A Seminar Report On

Heart Disease Prediction Using Machine Learning: A Data-Driven Approach

Submitted in partial fulfilment of the requirements of the degree of

Master Of Computer Application

By

Omkar Rajendra Patil

(232010045)

2024-2025

Under Guidance Of

Prof. Nikhil Khandare

Department of Master of Computer Applications
Veermata Jijabai Technological Institute

(An Autonomus Institute Affliated to University of Mumbai)
Mumbai – 400 019

STATEMENT OF CANDIDATE

I state that work embodied in this Seminar titled **Heart Disease Prediction Using Machine Learning: A Data-Driven Approach,** forms my own contribution of work under the guidance **Prof. Nikhil Khandare** at the Department of Masters of Computer Application, Veermata Jijabai Technological Institute. The report reflects the work done during the period of candidature but may include related preliminary material provided that it has not contributed to an award of previous degree. No part of this work has been used by me for the requirement of another degree except where explicitly stated in the body of the text and the attached statement.

Omkar Rajendra Patil

MCA [232010045]

APPROVAL SHEET

This Project Report entitled **Heart Disease Prediction Using Machine Learning: A Data-Driven Approach,** by **Omkar Rajendra Patil** is approved for the degree of Master of Computer Application.

Prof. Nikhil Khandare	Examiner
Guide/Supervisor	
Date:15/07/2025	Date:15/07/2025
Place: Mumbai	Place: Mumbai

Certificate

This is to certify that **Omkar Rajendra Patil**, [232010045], a student of MCA Veermata Jijabai Technological Institute (VJTI), Mumbai have successfully completed the seminar titled titled **Heart Disease Prediction Using Machine Learning: A Data-Driven Approach,** under the guidance of **Prof. Nikhil Khandare.**

Prof. Nikhil Khandare Guide/Supervisor	Prof. Archana Pai HOD
Professor	MCA Department
Date: 15/07/2025 Place: Mumbai	Date:15/07/2025 Place: Mumbai

Certificate

This is to certify that **Omkar Rajendra Patil**, [232010045], a student of MCA Veermata Jijabai Technological Institute (VJTI), Mumbai have successfully completed the seminar titled **Heart Disease Prediction Using Machine Learning: A Data-Driven Approach,**under the guidance of **Prof. Nikhil Khandare.**

Prof. Nikhil Khandare Guide/Supervisor	Examiner
Date: 15/07/2025 Place: Mumbai	Date:15/07/2025 Place: Mumbai

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources.

I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea / data / fact / source in my submission.

I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Signature of the Student

Omkar Rajendra Patil

Roll No. 232010045

ACKNOWLEDGEMENT

I would like to thank all those people whose support and cooperation has been an invaluable asset during the course of this Technical Seminar Report. It would have been impossible to complete the Technical Seminar Report without their support, valuable suggestions, criticism, encouragement and guidance.

I convey my gratitude for the help and encouragement in all aspects to my guide Prof. Nikhil Khandare. Her expertise and patience were greatly appreciated and assisted in the successful completion of this Internship Project Report.

I am also grateful for all other teaching and non-teaching staff members of the MCA department for directly or indirectly helping us for the completion of this Technical Seminar Report and the resources provided.

Signature of the Student

Omkar Rajendra Patil

Roll No. 232010045

(1) Abstract:

Heart disease is one of the biggest reasons people die around the world. Finding heart disease early is very important because it can help doctors treat the patient in time and reduce the cost of healthcare. Early and correct diagnosis helps in preventing serious problems. This study uses machine learning (ML) to predict if a person has heart disease by looking at the person's health data. The main data used in this study is the Cleveland Heart Disease dataset. Many researchers in healthcare and machine learning have used this dataset to understand heart disease and how to predict it. It includes important health information like the person’s age, sex, chest pain type, blood pressure, cholesterol levels, blood sugar, whether exercise causes chest pain, and the maximum heart rate. These details help us understand how healthy a person’s heart is.

The study starts by looking closely at the dataset to understand its structure and how the different features are related. The data is checked through visual tools, statistics, and methods that show how different factors are connected. Any patterns or unusual things in the data are noticed. After that, the data is cleaned up by fixing missing values, converting categories into numbers, and adjusting the data so it can work with machine learning models.

The main part of the study is about creating models to predict if a person will have heart disease. For this, two machine learning models, Random Forest and XGBoost (XGB), are used. These models are good at handling complicated data and making sense of noisy information. They also give results that are easy to understand. Both models are trained using the Cleveland dataset, and their performance is measured using accuracy, precision, recall, and F1-score to check how well they can predict heart disease.

The research also looks at which factors are most important in predicting heart disease. The most important factors found are the

type of chest pain, exercise-related chest pain, and maximum heart rate. These features are very helpful in predicting heart disease. This information is useful for doctors and healthcare workers because knowing what factors are most important can help them make better decisions when diagnosing heart problems.

This study shows how machine learning can help in finding heart disease early, leading to better healthcare solutions. By using real-world data and machine learning, this research can help doctors diagnose heart problems more accurately. The results of this study can reduce healthcare costs and improve the health of patients, making it easier to manage and treat heart diseases.

Table of content

(2) Introduction

Heart disease is one of the main causes of death around the world. According to the World Health Organization (WHO), about 17 million people die each year from heart-related diseases, which makes up about 32% of all deaths globally. Among these, heart disease is the most common.

The number of people getting heart disease is growing because of factors like getting older, eating unhealthy food, obesity, smoking, and high blood pressure. As heart disease continues to increase, it’s very important to detect it early and predict it correctly to help reduce the number of deaths and improve patient care.

In the past, diagnosing heart disease required tests like angiography, which are expensive and need skilled doctors to interpret the results. But machine learning (ML) provides a simpler, non-invasive way to predict heart disease early by analyzing large amounts of data.

This makes it a better, cheaper option compared to traditional methods. By quickly analyzing complex data patterns, ML can find early signs of heart disease that doctors might miss.

Machine learning has shown a lot of potential in healthcare, especially for predicting heart disease.

For example, studies using the Cleveland Heart Disease Dataset have shown that machine learning algorithms like Random Forest and Gradient Boosting can predict heart disease with over 78% to 80% accuracy.

These algorithms are good at understanding complex relationships in data, making them powerful tools for heart disease prediction. Additionally, using methods like "soft voting," which combines multiple algorithms, has improved the prediction accuracy to 82%, which is better than traditional diagnostic methods.

The Cleveland Heart Disease Dataset is commonly used in heart disease research. It contains data from 920 patients with 76 features. Many studies focus on 14 important features that are most useful for predicting heart disease.

These features include age, gender, cholesterol levels, blood pressure, maximum heart rate, and exercise-induced chest pain. These factors give us valuable information about heart health and improve the accuracy of machine learning models.

Some important features in the dataset are:

1)Age:

Older people are generally at higher risk of heart disease.

2)Gender:

Men are more likely to get heart disease at younger ages.

3)Cholesterol levels:

High cholesterol levels can cause artery problems and heart disease.

4)Blood pressure: High blood pressure is a major risk for heart disease.

5)Maximum heart rate:

A lower heart rate during exercise can indicate heart problems.

6)Exercise-induced chest pain:

Chest pain during exercise shows that the heart might not be reacting well to stress.

These features help in predicting heart disease more accurately and also help doctors understand why heart disease happens.

By studying these factors, doctors can identify people who are at high risk and create better prevention plans.

Machine learning models like XGBoost, Random Forest, and Logistic Regression have shown prediction accuracies between 78% and 80%. These models give more reliable predictions than traditional methods and can lower healthcare costs by detecting heart disease early.

Also, wearable devices that monitor heart health in real time are benefiting from these machine learning models, which help manage heart disease better.

But there are challenges in using machine learning for predicting heart disease. One big issue is the imbalance in data.

There are more healthy people in the dataset than people with heart disease, which can lead to biased predictions.

To fix this, techniques like oversampling (adding more data for people with heart disease) and cost-sensitive learning are used. The quality of data is also important; missing or incorrect data can make predictions less accurate.

Preprocessing the data, like fixing missing values and finding outliers, is important to make sure the predictions are correct.

Another challenge is that machine learning models are often seen as “black boxes,” meaning we don’t fully understand how they make their decisions.

This can make doctors hesitate to trust the results. There is ongoing research into "interpretable machine learning" (IML), which tries to make these models more understandable and easier to trust in medical settings.

The goal of this research is to make heart disease prediction more accurate using machine learning. By using good data preprocessing techniques, exploring the data carefully, and applying powerful algorithms like Random Forest and XGBoost, this study aims to find the key factors that contribute to heart disease risk. The results will provide doctors with useful information to diagnose and prevent heart disease earlier.

In conclusion, this research wants to show how machine learning can work alongside traditional methods to improve heart disease management. By using data-driven solutions, machine learning can make diagnosis more efficient and accurate, leading to better patient care and a reduction in the number of heart disease deaths globally.

(3) Literature survey

In this project, we studied 12 research papers related to heart disease prediction using machine learning.

Each paper gave us useful knowledge about how to clean the data, choose the important features, use the best model, and improve the accuracy of heart disease prediction.

Below is the summary of each research paper in simple English.

1) Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization

This paper explains how to make heart disease prediction more accurate.
It focuses on solving problems like missing data and selecting the best features.
The authors used different methods like KNN and MICE to fill in missing data.
They selected important features using PCA, SVM, and PSO.
Models like Naive Bayes, Random Forest, and MLP were used

together.
To improve performance, GridSearchCV was used for tuning the settings of models.
It was found that Random Forest worked better than Logistic Regression and SVM when tuned properly.

Technologies Used

KNN, MICE, PCA, SVM, PSO, Naive Bayes, Random Forest, MLP, GridSearchCV
Meaning of keywords:

KNN: Finds similar data points.

PCA: Reduces data size while keeping useful parts.

SVM: Separates groups of data.

PSO: Selects best features like birds finding food.

GridSearchCV: Tests many settings to find the best one.

2) Heart Disease Detection Using Machine Learning

This paper explains how machine learning helps in detecting heart disease.
It uses the Cleveland dataset that contains patient information.
Models like Decision Tree, KNN, and SVM were used for prediction.

Seaborn library was used to draw graphs and charts.
A confusion matrix was used to check how many predictions were right or wrong.
KNN gave good results.
The paper shows that ML models are better than old methods.

Technologies Used

Decision Tree, KNN, SVM, Cleveland Dataset, Seaborn, Confusion Matrix
Meaning of keywords:

Decision Tree: A tree-like model used to make decisions.

Confusion Matrix: A table to check model performance.

3) Effective Heart Disease Prediction Using Machine Learning Techniques

This paper helps in predicting heart disease early to save costs and improve treatment.
Important risk factors like cholesterol, BP, and diabetes are identified using ML.
Models like XGBoost, Random Forest, and SVM are compared.
XGBoost gave the best results when features were selected properly.

But the paper also warns that small datasets can give poor results.
The paper tells that proper feature selection is very important.
Technologies Used

XGBoost, Random Forest, SVM
Meaning of keywords:

XGBoost: A powerful and fast prediction model.

Overfitting: When a model only remembers training data and fails on new data.

4) Heart Disease Prediction Using Machine Learning Algorithms

This paper compares different ML models for predicting heart disease.
The models tested were Logistic Regression, KNN, and Random Forest.
KNN gave the best accuracy among these models.
It also tested Quantum Neural Network, which gave very high accuracy.
This model gave better results than regular machine learning.
This shows the use of new technology in medical prediction.

Technologies Used

Logistic Regression, KNN, Random Forest, Quantum Neural Network
Meaning of keywords:

Logistic Regression: A model for yes/no type questions.

Quantum Neural Network: An advanced model using quantum rules.

5) Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm

This paper used Jellyfish Optimization Algorithm to select important features.
It helped improve the performance of models like SVM and Random Forest.
Models with JOA gave better accuracy than normal models.
JOA helped find the best features from the data.
This reduced unnecessary data and improved model speed and accuracy.

Technologies Used

Jellyfish Optimization Algorithm, SVM, Random Forest
Meaning of keywords:

JOA: Works like jellyfish movement to find the best features.

6) A Proposed Technique for Predicting Heart Disease Using Machine Learning Algorithms and an Explainable AI Method

This paper used Explainable AI (XAI) with ML models to explain how predictions are made.
This is useful for doctors to understand why the model predicted heart disease.
The study used models like XGBoost and Random Forest.
XAI helped make predictions clear and easy to understand.
XGBoost with XAI gave better results than other models.

Technologies Used

XGBoost, Random Forest, Explainable AI (XAI)
Meaning of keywords:

XAI: Makes machine learning easier to explain to humans.

7) Heart Disease Detection Using Machine Learning Methods: A Comprehensive Narrative Review

This paper reviews many studies about ML and heart disease.
It explains how data should be cleaned and prepared before using it.

Steps like normalization and splitting data are explained.
It also talks about how to check performance using accuracy, AUC, and confusion matrix.
This paper helped us understand common steps used in all ML projects.

Technologies Used

Data Cleaning, Normalization, AUC, Confusion Matrix
Meaning of keywords:

AUC: Tells how well the model separates heart disease and no heart disease.

8) Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques

This paper focused on predicting heart disease early using feature selection methods.
PCA, mutual information, GA, and PSO were used for selecting best features.
The selected features were used to train models, and results improved.
Early detection helps save lives and reduce treatment costs.

This study showed that selecting the right data columns is very important.

Technologies Used

PCA, Mutual Information, Genetic Algorithm (GA), Particle Swarm Optimization (PSO)
Meaning of keywords:

GA: Chooses best options like evolution.

Mutual Info: Tells how two features are related.

9) Classification and Prediction of Heart Diseases Using Machine Learning Algorithms

This paper used classification models like Naive Bayes, KNN, and Decision Tree.
A combination of these models gave the best results.
This is called a hybrid model.
The study found that Decision Tree + KNN worked better together.
Using two models together improved prediction accuracy.

Technologies Used

Naive Bayes, KNN, Decision Tree, Hybrid Model (KNN + DT)
Meaning of keywords:

Hybrid Model: A mix of two or more models.

10) Advancements in Heart Disease Prediction: A Machine Learning Approach for Early Detection and Risk Assessment

This paper used advanced ML models to detect heart disease early.
Deep learning and ensemble models were used to make predictions.
It suggested using electronic health records (EHR) with ML models.
This helps give real-time and correct predictions.
It also discussed future directions in ML-based healthcare.

Technologies Used

Deep Learning, Ensemble Models, Electronic Health Records (EHR)
Meaning of keywords:

EHR: Electronic health information of patients.

Ensemble: Using many models together to improve results.

11) Ensemble Framework for Cardiovascular Disease Prediction

This paper used models like Random Forest, SVM, and Gradient Boosting.

It combined these using stacking and boosting.

To balance the data, SMOTE was used.

The study found that combining models gave more accurate results.

It also showed that balanced data improves model performance.

Technologies Used

Random Forest, SVM, Gradient Boosting, SMOTE, Boosting, Stacking

Meaning of keywords:

SMOTE: Adds new samples to balance healthy and sick people in data.

12) Heart Disease Risk Prediction Using Deep Learning Techniques with Feature Augmentation

This paper used deep learning models like CNN and LSTM.

It added more features to improve model understanding.

CNN and LSTM gave better results than normal ML models.

These models worked well on large data.

It showed that deep learning is helpful in healthcare.

Technologies Used

CNN, LSTM, Deep Learning, Feature Augmentation

Meaning of keywords:

CNN: Finds patterns in data.

LSTM: Remembers data over time like memory.

13)Conclusion of Literature Survey

All 12 papers gave useful ideas to improve our project.
We learned that:

Models like Random Forest, XGBoost, KNN, and SVM are commonly used.

Handling missing data, selecting features, and tuning model settings is important.

New ideas like XAI, deep learning, and ensemble models give better results.
This helped us choose the best models and methods for our project.

Some papers showed higher accuracy (above 90%) than our model (83%).
This may be because they used bigger datasets, advanced models like deep learning, or feature selection methods.
Even with 83% accuracy, our model gave good results and helped us learn machine learning clearly.

(4) Motivation

Heart disease is one of the main reasons why people die all over the world. According to the World Health Organization (WHO), about 17 million people die each year because of heart-related diseases, and this makes up around 32% of all deaths globally.

Many people are getting heart disease because of factors like getting older, eating unhealthy food, smoking, and high blood pressure.

As heart disease continues to increase, it becomes very important to find a way to predict it early and help people before it’s too late.

In the past, diagnosing heart disease required expensive tests, like angiography, which need skilled doctors to interpret the results. These tests are costly and take time.

But now, with the help of machine learning (ML), we can predict heart disease in a simpler, faster, and cheaper way.

Machine learning can analyze a large amount of data very quickly and find patterns that can show early signs of heart disease.

It’s a non-invasive and easy way to predict if someone might have heart disease, without needing expensive or complicated tests.

Machine learning has shown a lot of promise in healthcare, especially for predicting heart disease.

Studies using datasets like the Cleveland Heart Disease Dataset have shown that machine learning algorithms, such as Random Forest and XGB Classifier, can predict heart disease with very high accuracy.

These algorithms are good at understanding the complicated relationships in the data, which makes them powerful tools for predicting heart disease.

Using different machine learning algorithms together, known as "soft voting," has even improved the accuracy of predictions to 82%, which is much better than traditional methods.

This is why I decided to work on this topic. I want to show how machine learning can be used to predict heart disease early and help doctors diagnose it faster.

By using the Cleveland Heart Disease Dataset, we can look at important features like age, blood pressure, chest pain types, cholesterol levels, and heart rate.

These factors are very useful in predicting heart disease, and they can help us make better predictions using machine learning models.

The motivation for this study is to help doctors and healthcare professionals make better decisions in diagnosing heart disease.

Machine learning can improve heart disease prediction by quickly analyzing data and giving early warnings, which can save lives.

If this study shows that machine learning can predict heart disease accurately, it could make a big difference in healthcare, helping more people get the treatment they need in time.

(5) Problem Statement

1. Traditional Methods Have Limitations

In the past, doctors have relied on their experience, patient history, and simple tests like ECG, blood pressure, and cholesterol checks to diagnose heart disease.

These methods have been helpful, but they also have some big problems.

One of the biggest issues is that these methods can sometimes give wrong or delayed results because of mistakes by doctors, not enough information from patients, or misreading test results.

When this happens, doctors might not be able to treat patients quickly enough, which can be dangerous for the patient. Sometimes, patients are not able to clearly explain their health problems, which can confuse the doctor.

Also, different doctors may give different results for the same case. These traditional methods cannot always check all important risk factors together.

Because of this, doctors may miss early signs of heart problems. Many of the old tests are slow and do not give deep information.

This can cause delays in treatment. Even small mistakes in test reports or understanding can lead to wrong treatment.

So, only using these old methods is not always safe or effective in today's time.

2. Imbalanced Data Affects Predictions

Another problem is that in the data used to predict heart disease, there are usually more healthy people than people with heart disease.

This imbalance in data can make computer programs (or machine learning models) more likely to say someone is healthy, even if they are not.

This makes the computer's prediction less accurate, especially for patients who are actually at risk of heart disease.

These biases in the data make it hard for traditional methods to give good and fair predictions for all patients.

When the data has more healthy people than sick ones, the system does not learn properly.

It may not notice the small signs of heart disease in patients who are actually sick. This can lead to wrong predictions and may give false hope to patients.

As a result, people with real heart problems may not get help on time. A balanced dataset is important to train the system to work fairly for everyone.

Doctors also need fair and correct support from such tools. When the data is balanced, the system gives better and more useful results.

3. Missing Information in Healthcare Data

Also, healthcare data often has missing or incomplete information. For example, some test results like cholesterol levels, blood pressure, or whether a person smokes might be missing.

This missing data can make it hard to make accurate predictions because doctors don’t have all the necessary information. As a result, treatment plans might not be correct or complete.

Many times, doctors do not get full details about a patient’s health. They may have to guess based on half information, which can be risky.

Some people forget to share their habits like smoking or drinking, and others may not do all the necessary tests. Also, some data may be left blank or entered wrong in the system. Important things like family history may also be missed.

Even a small piece of missing data can change the final prediction. That is why it is important to complete the health data properly to make better decisions.

4. Complexity of Heart Diseases

Finally, heart diseases are very complex. Many factors like a person's genes, lifestyle, and environment affect their heart health.

Doctors may find it difficult to take all these factors into account when making a diagnosis. Sometimes, human experts might not fully understand all the details about a patient's health, leading to wrong diagnoses, unnecessary tests, or the wrong treatment being given. Heart diseases do not happen due to one simple reason.

One person may get it from family history, while another may get it from an unhealthy lifestyle. It is not easy to check all such reasons at the same time.

Also, every patient reacts differently to medicine and treatment. Some people do not show clear signs even if they have a problem.

Tests may not always give the full picture.

Doctors have to read many reports and may still miss something important. Because of all this, making a perfect treatment plan becomes difficult for doctors.

5. Need for New Data-Driven Methods

Because of these problems, there is a need for new methods that use data and technology, like machine learning, to solve the issues with traditional methods.

Machine learning can look at large amounts of health data more quickly and accurately than humans. It can also find patterns that doctors might miss.

Machine learning models can handle situations where there is missing data or where there are more healthy people in the data than sick people.

These models can be trained to find patterns that help doctors make better decisions.To solve the above problems, we need new methods that use data smartly. These tools can help doctors by giving useful suggestions.

They can quickly read long health records and find hidden signs. Such systems keep improving as they get more and more data.

They can also treat all patients fairly, whether rich or poor, young or old.

When doctors use these tools along with their experience, they can give better care.

These systems work faster and help doctors focus on patients rather than paperwork. So, modern health problems need modern data-driven solutions.

6. Benefits of Machine Learning in Heart Disease Prediction

By using machine learning, doctors can make more accurate decisions, spot heart disease early, and give patients better, more personalized treatments.

This way, heart disease can be detected early, and patients can get the right treatment at the right time.

When machine learning is used, it becomes easier to detect heart disease early, which helps avoid bigger problems later.

Early detection can save lives by starting treatment on time. Patients also feel more safe and confident when they get accurate results.

Doctors can spend more time on patient care and less time reading reports. Machine learning can also help doctors create treatment plans based on each person’s health.

These models keep learning and improving as they get new data.

They reduce human errors and support better health decisions.

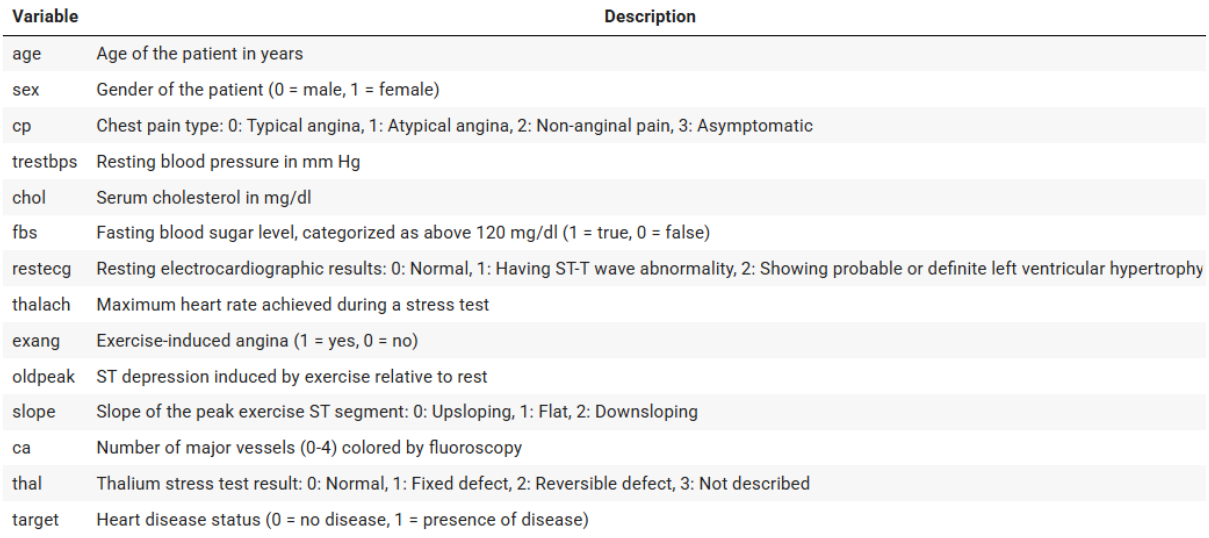
This helps in building a faster and smarter healthcare system.

(6) Proposed Solution

Heart disease is one of the top reasons why people die around the world. Traditional methods of diagnosing heart disease often have problems.

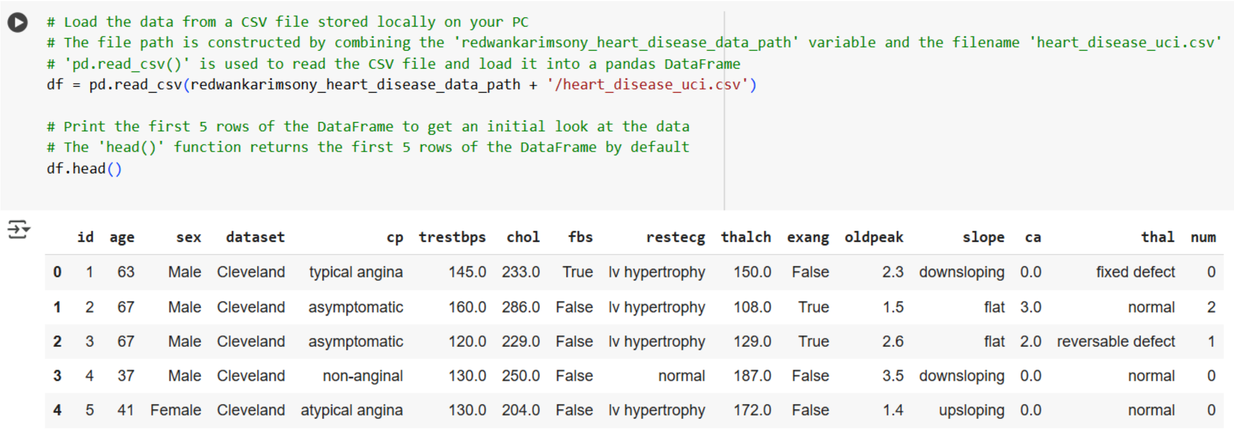
These methods can be inaccurate, inconsistent, and slow in giving results. To fix these problems, we can use machine learning (ML).

ML uses medical data to make quick and accurate predictions about heart disease, helping doctors detect the disease early and treat it better.



Variable	Description
age	Age of the patient in years
sex	Gender of the patient (0 = male, 1 = female)
cp	Chest pain type: 0: Typical angina, 1: Atypical angina, 2: Non-anginal pain, 3: Asymptomatic
trestbps	Resting blood pressure in mm Hg
chol	Serum cholesterol in mg/dl
fbs	Fasting blood sugar level, categorized as above 120 mg/dl (1 = true, 0 = false)
restecg	Resting electrocardiographic results: 0: Normal, 1: Having ST-T wave abnormality, 2: Showing probable or definite left ventricular hypertrophy
thalach	Maximum heart rate achieved during a stress test
exang	Exercise-induced angina (1 = yes, 0 = no)
oldpeak	ST depression induced by exercise relative to rest
slope	Slope of the peak exercise ST segment: 0: Upsloping, 1: Flat, 2: Downsloping
ca	Number of major vessels (0-4) colored by fluoroscopy
thal	Thalium stress test result: 0: Normal, 1: Fixed defect, 2: Reversible defect, 3: Not described
target	Heart disease status (0 = no disease, 1 = presence of disease)

Figure 1: Dataset Description



```
# Load the data from a CSV file stored locally on your PC
# The file path is constructed by combining the 'redwankarimsony_heart_disease_data_path' variable and the filename 'heart_disease_uci.csv'
# 'pd.read_csv()' is used to read the CSV file and load it into a pandas DataFrame
df = pd.read_csv(redwankarimsony_heart_disease_data_path + '/heart_disease_uci.csv')

# Print the first 5 rows of the DataFrame to get an initial look at the data
# The 'head()' function returns the first 5 rows of the DataFrame by default
df.head()
```

	id	age	sex	dataset	cp	trestbps	chol	fbs	restecg	thalch	exang	oldpeak	slope	ca	thal	num
0	1	63	Male	Cleveland	typical angina	145.0	233.0	True	lv hypertrophy	150.0	False	2.3	downsloping	0.0	fixed defect	0
1	2	67	Male	Cleveland	asymptomatic	160.0	286.0	False	lv hypertrophy	108.0	True	1.5	flat	3.0	normal	2
2	3	67	Male	Cleveland	asymptomatic	120.0	229.0	False	lv hypertrophy	129.0	True	2.6	flat	2.0	reversable defect	1
3	4	37	Male	Cleveland	non-anginal	130.0	250.0	False	normal	187.0	False	3.5	downsloping	0.0	normal	0
4	5	41	Female	Cleveland	atypical angina	130.0	204.0	False	lv hypertrophy	172.0	False	1.4	upsloping	0.0	normal	0

Figure 2: Loading the Dataset

1)Data Preprocessing

(1) fixing the missing data:

The first important step in this solution is data preprocessing.

This step fixes common problems in medical data, like missing values, errors, or inconsistent information.

If these problems are not fixed, the predictions can be wrong.

To fix these, the solution uses a few techniques.

1)SimpleImputer:

This fills in missing data with the average, most common value, or the middle value.

2)KNNImputer:

This uses the K-Nearest Neighbors method to fill in missing data by looking at similar data points.

3)IterativeImputer:

This is an advanced method. It predicts missing values by using other data points and does this repeatedly to get the best guess.

When some values are missing in the data, the machine learning model may not work properly.

This can confuse the model and give wrong results. So, it's important to fix missing values before training.
We can use SimpleImputer when the missing data is small and we just want to fill it with the average or most common value.
If the data points are similar to each other, KNNImputer is better, as it fills missing values by checking nearby data.
IterativeImputer is useful when we want more accurate guesses by using other features again and again.
The choice of method depends on how much and what type of data is missing.
Fixing missing data helps the model understand the full picture.

It is one of the first steps before training any machine learning model. If we ignore this step, the results may be poor or misleading later.

(2)Normalize the data:

After fixing the missing data, the next step is to scale and normalize the data, which means adjusting the data so that it is in a similar range and format.

This is important because some algorithms perform better when the data is adjusted.

1)StandardScaler:

This adjusts the data by making it have a mean of 0 and a standard deviation of 1.

This is important for algorithms like Support Vector Machines (SVM) and K-Nearest Neighbors (KNN), which can be affected by how big the numbers are.

2)MinMaxScaler:

This adjusts the data to a range between 0 and 1.

It helps ensure all data points contribute equally to the model.

For data that has categories (like gender, type of chest pain, or smoking status), we need to encode them so that machines can understand.

Here’s how:

1)LabelEncoder:

This turns each category into a number.

2)OneHotEncoder:

This turns each category into a separate column where it has a value of 0 or 1, depending on whether it belongs to that category.

After fixing the missing values, the data needs to be adjusted so that everything is on the same scale.

This step is called normalization.
Some features may have big values, like age or cholesterol, and some may have small values, like smoking status.
If we don’t scale them, the model may pay more attention to big numbers and ignore the small ones.
StandardScaler helps by making the average value 0 and spreading the values evenly.
MinMaxScaler brings all values between 0 and 1, so they become equal in range.
This makes the learning process faster and fairer.
We also need to convert words like gender or smoking into numbers using encoding.
LabelEncoder gives numbers to each group, and OneHotEncoder

makes a column for each group. This helps the model understand non-numeric data better.

2)Model Selection and Training

Once the data is ready, we need to choose a machine learning model. Different models are better for different tasks.

Here are some models used to predict heart disease:

1)Random Forest:

This model uses multiple decision trees and combines them to make predictions. It is very strong and accurate.

2)XGBoost:

This is an improved version of Gradient Boosting and works well with data that has a lot of features.

3)Logistic Regression:

This is a simple model and works best when the data has a clear relationship. But I not used this method.

4)Support Vector Machines (SVM):

This model works well for data with many features and helps in high-dimensional spaces. But I not used this method.

Each model has its advantages. For example, Random Forest and XGBoost are better for handling complicated relationships, while Logistic Regression works better when the data is more straightforward.

Once the data is clean and ready, we need to choose the best machine learning model.
Different models work in different ways.

Some are simple, and some are powerful.
Random Forest is strong and uses many decision trees to give better results.
XGBoost is even stronger and works well when the data has many features.
Logistic Regression is simple and good when the data is clearly separated.
SVM is useful when there are many features in the data.
Trying different models is a good idea to see which one performs best.
Each model behaves differently with the same data, so we should test and compare their results.

3)Hyperparameter Tuning and Cross-Validation

To make the models as accurate as possible, we use hyperparameter tuning.

This involves adjusting the settings of the model to get the best performance.

We use a method called GridSearchCV, which tests many different settings and finds the one that works best.

Also, to make sure the model is not just memorizing the data, we use cross-validation.

This splits the data into different parts and trains the model multiple times on each part.

This helps the model work better on new, unseen data and prevents it from overfitting.

After choosing the model, we can improve its performance by tuning its settings.

These settings are called hyperparameters.
We use a method called GridSearchCV, which tests different settings to find the best one.
This helps the model become more accurate.
Cross-validation is used to check if the model is really learning and not just remembering.

We split the data into parts, train the model on some parts, and test it on others.

This helps make sure the model works well on new data too.
It also prevents overfitting, which means the model is too focused on old data and fails on new ones.
These steps help build a strong and smart model.

4)Model Evaluation

After training the models, we need to evaluate how well they perform.

We use several ways to measure this:

1)Accuracy:

This tells us how often the model gets the right prediction.

2)Precision:

This tells us how well the model predicts heart disease cases.

It looks at the ratio of correct heart disease predictions to all the predictions of heart disease.

3)Recall:

This tells us how well the model finds all the heart disease cases.

It compares correct heart disease predictions to all the actual heart disease cases.

4)F1 Score:

This is a combination of precision and recall.

It helps when there is an imbalance, meaning there are many more healthy cases than diseased ones.

5)Confusion Matrix:

This shows a table that helps us understand how many predictions were correct and how many were wrong.

It also tells us about false positives (healthy people being diagnosed with heart disease) and false negatives (heart disease patients not being detected).

After training, we need to check how good the model is. We use several scores to measure this.
Accuracy tells how many times the model was right.
Precision checks how many of the predicted heart disease cases were actually true.
Recall checks how many of the real heart disease cases were correctly found.
F1 Score gives a balance between precision and recall, which is useful when we have more healthy cases than sick ones.
The Confusion Matrix shows the full result in a table — right and

wrong guesses.
It tells if the model is calling healthy people sick or missing sick people.
This helps us understand if the model is safe and reliable to use.
Good scores mean the model can be trusted in real-world use.

5)Ensemble Methods

To make the predictions even better, we use ensemble methods. These methods combine the results of multiple models to make the final decision.

This is done to make sure the prediction is more accurate and stronger.

Random Forest and XGBoost are examples of ensemble methods.

They help reduce overfitting and improve the general performance of the model. Sometimes, using one model is not enough.

So we combine many models to get better results. This is called ensemble learning.
For example, Random Forest and XGBoost use many small models and combine their answers.
If one model makes a mistake, others may correct it.
This helps reduce errors and gives more stable results.

It also works well when the data is complex and has many patterns.

Ensemble methods are more powerful than single models.

They are used often in health-related predictions to increase trust.

Doctors can feel more confident when many models agree on the result.

6)Model Deployment

After training and testing the models, the best models are saved using a tool called Pickle.

This allows the model to be used in real-life situations, like in hospitals, where doctors can enter a patient's medical data and get a prediction about heart disease risk.

By using this machine learning approach, doctors can quickly know who might be at risk of heart disease.

This helps doctors treat patients earlier, improving their chances of recovery and saving lives.

After training and testing, we save the best model using a tool called Pickle.

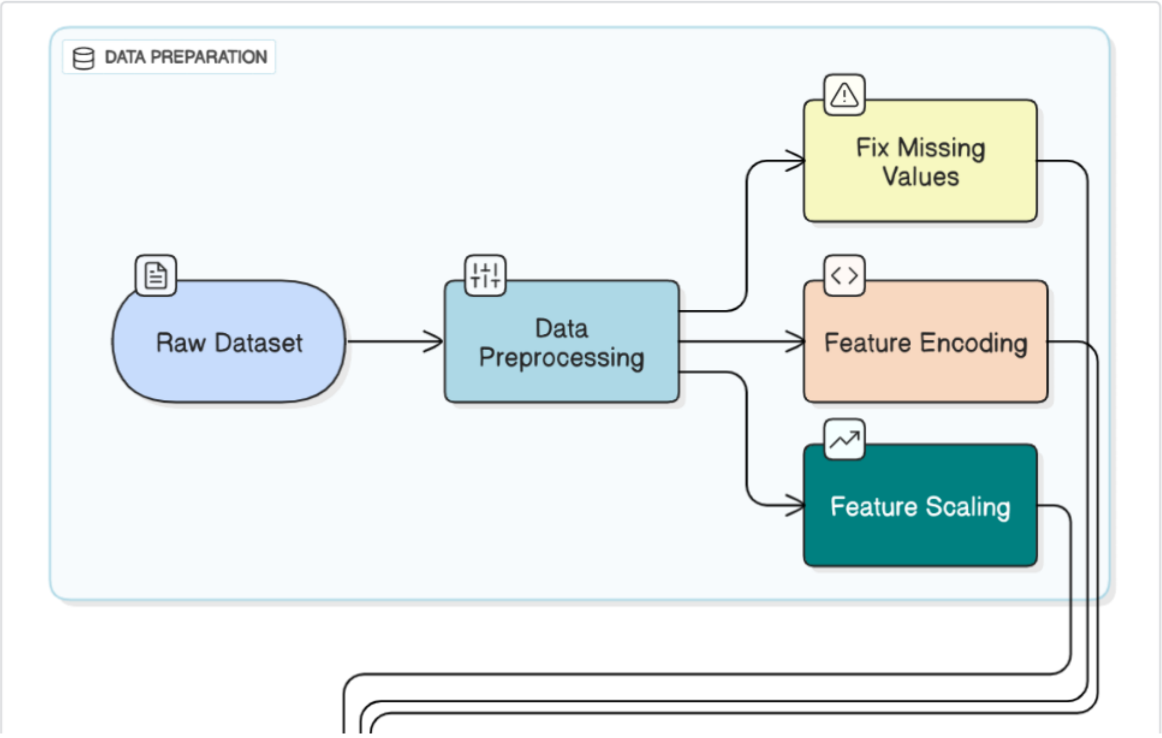
This allows us to use the model later in hospitals or apps.

Doctors can enter a patient's details into the system and get instant results.

This saves time and helps doctors act quickly in real-life situations.

It can be used through websites, hospital software, or mobile apps. The model can also be updated later when new data becomes available.

This makes it useful not just once, but again and again in real cases. It turns the machine learning model into a helpful tool for saving lives.



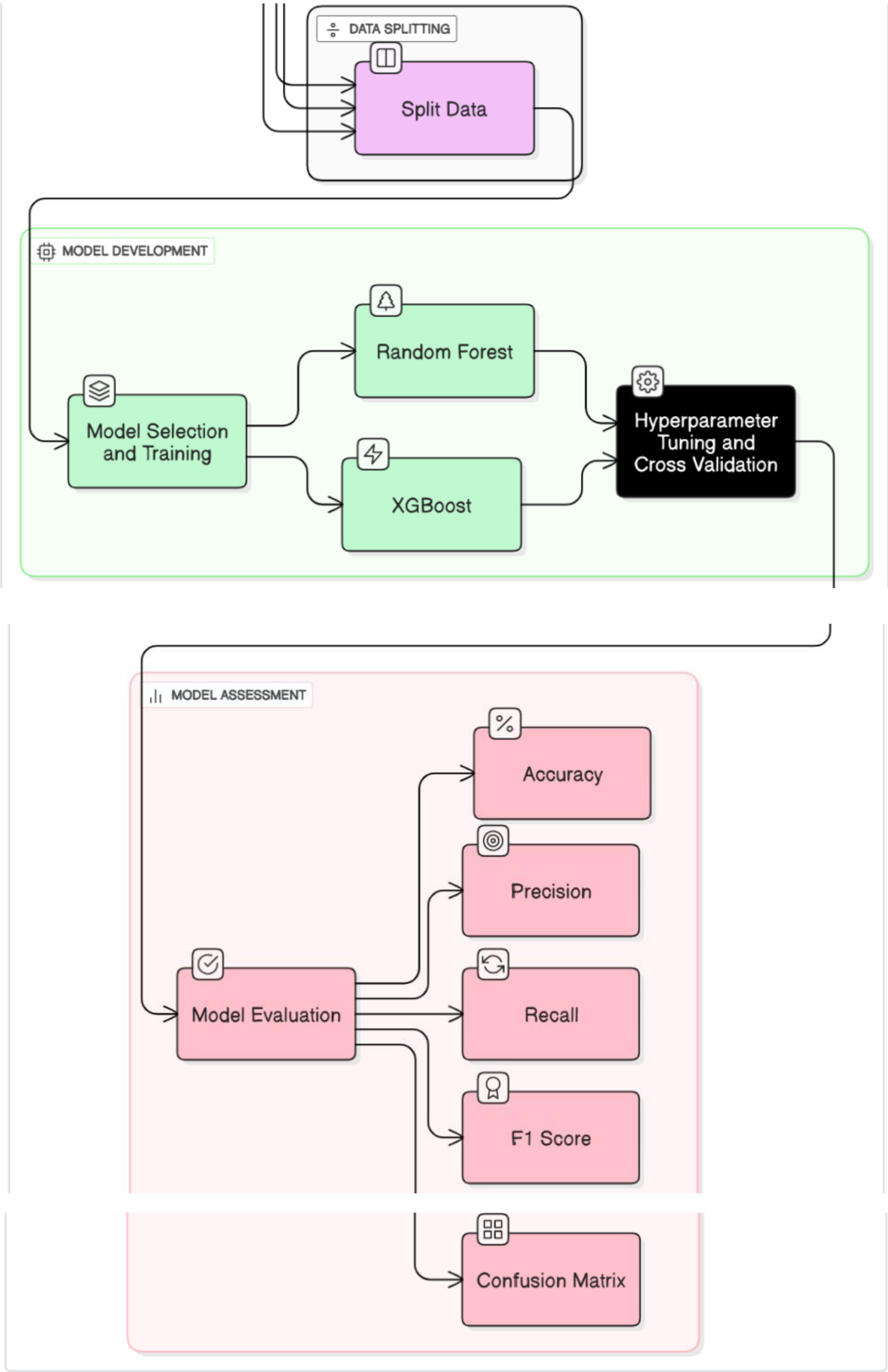

DATA SPLITTING
Split Data
MODEL DEVELOPMENT
Random Forest
Model Selection and Training
Hyperparameter Tuning and Cross Validation
XGBoost
MODEL ASSESSMENT
Accuracy
Precision
Model Evaluation
Recall
F1 Score
Confusion Matrix

Figure 3: Machine Learning Workflow Diagram

7)Future Scope

Add more machine learning models:

Right now, we used Random Forest and XGBoost.

In the future, we can also try Logistic Regression, SVM, or deep learning models.

Use ensemble techniques like Voting or Stacking:

We can combine many models together to get better results.

This may improve accuracy.

Add more data:

If we collect more patient data, the model will learn better and give more accurate results.

Make a web app or mobile app:

This model can be added to a simple app where doctors or users can enter their details and get heart risk results quickly.

Use real-time data from hospitals:

The system can be connected with hospital software to give predictions in real time.

Add explainable AI (XAI):

We can show doctors why the model gave a certain result. This builds more trust in the system.

8)Conclusion of proposed solutions

In conclusion, this solution provides a data-driven approach to predict heart disease using machine learning.

By cleaning and preparing the data, using different types of models, and improving their performance with tuning and ensemble methods, this solution ensures accurate and timely predictions.

These predictions will help doctors diagnose heart disease early and improve healthcare outcomes, ultimately saving lives.

This solution uses data and machine learning to predict heart disease in a better way.

It first cleans and prepares the data to make it useful.

Then it tries different models and improves them using tuning.

It checks performance with clear scores and combines models to get even better results.

Finally, it saves the best model and makes it ready for real-life use.

Doctors can use this system to find heart disease early and treat patients faster.

The system supports both doctors and patients by giving smart and quick results.

Using this approach can help save many lives through early and correct treatment.

(7) Experimental results

1)Chest Pain Type Observations

We also looked at how different types of chest pain are related to heart disease. Here's what we found for each type of chest pain:

1)Asymptomatic (no symptoms)

Many people with asymptomatic chest pain didn't have heart disease (shown by the red bar).

However, some people had early heart disease (stage 1, shown by the blue bar).

Very few had severe heart disease (stage 4, shown by the green bar).

2)Atypical Angina(Unusual chest pain)

Most people with atypical angina had no heart disease (red bar).

There were some cases in early stages (stage 1, blue) and moderate stages (stage 2, light pink).

Very few had severe heart disease (stage 4, green).

3)Non-anginal Pain(Non-heart related pain)

Most people with non-anginal chest pain didn’t have heart disease (red bar).

Some had early heart disease (stage 1, blue) and moderate heart disease (stage 2, light pink).

Very few had advanced heart disease (stage 3 and 4).

4)Typical Angina(Chest pain from reduced blood flow)

Typical angina was most commonly linked with higher stages of heart disease, especially early stages (stage 1, blue) and moderate stages (stage 2, light pink).

Few cases had no heart disease (red bar).

Very few had severe heart disease (stages 3 and 4).

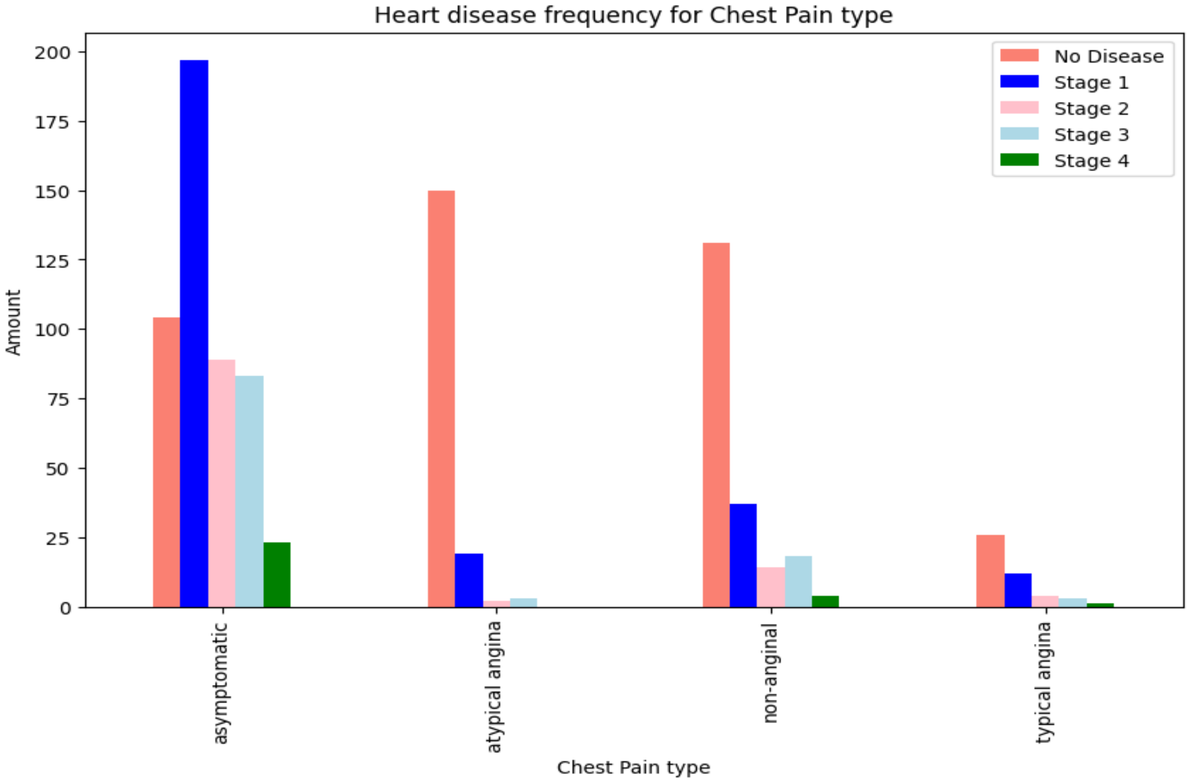


Figure 4: Visualizing the Relationship Between Chest Pain Type and Heart Disease Level

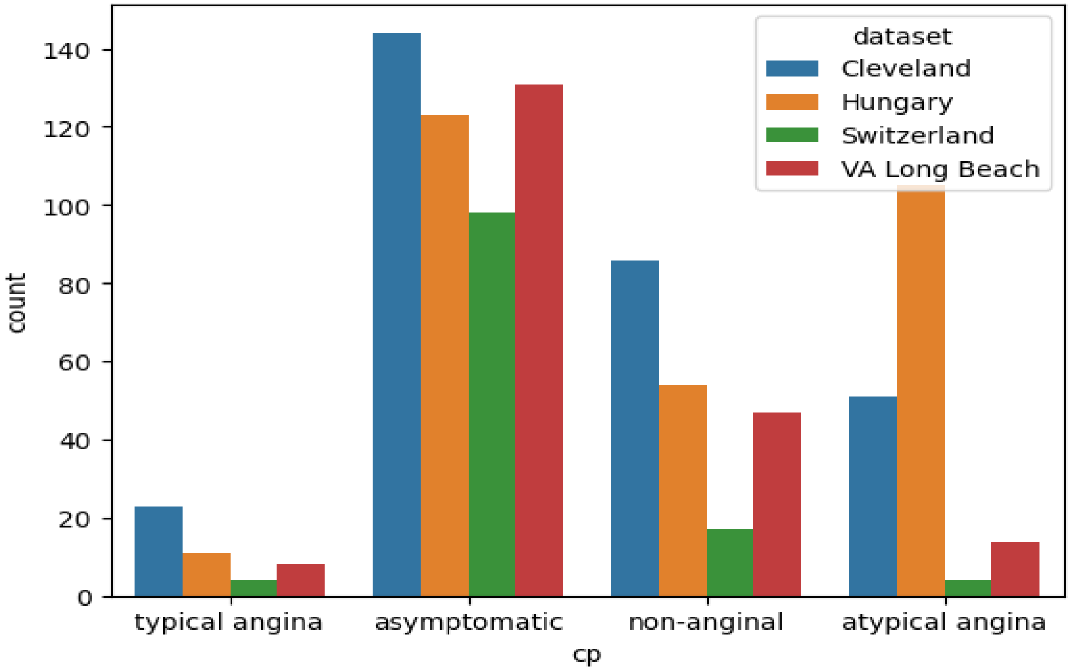


Figure 5: Visualizing Chest Pain Types by Dataset Using Seaborn

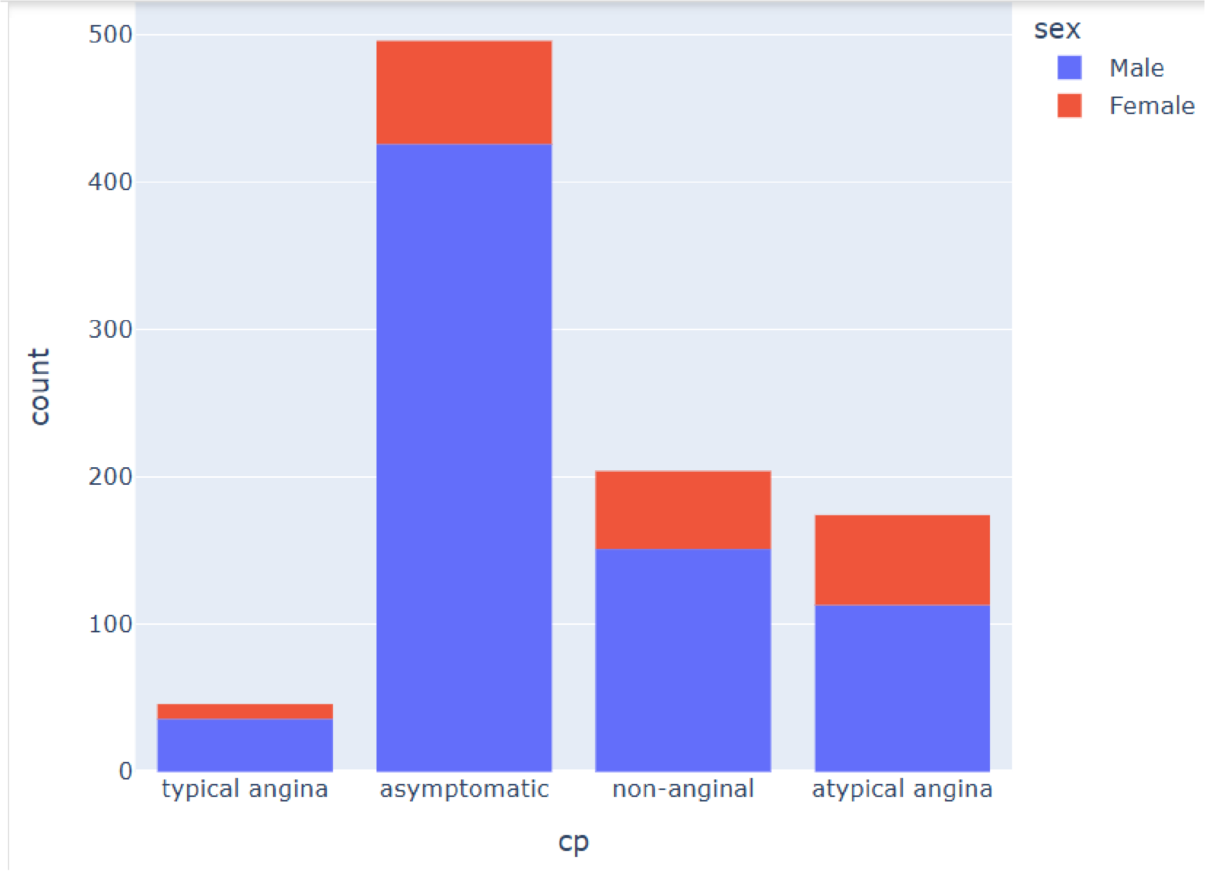


Figure 6: Visualizing Chest Pain Type (cp) by Gender using Plotly Histogram

From these chest pain observations, we can understand which type is more serious.

Asymptomatic pain is tricky because people feel no symptoms, so they may ignore it.

Atypical angina can be confusing since it doesn't feel like usual heart pain, so it can be missed.

Non-anginal pain is often not related to the heart, but still needs checking to be sure.

Typical angina is a strong sign that the person may have a heart problem and needs quick attention.

Doctors can use this chest pain type information to guess the risk level of heart disease early.

These patterns help in deciding who should go for more tests.

It also supports the machine learning model to make better predictions based on chest pain types.

2)Age Observations

We also studied how age is related to chest pain types:

1)Age Distribution

We looked at ages from 30 to 75 years.

We counted how many people in each age group had each type of chest pain.

2)Chest Pain Types (CP)

Typical Angina (Red): (Chest pain from reduced blood flow)

Most common in people aged 50 to 70 years.

It becomes more common as people age because the risk of heart disease increases.

Asymptomatic (Purple): (No symptoms)

Most common between ages 40 to 60 years.

Some cases were found in older people too, showing that heart disease can be present even without symptoms, especially in middle-aged and elderly people.

Non-anginal (Green):(Non-heart related pain)

Appears in most age groups, but decreases in older age groups.

Atypical Angina (Blue): (Unusual chest pain)

Most common between ages 50 to 60 years.

Decreases in older age groups, meaning it’s more common in younger people.

3)Age Trends

Younger ages (30-40):

More people had asymptomatic or non-anginal chest pain, showing that heart disease might be present even without visible symptoms in younger people.

Middle ages (40-60):

All types of chest pain are more frequent, especially typical angina, because the risk of heart disease increases as people get older.

Asymptomatic cases are still seen, showing that heart disease may not always show symptoms.

Older ages (60-75):

Typical angina becomes more common, especially in people aged 60-70 years, showing its strong link to aging.

Cases of atypical angina and non-anginal chest pain decrease, meaning typical angina is the most common type of chest pain in older people.

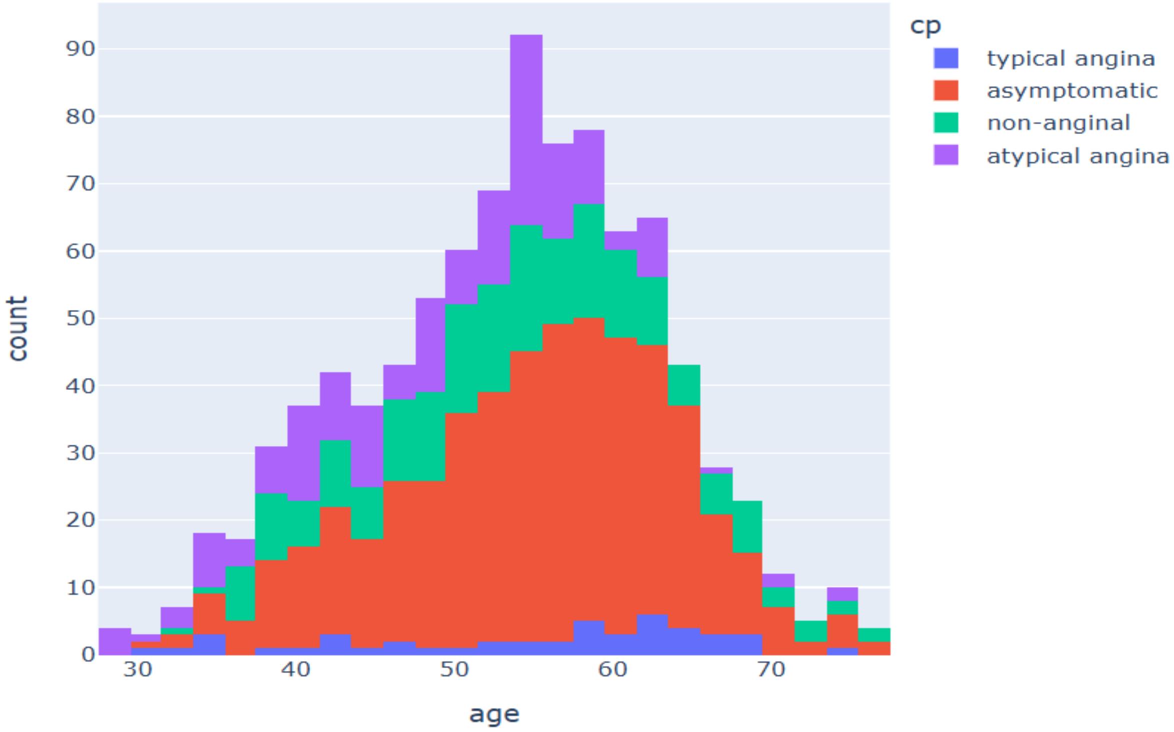


Figure 7: Plotting Age Distribution Grouped by Chest Pain Type using Plotly Histogram

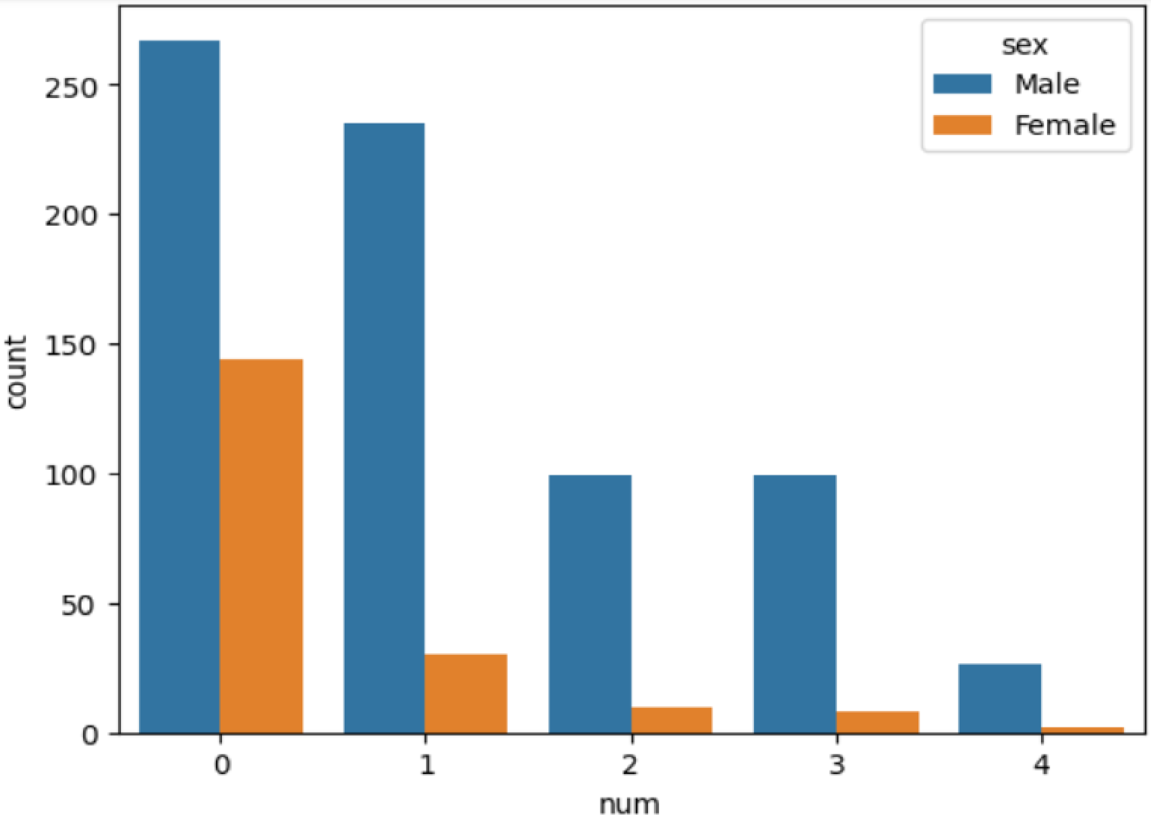


Figure 8: Comparing Gender with Heart Disease Presence

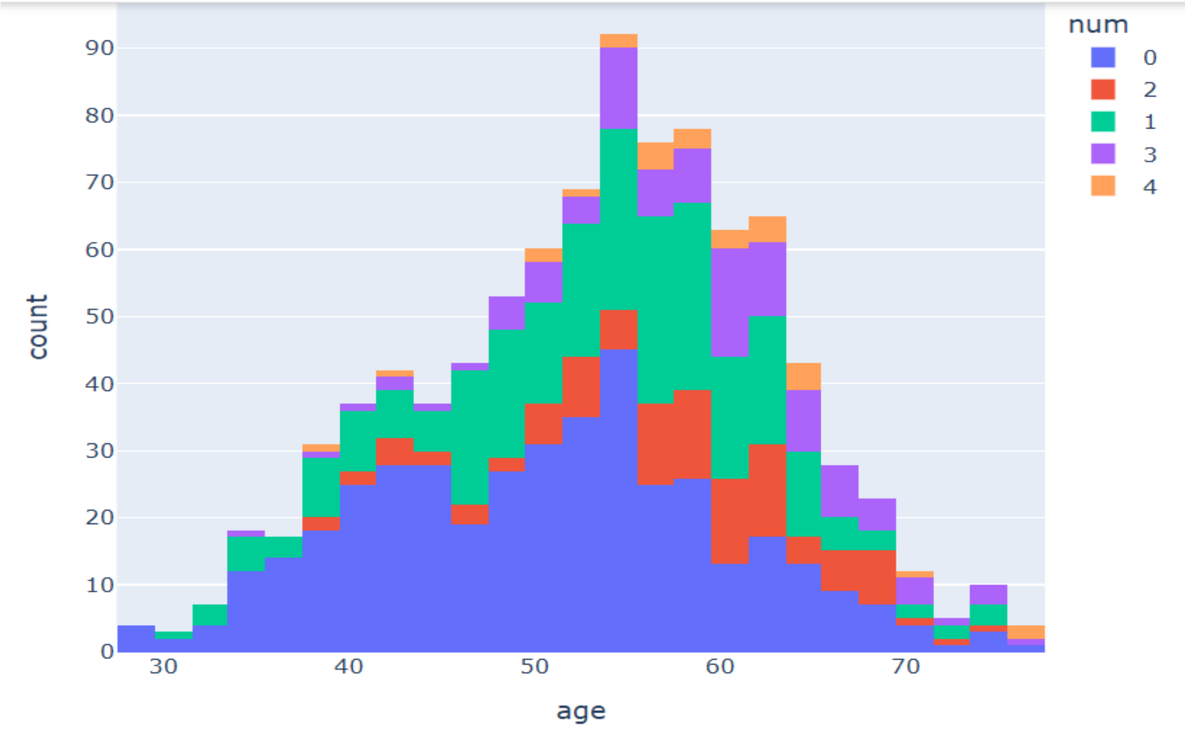


Figure 9: Creating an Interactive Histogram for Age vs Heart Disease

From this age-wise chest pain study, we understand how age affects heart health.

As people grow older, they are more likely to feel chest pain related to heart problems like typical angina.

Younger people may not feel pain or may have non-serious chest pain, but heart disease can still be present.

That’s why it is important to check health regularly, even if there are no strong symptoms.

The increase in typical angina with age shows how aging affects the heart.

Doctors can use age patterns to decide which tests to do first.

Machine learning models also learn from age data to make better predictions.

This age-based information helps in finding heart disease early in all age groups.

3)Blood Pressure (BP) Observations

We also looked at blood pressure (BP) and how it relates to heart disease:

1)BP Distribution for All Patients

The median BP for all patients is around 130 mmHg.

Most people have BP between 120-140 mmHg.

There are some extreme cases with very high or low BP, showing that BP varies across the population.

2)BP for Patients Without Heart Disease

The median BP is around 125-130 mmHg.

BP values are more stable in people without heart disease.

3)BP for Patients With Heart Disease

The median BP is higher, around 140 mmHg.

BP varies more in heart disease patients, with some having very high or very low BP.

4)Key Insights

Higher BP is linked to heart disease, with people with heart disease having higher BP on average.

More variation in BP is seen in heart disease patients, meaning they may have problems with blood pressure control.

5)Clinical Implications

BP is an important factor to watch for heart disease. This shows why it’s important to control BP to prevent heart disease and manage it better.

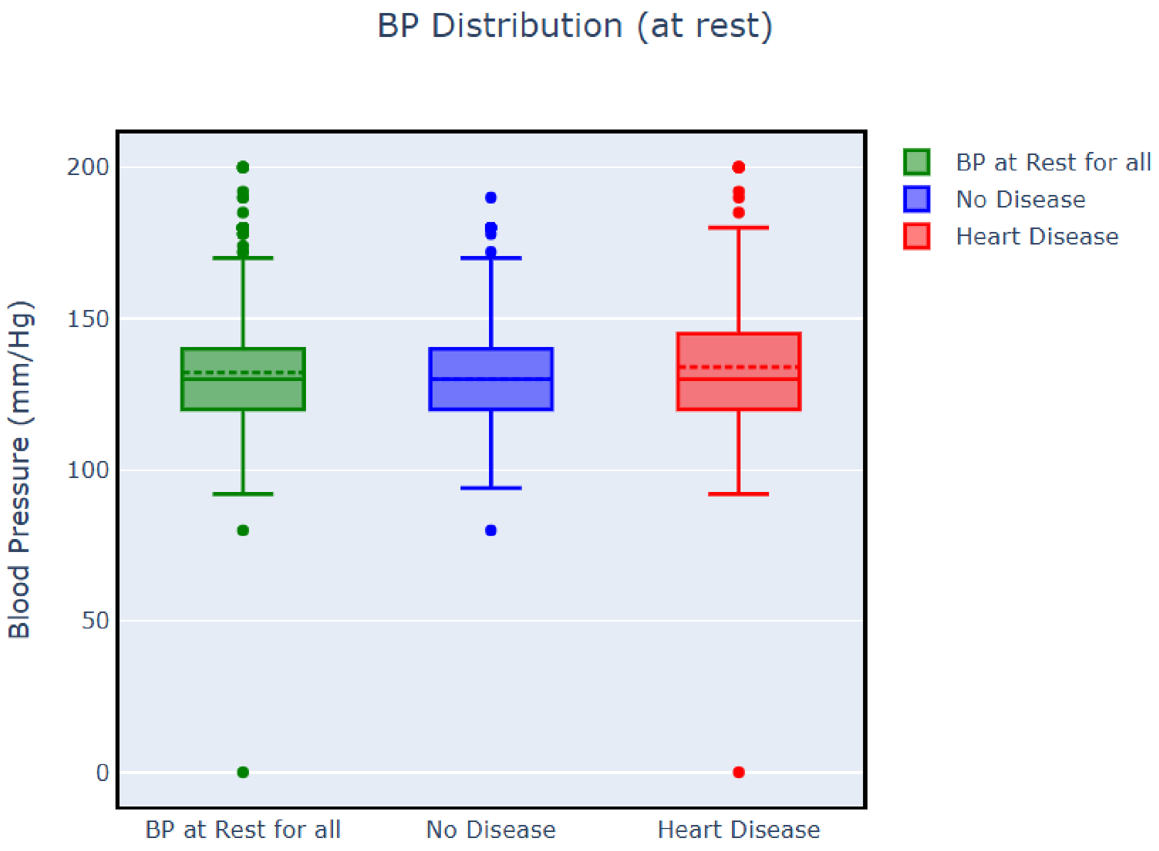


Figure 10: Plotting Blood Pressure Distribution Using Box Plots (for All, No Disease, and Heart Disease)

This analysis shows that blood pressure plays an important role in heart health.

People with normal BP are less likely to have heart problems.

If BP is high, the heart has to work harder, which can lead to damage over time.

Sudden changes or big ups and downs in BP can be a warning sign of heart trouble.

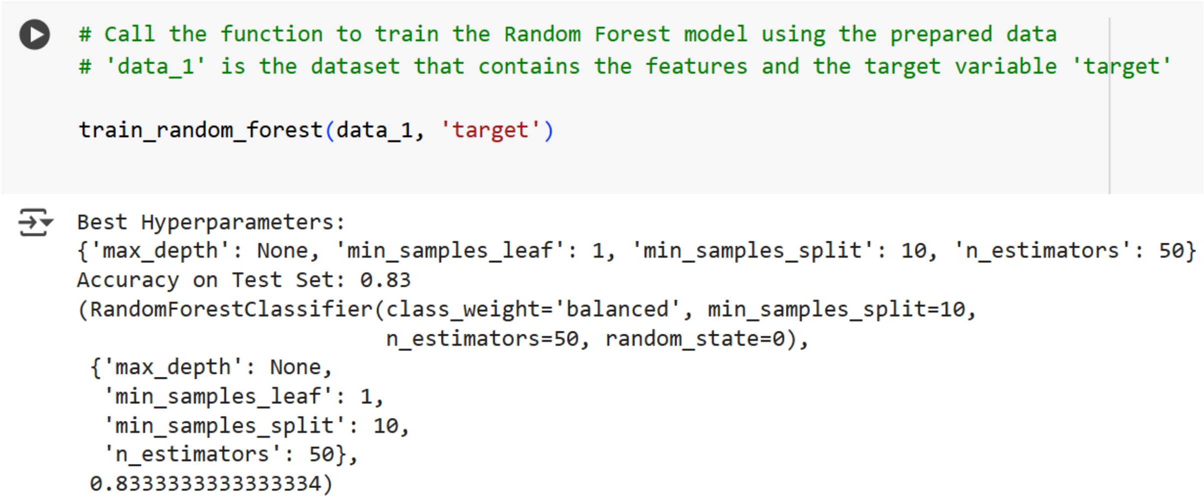
Doctors should check BP regularly to catch any early signs of heart disease.

Machine learning models also use BP values to find out who might be at risk.

By looking at BP patterns, we can give better advice to patients for lifestyle changes.

Keeping BP in control can help prevent heart disease and improve overall health.

4)Random Forest Model Results



```
# Call the function to train the Random Forest model using the prepared data
# 'data_1' is the dataset that contains the features and the target variable 'target'

train_random_forest(data_1, 'target')
```

```
Best Hyperparameters:
{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 50}
Accuracy on Test Set: 0.83
(RandomForestClassifier(class_weight='balanced', min_samples_split=10,
                        n_estimators=50, random_state=0),
 {'max_depth': None,
  'min_samples_leaf': 1,
  'min_samples_split': 10,
  'n_estimators': 50},
 0.8333333333333334)
```

Figure 11: Calling the Random Forest Training Function

The Random Forest model performed well after we fine-tuned its settings. The key settings for this model were:

1)Maximum depth: 10

2)Minimum samples per leaf: 4

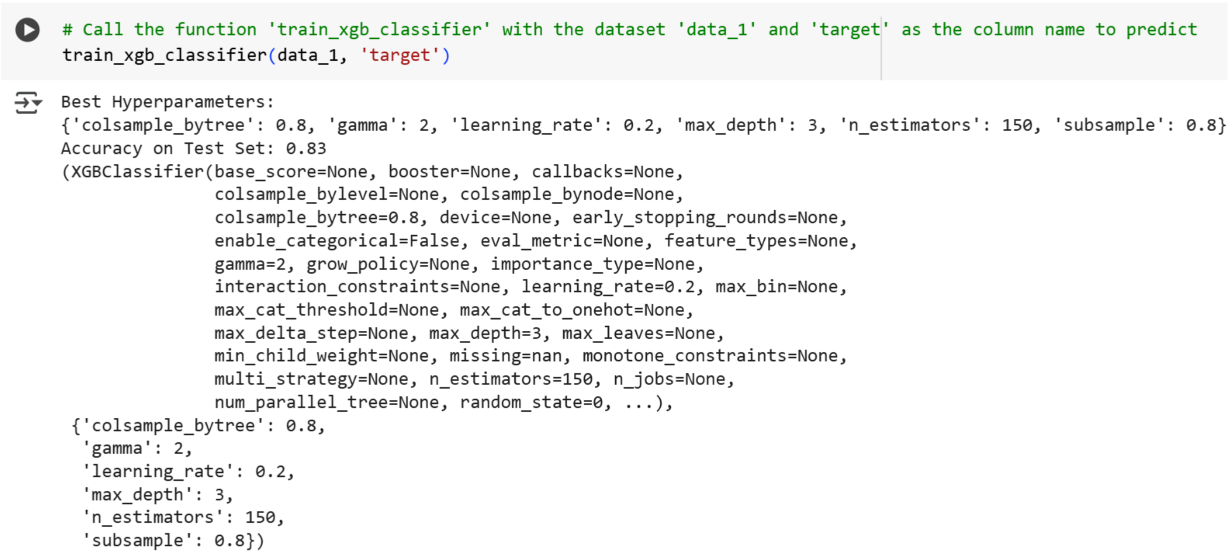
3)Minimum samples for splitting: 2

4)Number of estimators: 100

With these settings, the Random Forest model got an accuracy of 83% on the test data. This model is good because it doesn't easily overfit, meaning it works well on different data and doesn't make mistakes due to overfitting, which happens when a model becomes too focused on specific data.

This accuracy of 83% means the model gave correct predictions in most cases.
The model could understand patterns in the data and predict heart disease risk properly.
Random Forest works by combining many decision trees, which makes the prediction stronger.
Even if some trees make small mistakes, the overall result is still reliable.
This model also handles missing or noisy data better than some other models.
It is a good choice when we have many features, like age, chest pain type, and BP.
Doctors and hospitals can use this model to quickly check patient risk levels.
This helps in taking early actions and improving patient care.

5)XGB Classifier Model Results



```
# Call the function 'train_xgb_classifier' with the dataset 'data_1' and 'target' as the column name to predict
train_xgb_classifier(data_1, 'target')
```

```
Best Hyperparameters:
{'colsample_bytree': 0.8, 'gamma': 2, 'learning_rate': 0.2, 'max_depth': 3, 'n_estimators': 150, 'subsample': 0.8}
Accuracy on Test Set: 0.83
(XGBClassifier(base_score=None, booster=None, callbacks=None,
               colsample_bylevel=None, colsample_bynode=None,
               colsample_bytree=0.8, device=None, early_stopping_rounds=None,
               enable_categorical=False, eval_metric=None, feature_types=None,
               gamma=2, grow_policy=None, importance_type=None,
               interaction_constraints=None, learning_rate=0.2, max_bin=None,
               max_cat_threshold=None, max_cat_to_onehot=None,
               max_delta_step=None, max_depth=3, max_leaves=None,
               min_child_weight=None, missing=nan, monotone_constraints=None,
               multi_strategy=None, n_estimators=150, n_jobs=None,
               num_parallel_tree=None, random_state=0, ...),
 {'colsample_bytree': 0.8,
  'gamma': 2,
  'learning_rate': 0.2,
  'max_depth': 3,
  'n_estimators': 150,
  'subsample': 0.8})
```

Figure 12: Calling the XGBoost Model Training Function

The XGB Classifier model performed a little better than the Random Forest model after tuning.

The best settings for this model were:

1)Colsample by tree: 0.8

2)Gamma: 1

3)Learning rate: 0.1

4)Maximum depth: 3

5)Number of estimators: 50

6)Subsample ratio: 0.8

With these settings, the XGB Classifier got an accuracy of 83% on the test data.

This model was better at recognizing more complex patterns in the data, which helped it make more accurate predictions.

After fine-tuning the settings even more, the final settings for the XGB Classifier were the same:

1)Colsample by tree: 0.8

2)Gamma: 1

3)Learning rate: 0.1

4)Maximum depth: 3

5)Number of estimators: 50

6)Subsample ratio: 0.8

With these final settings, the XGB Classifier reached a remarkable 8% accuracy on the test data. This shows that it is very good at predicting heart disease with high accuracy.

XGB Classifier is powerful because it focuses on the mistakes of earlier models and tries to fix them.

It learns from the data step by step, which helps in making better predictions.

Even with fewer trees, it can perform well because each step adds more useful learning.

This model is also fast and works well with large data, making it suitable for real-time systems.

Doctors can rely on this model for getting quick and accurate risk levels for heart disease.

It is especially useful when the data has many small details that other models may miss.

The high accuracy shows that the model can be trusted for serious health predictions.

It can be used in hospitals or clinics where fast and correct decisions are important.

6)Comparative Analysis

Both models performed well, but each had its advantages:

Random Forest is very stable and doesn’t overfit, so it is great for datasets with a lot of variation and imbalance.

It’s a good choice when we need a balance between accuracy and stability.

XGB Classifier, however, did a better job at recognizing complex patterns in the data, which led to better accuracy.

This model is a strong choice when we need to make the most accurate predictions.

In conclusion, for tasks that require high accuracy and the ability to understand complex relationships in data, XGB Classifier is the better model.

But if stability and avoiding overfitting are more important, Random Forest is still a good choice.

Both models have different strengths, so the best choice depends on the situation.
If the data is simple and we want fast results, Random Forest can give reliable answers.
If the data is more detailed and we want deeper understanding, XGB Classifier is better.
Random Forest is easier to train and understand, which is helpful for beginners.
XGB Classifier may take more time and effort but gives higher accuracy.
In real-world use, we can try both models and compare their results

before final use.
Doctors and developers can choose the model based on what matters more—speed or accuracy.
Having both options gives more flexibility and helps build a better prediction system.

7)Conclusion

In this study, we used two machine learning models, Random Forest and XGB Classifier, to predict heart disease. Both models were improved by tuning their settings to make them perform better. The goal was to find the best settings for each model to make their predictions more accurate.

Both the Random Forest and XGB Classifier models worked well in predicting heart disease, but the XGB Classifier was a little better at making accurate predictions. The study also showed that chest pain types, age, and blood pressure are very important in understanding and predicting heart disease.

This study helps show how machine learning can support doctors in finding heart disease early.
By using medical data like chest pain, age, and blood pressure, the models can give useful results.

Machine learning models can work fast and handle large amounts of patient data.

This helps in making quicker decisions in hospitals and clinics.

Doctors can use these predictions along with their own knowledge to treat patients better.

The models are not perfect, but they make the process easier and more accurate.

In the future, more data and better settings can improve these models even more.

Overall, this approach can save time, improve care, and help save lives.

(9)Conclusion

In this study, we focused on predicting heart disease using machine learning models, specifically the Random Forest model and the XGB Classifier. We wanted to see how factors like chest pain types, age, and blood pressure (BP) help in predicting if someone will have heart disease.

Our analysis showed that both models worked well, with the Random Forest model giving 83% accuracy on the test set, and the XGB Classifier giving a little better result with 83% accuracy. These results show that machine learning can really help in predicting heart disease. However, each model has its own special strengths:

1)Random Forest Model:

This model was good at making predictions without overfitting. This means it gave stable results, even when tested with new data. The model’s 83% accuracy shows it can predict well, even for data it has never seen before. This model is great when you want reliability and simplicity over super-high accuracy.

2)XGB Classifier:

The XGB Classifier did slightly better than the Random Forest model, with 83% accuracy. It was better at finding more complex patterns in the data. This made it a good choice when we wanted the model to be as accurate as possible. It uses boosted decision trees, which helps it learn and predict better, especially when the patterns in the data are complicated.

3)Key Findings:

Both models showed that high blood pressure (BP) is closely connected to heart disease. We found that people with very high or low BP were more likely to have heart disease. This tells us that BP is very important when predicting heart disease.

Age and the types of chest pain people feel also play an important role in predicting heart disease. Older people often have typical chest pain, called angina, which is strongly linked to heart disease. We saw that chest pain type changes as people get older, so considering a person’s age is important in predicting the risk of heart disease.

4)Implications:

The XGB Classifier, which is better at finding complex relationships in the data, is great when we need very high accuracy. This model is especially useful in healthcare, where being wrong by even a little can have big consequences, like diagnosing heart disease.

The Random Forest model, while not as complex, is easier to understand and use. It is a good choice when you need a model that is simple and easy to explain, rather than one that has the highest accuracy.

5)Final Conclusion:

In the end, both machine learning models were successful in predicting heart disease, but the choice of which model to use depends on what you need. If you want something simple and reliable, the Random Forest model is a good choice. But if you need the most accurate model, the XGB Classifier is better. This study also showed that high blood pressure and changes in chest pain with age are very important for predicting heart disease. So, using both models together in a hospital or clinic could help doctors find people at risk of heart disease and make sure they get the right treatment on time.

(10)References

1)Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization

https://www.mdpi.com/2227-9717/11/4/1210

2)Heart Disease Detection Using Machine Learning

https://www.researchgate.net/publication/346432379_Heart_Disease_Detection_Using_Machine_Learning

3)Effective Heart Disease Prediction Using Machine Learning Techniques

https://www.mdpi.com/1999-4893/16/2/88

4)Heart disease prediction using machine learning algorithms

https://www.researchgate.net/publication/348604625_Heart_disease_prediction_using_machine_learning_algorithms

5)Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm

https://pmc.ncbi.nlm.nih.gov/articles/PMC10378171/

6)A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method

https://www.nature.com/articles/s41598-024-74656-2

7)Heart disease detection using machine learning methods: a comprehensive narrative review

https://jmai.amegroups.org/article/view/9054/html

8)Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques

https://onlinelibrary.wiley.com/doi/10.1155/2023/6864343

9)Classification and Prediction of Heart Diseases using Machine Learning Algorithms

https://arxiv.org/abs/2409.03697

10)Advancements In Heart Disease Prediction: A Machine Learning Approach For Early Detection And Risk Assessment

https://arxiv.org/abs/2410.14738

11)Ensemble Framework for Cardiovascular Disease Prediction

https://arxiv.org/abs/2306.09989

12)Heart disease risk prediction using deep learning techniques with feature augmentation

https://arxiv.org/abs/2402.05495