing actionable insights and recommendations based on modelpredictions.

Global Collaboration and Data Sharing: Facilitating global collaboration and data sharing initiatives can help create more robust models by leveraging diverse datasets from different regions and demographics.

In conclusion, the future scope of heart disease prediction models in machine learning is expansive and holds great potential for revolutionizing healthcare practices.

APPENDIX 1 SAMPLE CODE

```
# MODEL TRAINING

import pandas as pd
import pickle
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
# loading and reading the dataset
heart = pd.read_csv("C:\\Users\\Thinesh\\Desktop\\heart_cleveland_upload.csv")
# creating a copy of dataset so that will not affect our original dataset.
heart_df = heart.copy()
# Renaming some of the columns
heart_df = heart_df.rename(columns={'condition':'target'})
print(heart_df.head())
x= heart_df.drop(columns= 'target')
y= heart_df.target
# splitting our dataset into training and testing for this we will use
#train_test_split library.
x_train, x_test, y_train, y_test= train_test_split(x, y, test_size= 0.25, random_state=42)
#feature scaling
scaler= StandardScaler()
x_train_scaler= scaler.fit_transform(x_train)
x_test_scaler= scaler.fit_transform(x_test)
# creating K-Nearest-Neighbor classifier
model=RandomForestClassifier(n_estimators=20)
model.fit(x_train_scaler, y_train)
y_pred=model.predict(x_test_scaler)
```

```
p =model.score(x_test_scaler,y_test)
print(p)
print('Classification Report\n', classification_report(y_test, y_pred))
print('Accuracy: {}%\n'.format(round((accuracy_score(y_test, y_pred)*100),2)))

# Save the trained model and scaler to pickle files
with open('heart_model.pkl', 'wb') as model_file:
    pickle.dump(model, model_file)

with open('scaler.pkl', 'wb') as scaler_file:
    pickle.dump(scaler, scaler_file)
```

HTML CODE

```
<!DOCTYPE html>
<html>
<head>
    <title>Heart Disease Prediction</title>
    <link rel="stylesheet" href="{{ url_for('static', filename='css/styles.css') }}">
    <style>
        body {
            font-family: Arial, sans-serif;
            background-image: url("C:\\Users\\Downloads\\desktop-wallpaper-human-heart-10-
thumbnail.jpg");
            background-size: cover;
            background-position: center;
            background-repeat: no-repeat;
            margin: 0;
            padding: 0;
        }
        .container {
            width: 100%;
            max-width: 600px;
            margin: 50px auto; /* Adjust the top margin as needed */
            padding: 20px;
            background-color: rgba(0, 0, 0, 0.5);
            border-radius: 10px;
```

```
    box-shadow: 0 2px 5px rgba(0, 0, 0, 0.3);
}
h1 {
    text-align: center;
    margin-bottom: 20px;
    color: white; /* Text color */
}
form {
    margin-bottom: 20px;
}
label {
    display: block;
    margin-bottom: 5px;
    font-weight: bold;
    color: white; /* Text color */
}
input[type="text"],
input[type="radio"] {
    width: 100%;
    padding: 10px;
    margin-bottom: 10px;
    border: 1px solid #ccc;
    border-radius: 5px;
    box-sizing: border-box;
}
input[type="submit"] {
    width: 100%;
    padding: 10px;
    background-color: #ff4d4d; /* Red color for button */
    color: white;
    border: none;
    border-radius: 5px;
    cursor: pointer;
    opacity: 0.8;
}
input[type="submit"]:hover {
    background-color: #ff3333; /* Darker red on hover */
```

```
            opacity: 1;
        }
        .result {
            margin-top: 20px;
            font-weight: bold;
            color: white; /* Text color */
        }
    </style>
</head>
<body>
    <div class="container">
        <h1>Heart Disease Prediction</h1>
        <form method="POST" action="/">
            <label for="age">Age (age):</label>
            <input type="text" id="age" name="age" required>

            <label for="sex">Sex (sex):</label><br>
            <input type="radio" id="male" name="sex" value="1" required>
            <label for="male">Male</label>
            <input type="radio" id="female" name="sex" value="0">
            <label for="female">Female</label>

            <label for="cp">Chest Pain Type (cp):</label>
            <input type="text" id="cp" name="cp" required>

            <label for="trestbps">Resting Blood Pressure (trestbps):</label>
            <input type="text" id="trestbps" name="trestbps" required>

            <label for="chol">Serum Cholesterol (chol):</label>
            <input type="text" id="chol" name="chol" required>

            <label for="fbs">Fasting Blood Sugar (fbs):</label>
            <input type="text" id="fbs" name="fbs" required>

            <label for="restecg">Resting Electrocardiographic Results (restecg):</label>
            <input type="text" id="restecg" name="restecg" required>
```

```
            <label for="thalach">Maximum Heart Rate Achieved (thalach):</label>
            <input type="text" id="thalach" name="thalach" required>

            <label for="exang">Exercise Induced Angina (exang):</label>
            <input type="text" id="exang" name="exang" required>

            <label for="oldpeak">ST Depression Induced by Exercise (oldpeak):</label>
            <input type="text" id="oldpeak" name="oldpeak" required>

            <label for="slope">Slope of Peak Exercise ST Segment (slope):</label>
            <input type="text" id="slope" name="slope" required>

            <label for="ca">Number of Major Vessels Colored by Fluoroscopy (ca):</label>
            <input type="text" id="ca" name="ca" required>

            <label for="thal">Thalassemia (thal):</label>
            <input type="text" id="thal" name="thal" required>

            <input type="submit" value="Predict">
        </form>

        {% if result %}
        <div class="result">
            <h2>{{ result }}</h2>
        </div>
        {% endif %}
    </div>
</body>
</html>
```

#FLASK APPLICATION

```
from flask import Flask, render_template, request
import pandas as pd
from sklearn.preprocessing import StandardScaler
import pickle
import os
from flask import Flask, render_template, request, redirect, url_for

app = Flask(__name__, static_url_path='/static', static_folder='static')

model_path = os.path.join("C:\\Users\\Thinesh\heart_model.pkl")
scaler_path = os.path.join("C:\\Users\\Thinesh\\scaler.pkl")

# Load the trained model and scaler
model = pickle.load(open(model_path, 'rb'))
scaler = pickle.load(open(scaler_path, 'rb'))

@app.route('/', methods=['GET', 'POST'])
def index():
    if request.method == 'POST':
        age = float(request.form['age'])
        sex = int(request.form['sex'])
        cp = int(request.form['cp'])
        trestbps = float(request.form['trestbps'])
        chol = float(request.form['chol'])
        fbs = int(request.form['fbs'])
        restecg = int(request.form['restecg'])
        thalach = float(request.form['thalach'])
        exang = int(request.form['exang'])
        oldpeak = float(request.form['oldpeak'])
        slope = int(request.form['slope'])
        ca = int(request.form['ca'])
        thal = int(request.form['thal'])

        # Create a DataFrame from user input
```

```
        user_data = pd.DataFrame([[age, sex, cp, trestbps, chol, fbs, restecg, thalach,
exang, oldpeak, slope, ca, thal]],
                                columns=['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalach',
'exang', 'oldpeak', 'slope', 'ca', 'thal'])

        # Scale the user data
        user_data_scaled = scaler.transform(user_data)

        # Predict using the trained model
        prediction = model.predict(user_data_scaled)

        if prediction[0] == 1:
            result = "The model predicts that you have a risk of heart disease."
        else:
            result = "The model predicts that you do not have a risk of heart disease."

        return render_template('index.html', result=result)

    return render_template('index.html', result=None)

if __name__ == '__main__':
    app.run(debug=True)  # Change debug=False for production us
```

APPENDIX 2 SCREENSHOTS

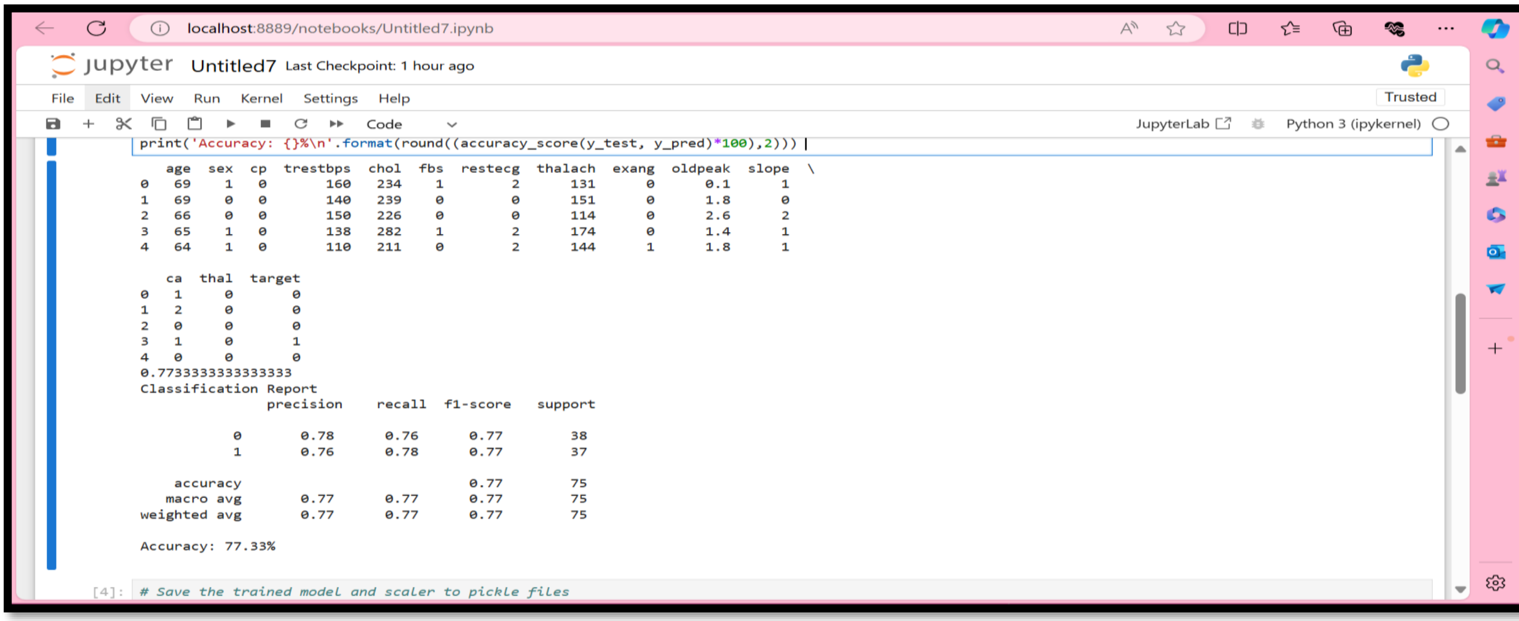


Figure No.A.2.1.Model Training

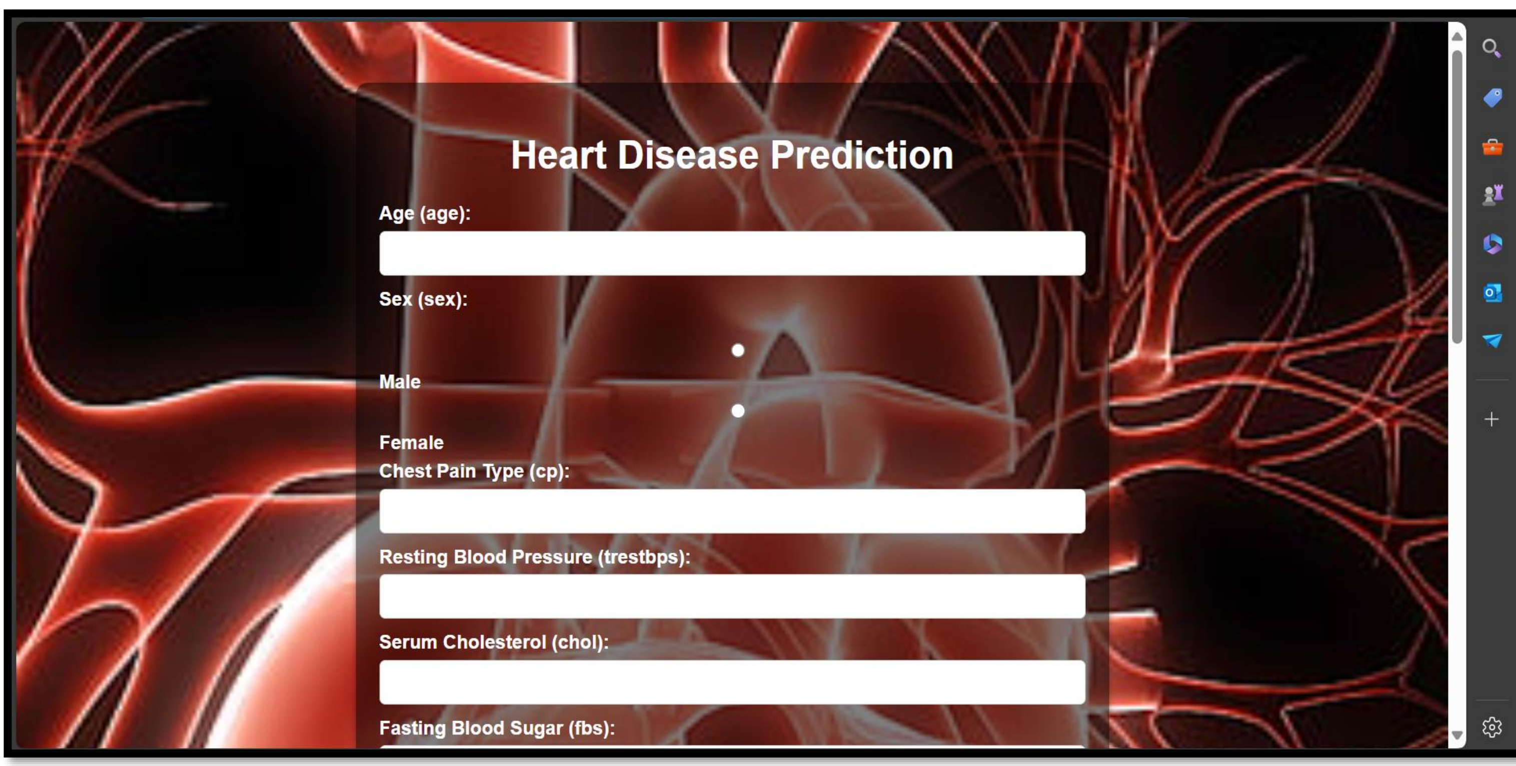


Figure No.A.2.1.Heart Disease Prediction

REFERENCES

1. Apurb Rajdhan, Avi Agarwal, Milan Sai, Dundigalla Ravi and Poonam Ghuli, "Heart Disease Prediction using Machine Learning", International Journal Of Engineering Research Technology(IJERT), vol. 09, no. 04, April 2020.
2. Baban Rindhe, Nikita Ahire, Patil, Rupali Gagare, Shweta Darade and Manisha, "Heart Disease Prediction Using Machine Learning", International Journal of Advanced Research in ScienceCommunication and Technology, 2021.
3. G. Choudhary and S. Narayan Singh, "Prediction of Heart Disease using Machine Learning Algorithms", International Conference on Smart Technologies in Computing Electrical and Electronics (ICSTCEE), 2020.
4. Gavhane, G. Kokkula, I. Pandya and K. Devadkar, "Prediction of Heart Disease Using Machine Learning", Second International Conference on Electronics Communication and Aerospace Technology (ICECA), 2018.
5. P. Ramprakash, R. Sarumathi, R. Mowriya and S. Nithya Vishnupriya, "Heart Disease Prediction Using Deep Neural Network", International Conference on Inventive Computation Technologies (ICICT), 2020.
6. M. S. Raja, M. Anurag, C. P. Reddy and N. R. Sirisala, "Machine Learning Based Heart Disease Prediction System", International Conference on Computer Communication and Informatics (ICCCI), 2021.