

Early Detection of Heart Disease Using Random Forest Classifier

- 2410040036
- 2410040085
- 2410040019
- 2410040079
- 2410040064

Introduction

- Heart disease remains the leading cause of mortality globally.
- Early diagnosis can reduce complications and improve survival rates.
- Traditional medical diagnosis is accurate but may be time-consuming, costly, and dependent on expertise.
- Machine Learning (ML) provides a promising approach to support doctors in predicting risks faster and with data-driven accuracy.
- Our project explores ML for early detection of heart disease.

Problem Statement

- ▶ Current diagnostic techniques are:
 - ▶ Manual and require medical expertise.
 - ▶ Sometimes fail to detect early-stage heart problems.
- ▶ Patients at risk may go undiagnosed until severe symptoms appear.
- ▶ Hence, there is a pressing need for:
 - ▶ An automated predictive system.
 - ▶ High accuracy in classification.
 - ▶ Reliability and scalability in real-world healthcare.

Objectives



To systematically analyze patient health datasets containing clinical parameters.



To build a predictive system using the Random Forest Classifier.



To evaluate the model with Accuracy, Precision, Recall, and F1-score.



To compare the Random Forest model against baseline ML models like Logistic Regression, Decision Tree, or SVM.

Literature Survey

Previous research shows:

- Logistic Regression and Decision Trees achieve moderate accuracy (~70–80%).
- SVM provides good results but may overfit with noisy data.

Random Forest

- An ensemble learning algorithm, combines multiple decision trees for higher stability and accuracy.

Research Gap:

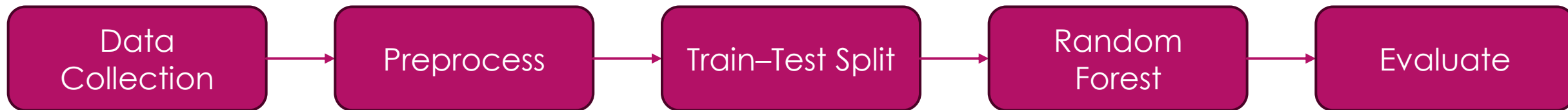
- Need for generalizable and reliable models suitable for diverse patient data.
- Our project addresses this by leveraging Random Forest's strengths.

Dataset

- ▶ Source: UCI Heart Disease / Kaggle
- ▶ Features: Age, Sex, Resting BP, Cholesterol, ECG, Max HR, etc.
- ▶ Target: Presence/Absence of heart disease
- ▶ Preprocessing: missing values, scaling, feature selection

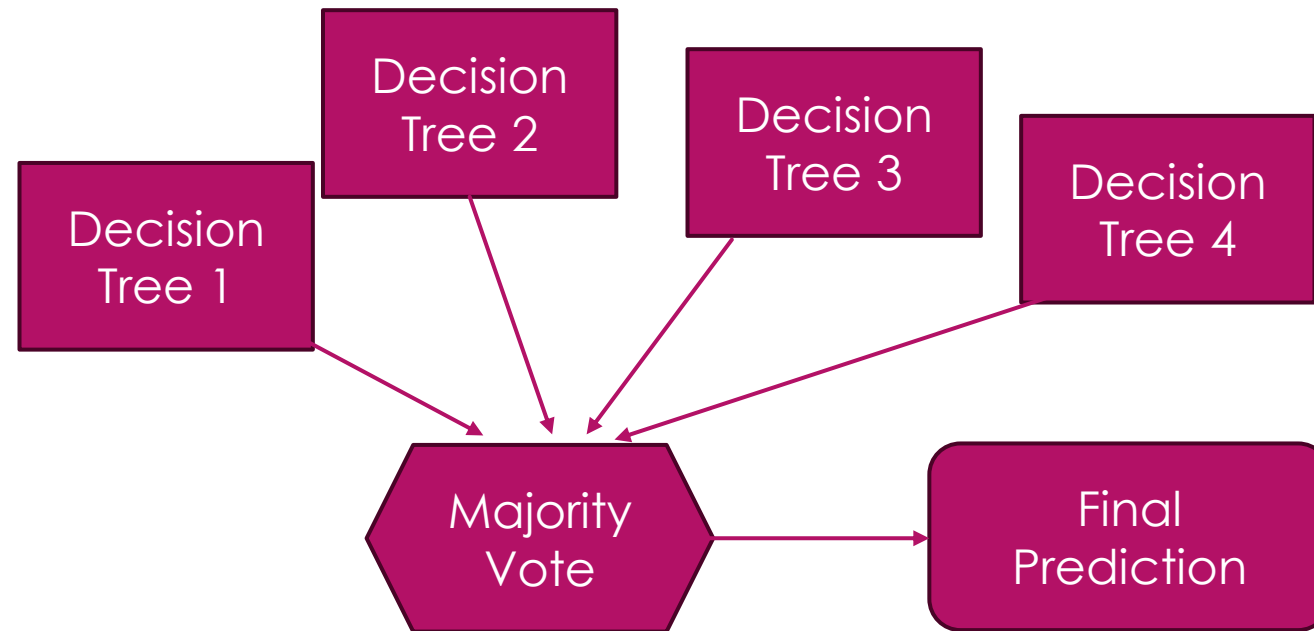
Methodology

- ▶ Data Collection
- ▶ Preprocessing (cleaning & normalization)
- ▶ Train–Test Split
- ▶ Model: Random Forest Classifier
- ▶ Evaluation: Confusion Matrix, ROC–AUC, F1



Random Forest Classifier

- ▶ Random Forest is an ensemble method combining multiple decision trees.
- ▶ Works using:
 - ▶ Bagging (Bootstrap Aggregation) → Random subsets of data.
 - ▶ Random Feature Selection → Reduces correlation between trees.
- ▶ Each tree votes → Final prediction is based on majority voting.
- ▶ Advantages:
 - ▶ High accuracy.
 - ▶ Robust against overfitting.
 - ▶ Handles high-dimensional medical datasets efficiently.



THANK YOU