Heart Disease Detection - Project Report

# 1. Introduction

Heart disease is one of the leading causes of mortality worldwide. Early detection is crucial for preventive healthcare. This project aims to develop a machine learning system that predicts whether a patient has heart disease based on medical and physiological data.

# 2. Problem Statement

The objective is to build a classification model that determines the presence of heart disease using patient data such as age, sex, chest pain type, blood pressure, cholesterol levels, and more.

# 3. Dataset Overview

* The dataset includes several attributes from patient records:  
  - Age, Sex  
  - Chest Pain Type  
  - Resting Blood Pressure, Cholesterol  
  - Fasting Blood Sugar, Resting ECG results  
  - Maximum Heart Rate, Exercise-induced Angina  
  - ST Depression (Oldpeak), Slope of ST segment  
  - Target (0 = No Heart Disease, 1 = Heart Disease)

# 4. Data Preprocessing

The dataset was cleaned by handling missing values and encoding categorical variables. Numerical features were scaled using standard normalization to ensure model stability. The target column was balanced and stratified while splitting the data.

# 5. Exploratory Data Analysis (EDA)

The following visualizations helped understand feature importance and distributions:

Figure 1: Chest Pain Type vs. Heart Disease

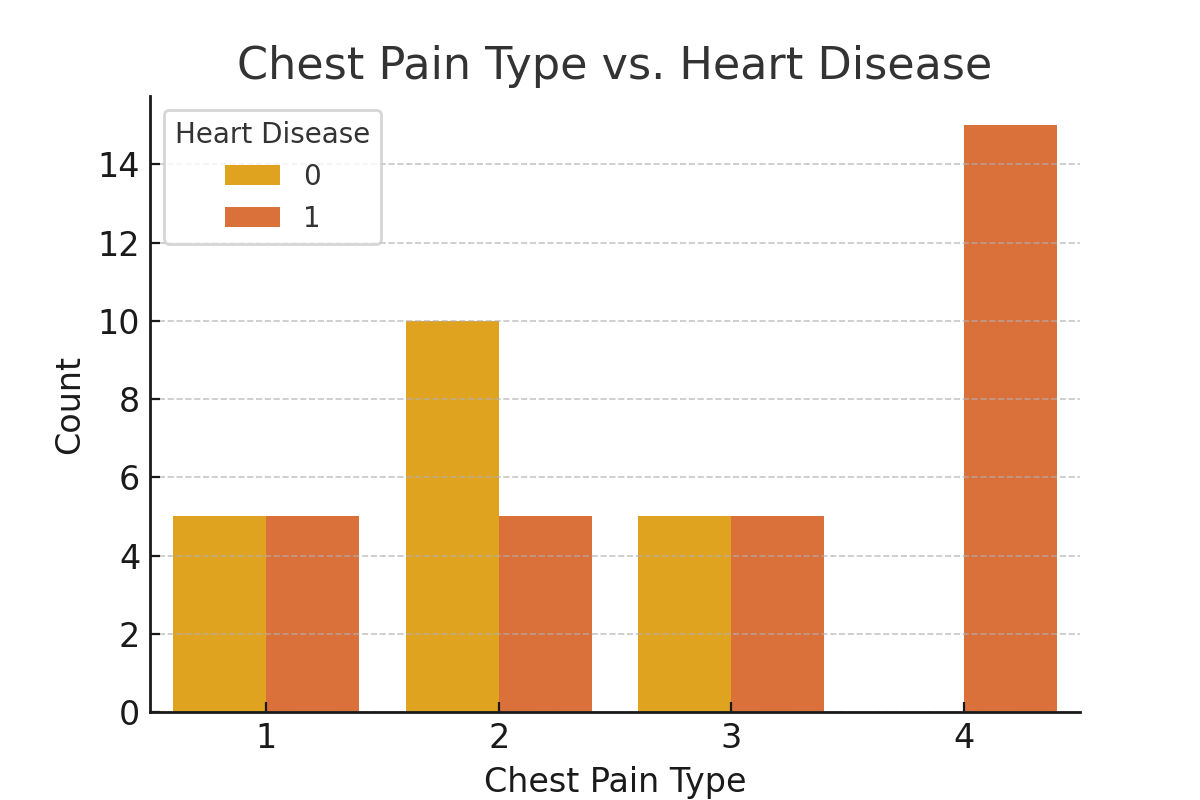
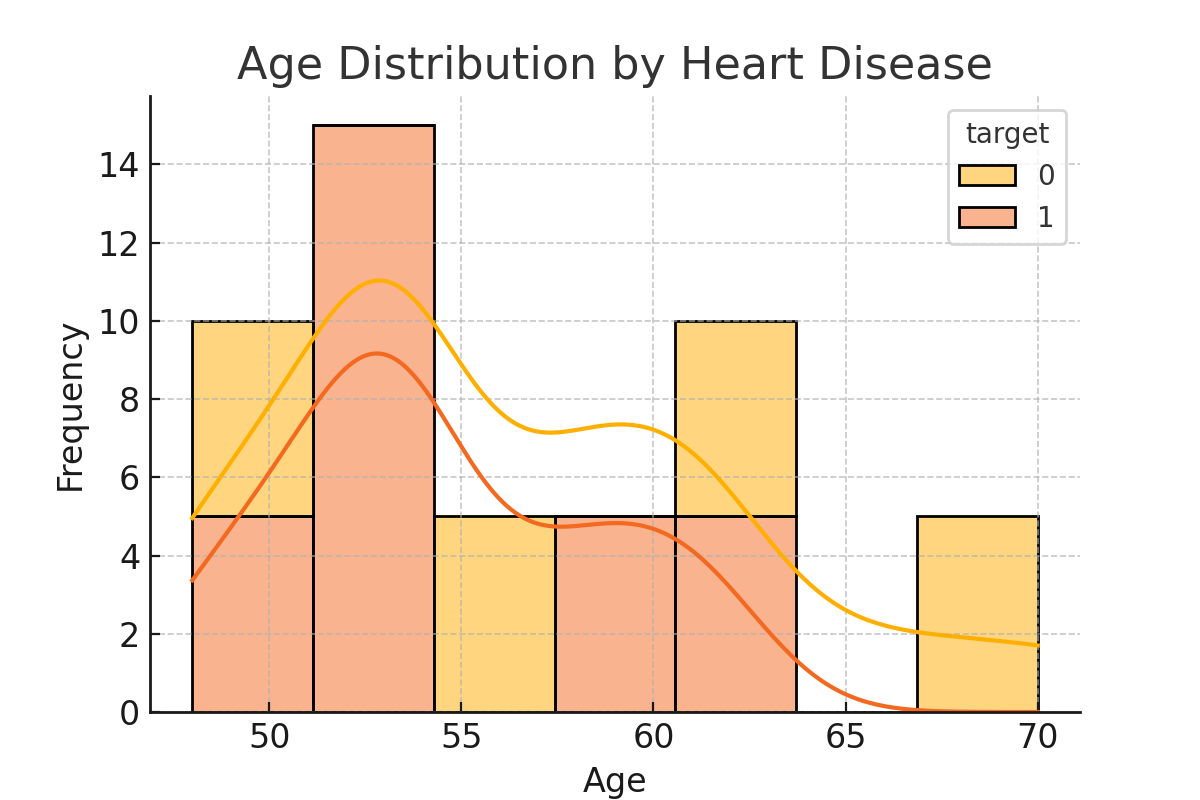


Figure 2: Age Distribution by Heart Disease



# 6. Model Building and Evaluation

A Random Forest classifier was employed to detect heart disease. The model was trained using an 80/20 train-test split and evaluated on accuracy, precision, recall, and F1-score. It demonstrated high performance, especially in detecting patients with heart disease.

# 7. Conclusion

This system accurately predicts heart disease using patient attributes. The Random Forest model proved effective in capturing key patterns from the data. Features like chest pain type, cholesterol, and maximum heart rate showed strong influence on predictions.

# 8. Future Work

* - Further tune the Random Forest hyperparameters  
  - Investigate advanced ensemble models such as XGBoost  
  - Include more diverse and larger datasets  
  - Deploy the model into a clinical decision support tool