Predictive Analytics for early detection of cardiovascular diseases using machine learning techniques

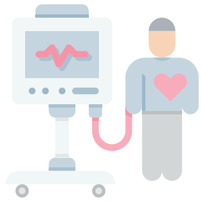
Team Members

Mannem Pardhava (Y22CB041)
Edara Karthik (Y22CB019)

Project Guide

Dr. T. Anuradha

Abstract



Heart disease is among the most common chronic conditions, and early detection is paramount in reducing mortality rates by mitigating the severity of cardiovascular diseases (CVD). Machine learning algorithms turn out to provide an effective way for identifying risk indicators. The approach presents a new technique to integrate Recursive Feature Elimination in different machine learning models like Logistic regression, Support Vector Machine, Random Forest, and K-nearest Neighbours etc for more accurate prediction of cardiac disease by parameter tuning. In the proposed approach, RFE will be applied to select the most significant features, and then the models will be trained with these features. These individual model predictions are further combined into a final prediction using an ensemble method, specifically a Voting Classifier. The models were evaluated against patients' health records characterized by key CVD features. The results show that with the implementation of the ensemble method, the accuracy of prediction significantly increases to the point where it performs better than the individual models in isolation. This proposal shows that the combination of the RFE-based feature selection approach with ensemble machine learning models can be effectively used in predicting cardiovascular disease and hence, provides a means for its early diagnosis, reducing fatalities associated with heart diseases.

Introduction

Heart disease is a major health concern worldwide, and early detection is crucial for improving treatment outcomes and reducing mortality. Traditional methods for diagnosing heart disease can be time-consuming and invasive. We aims to develop a more accurate and efficient approach using machine learning techniques. By leveraging the power of machine learning algorithms, we can analyze large datasets of patient health records to identify patterns and risk factors associated with heart disease.

MEDICAL REPORT

Existing System

The existing machine learning system is based on gradient boosting, a machine learning technique that combines multiple weak learners into a single to improve the accuracy.

Disadvantages

- Increased Risk of Overfitting
- Sequential Model Building
- Higher Computational Cost
- Handling Noisy Data
- Challenging for Scaling

Proposed System

The proposed machine learning system aims to overcome the limitations of the existing system by implementing Voting Classifier technique.

Advantages

- **Reduced Risk of Overfitting**
- **Parallel Model Building**
- **Faster Training Times**
- **Robustness to Noisy Data**
- **More Scalable**

Hardware Requirements

HP Pavilion 15

Intel i5 Processor

16GB RAM

Software Requirements

Windows 11

Google colab

VS Code Software

Python

System Architecture

Heart Disease Dataset

Data Engineering

Feature Selection using RFE

Training Models

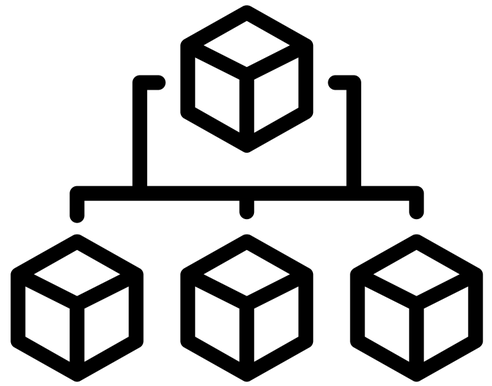
Ensembling Classifier

Performance Estimation

AUROC analysis

Comparision

MODULES



User Interface and Input Handling Module

The User Interface module lets users enter medical report details and see predicted result, using Streamlit.

Mapping input features with RFE features

The inputs given by the user are mapped with the features selected in RFE and send for the prediction

Prediction by the model

This module consists the model which predicts the result and returns to the User Interface

Algorithm

Step 1: Load Dataset

-Load dataset with features: Age, Sex,
ChestPainType, RestingBP, Cholesterol, etc.

Step 2: Data Preprocessing

-Handle missing values
-Split into training (X_train, y_train) and testing (X_test, y_test).

Step 3: Feature Selection with RFE

-Use Random Forest as the base estimator
-Apply Recursive Feature Elimination (RFE) to select the top features

Step 4: Hyperparameter Tuning

-Tune Random Forest and SVM using GridSearchCV and RandomizedSearchCV

Step 5: Model Training

-Train models individually

-Evaluate performance (accuracy, precision, recall, F1-score)

Algorithm

Step 6: Ensemble Learning with Voting Classifier

-Combine models with a Voting Classifier (soft voting)
-Train the ensemble on selected features

Step 7: Model Evaluation

-Calculate accuracy, precision, recall, F1-score, and AUC
-Compare ensemble with individual models

Step 8: ROC Curve Plotting

-Plot ROC curves for each model

-Calculate AUC for comparison

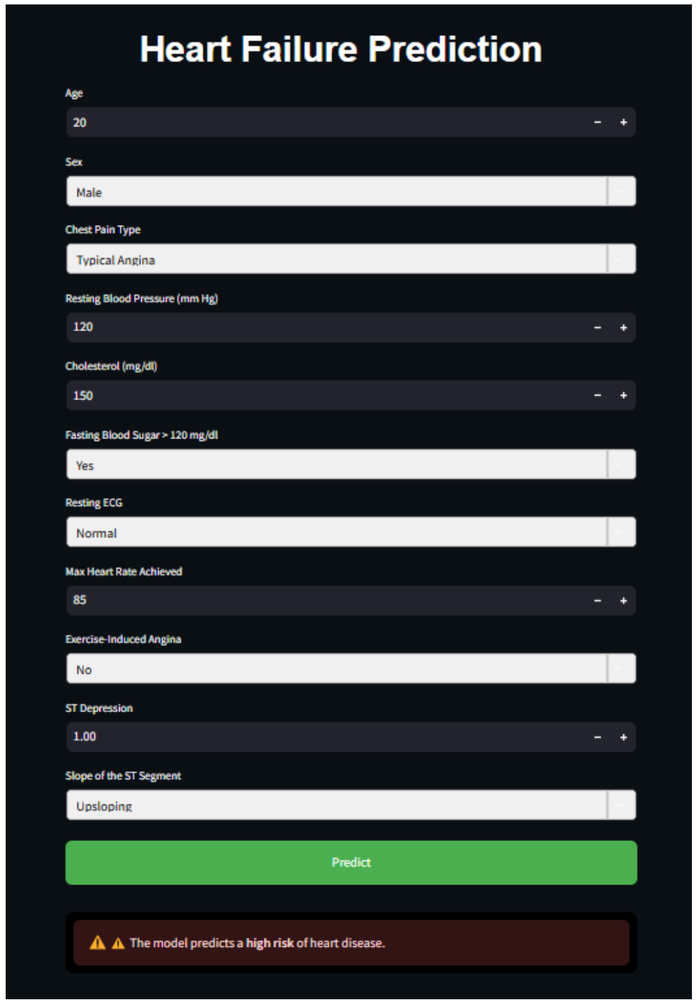
Step 9: Performance Comparision

-Select the best-performing model based on evaluation

Result

```
print(f'Ensemble Accuracy: {accuracy_score(y_test, y_pred_ensemble)}')
✓ 0.0s
Ensemble Accuracy: 0.9078260869565217
```

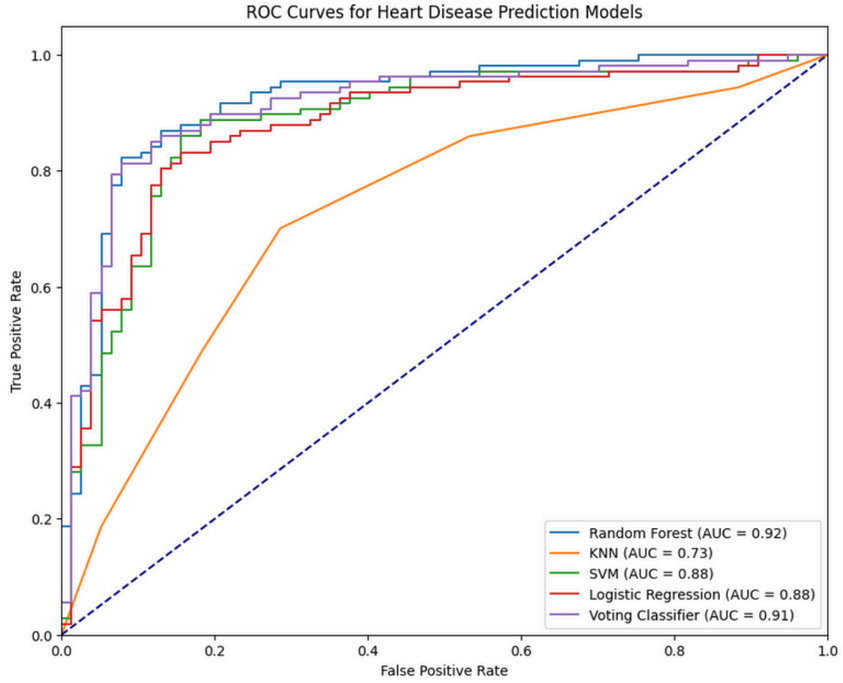
Accuracy



User Interface

```
print(f" Precision: {precision_weighted:.2f}\n")
print(f" Recall: {recall_weighted:.2f}\n")
print(f" F1-Score: {f1_weighted:.2f}")
✓ 0.0s
Precision: 0.85
Recall: 0.85
F1-Score: 0.85
```

Precision, Recall, F1-Score



ROC Curves

Our approach successfully demonstrates the power of machine learning in predicting heart failure by leveraging advanced techniques like Recursive Feature Elimination (RFE) for feature selection and ensemble learning with multiple classifiers.This model offers a valuable tool for early identification of heart disease, providing healthcare professionals with reliable decision-making support that can potentially reduce mortality rates associated with cardiovascular diseases.

REFERENCE

1.Chicco, D., & Jurman, G. (2020). "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone." BMC Medical Informatics and Decision Making, 20(1), 1-16.

2. Kora, P., & Kalva, S. K. (2015). "Hybrid machine learning algorithms for heart disease prediction." Procedia Computer Science, 70, 85-91.

3. Ahmad, F., et al. (2018). "Efficient heart disease prediction through feature selection." Procedia Computer Science, 132, 371-378.

4. Khedkar, D. D., & Kharat, P. M. (2017). "Prediction of heart disease using machine learning algorithms: A survey." International Journal of Advanced Research in Computer Science, 8(3).

5. Al-Milli, N. (2013). "Backpropogation neural network for prediction of heart disease." Journal of Theoretical and Applied Information Technology, 56(1), 131-135.

Thank You

For your attention and cooperation

-TEAM 7