Project Synopsis

Prediction of Heart Diseases using Random Forest

Submitted By:

Aakash Singh (Roll No.: 2100290109001)
Lalit Kishor (Roll No.: 2200290109008)
Suhel Khan (Roll No.: 2200290109016)

Under the Supervision of:

Prof. Gaurav Parashar
KIET Group of Institutions, Ghaziabad
Session: 2024–25

1. Introduction

Cardiovascular diseases are a leading cause of mortality globally. Accurate and early diagnosis is crucial for effective intervention and prevention. In this project, a machine learning-based model, particularly using the Random Forest algorithm, has been developed to predict heart disease risk using clinical and demographic data.

2. Objectives

• Build a machine learning model for heart disease prediction.
• Evaluate and compare various algorithms: Logistic Regression, SVM, Decision Trees, Random Forest, ANN.
• Identify significant health indicators influencing heart disease.
• Create an interactive web interface for real-time predictions.

3. Methodology

The dataset comprises 303 records with 14 clinical features. After preprocessing (handling missing values, normalization, and feature selection), models were trained and evaluated using metrics like accuracy, sensitivity, specificity, and AUC. Python and Jupyter Notebook were used for modeling; frontend integration was done using HTML, CSS, and JavaScript.

4. Results

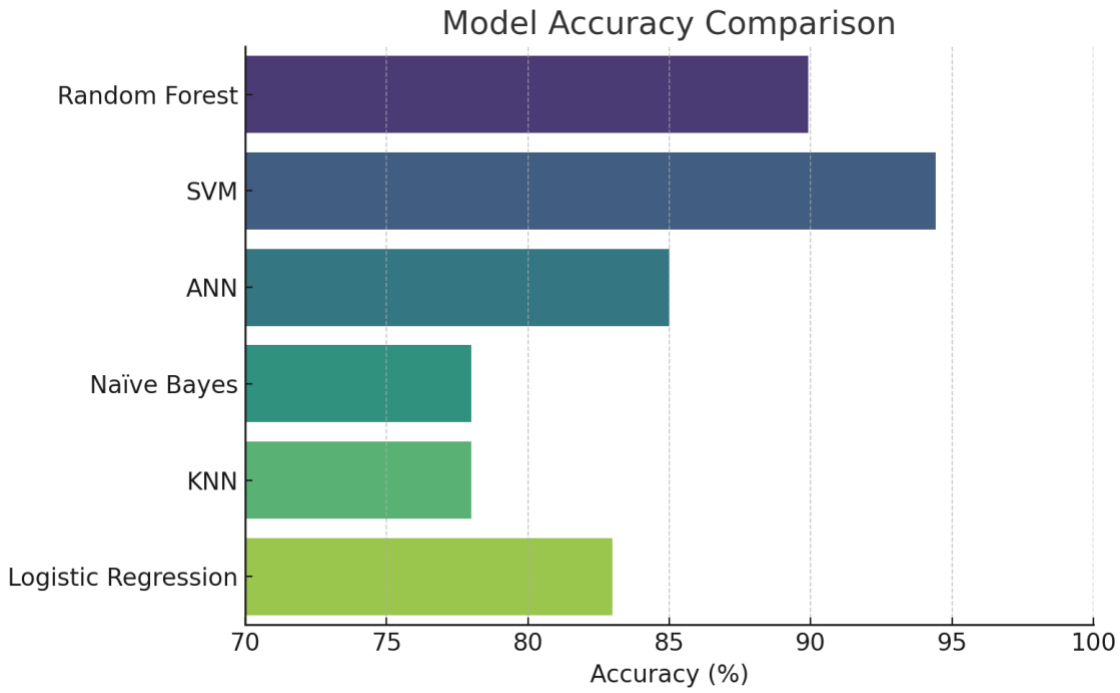
The Random Forest model achieved:
• Accuracy: 89.92%
• Sensitivity: 91.58%

• Specificity: 87.67%
• AUC: 94.16%

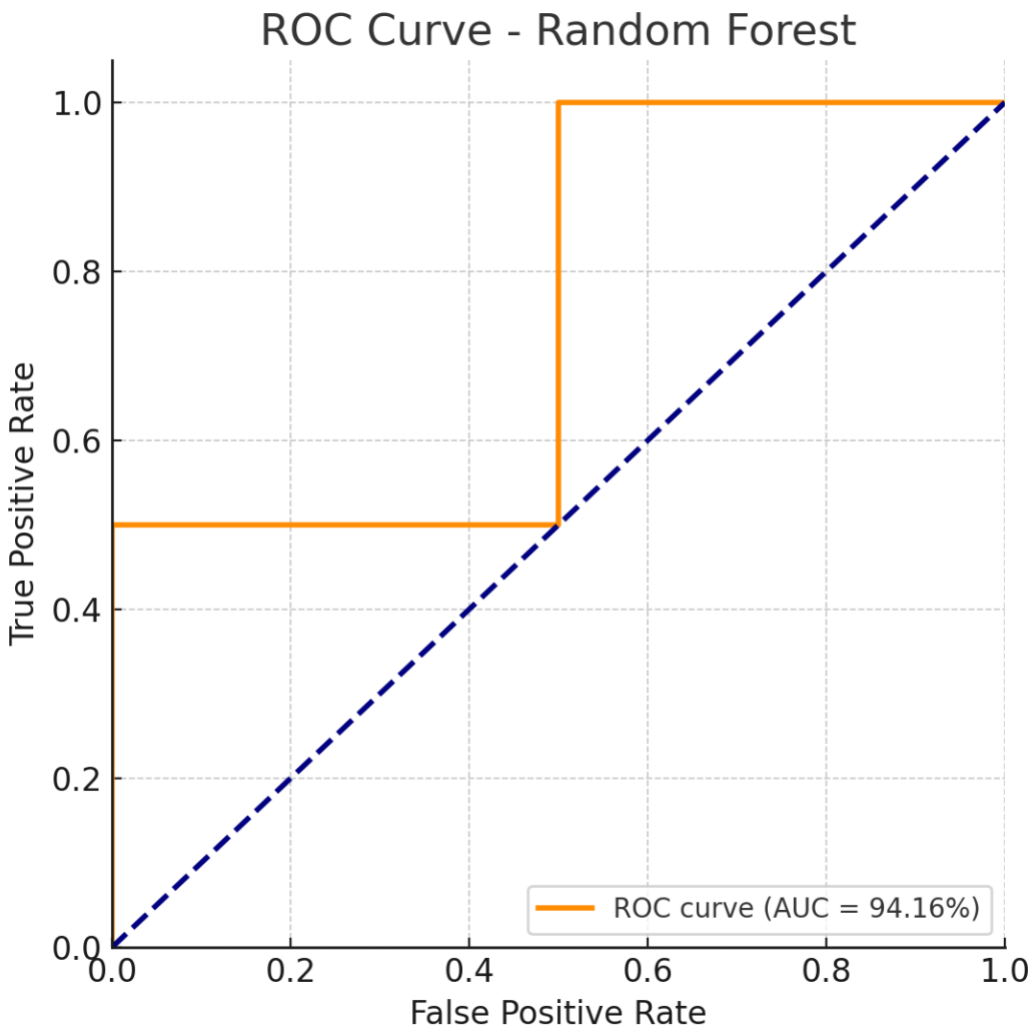
These metrics indicate a robust and generalizable model for early heart disease prediction.

5. Graphs and Visualizations

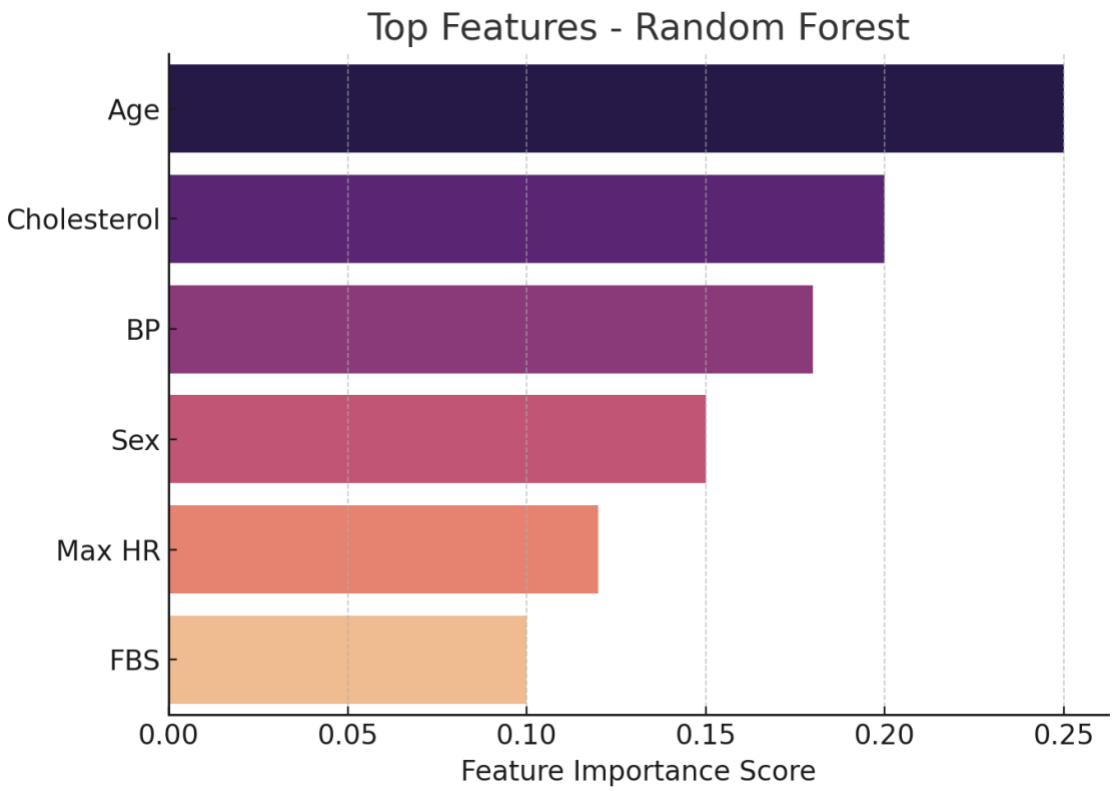
1. 5.1 Model Accuracy Comparison



2. 5.2 ROC Curve - Random Forest



3. 5.3 Feature Importance - Random Forest



6. Conclusion

The Random Forest algorithm demonstrated strong potential for heart disease prediction, validated by high accuracy and AUC scores. The model has been integrated into a user-friendly web application, offering real-time predictive support to users and healthcare practitioners.

7. Future Scope

- Real-time data integration from wearable devices and EHRs.
- Clinical validation and scalability testing.
- Expansion to predict other chronic diseases.
- Enhanced user engagement and health education features.