

Identification of Risk Factors for Heart Disease Using Statistical Methods and Machine Learning

A20521362

Young Sun Lee



Heart disease



Typical heart diseases:
Myocardial infarction, Heart failure, Arrhythmia



Heart disease is a condition that competes with cancer as the **leading cause of death**



Heart disease carries the risk of sudden death
→ **Prediction** is important



Objective:
To identify the **most important factors** for early detection



Dataset

- **Heart Failure Prediction Dataset**

from Kaggle

- **12 variables** and **918 observations**

11 predictive variables

1 response variable

Variables

Age: age of the patient [years]

Sex: sex of the patient

- M: Male
- F: Female

ChestPainType: chest pain type

- TA: Typical Angina
- ATA: Atypical Angina
- NAP: Non-Anginal Pain,
- ASY: Asymptomatic

RestingBP: resting blood pressure [mmHg]

Cholesterol: serum cholesterol [mm/dl]

FastingBS: fasting blood sugar

- 1: if FastingBS > 120 mg/dl
- 0: otherwise

RestingECG: resting electrocardiogram results

- Normal: Normal
- ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05)

MaxHR: maximum heart rate achieved [Numeric value between 60 and 202]

ExerciseAngina: exercise-induced angina

- Y: Yes
- N: No

Oldpeak: oldpeak [Numeric value measured in depression]

ST_Slope: the slope of the peak exercise ST segment

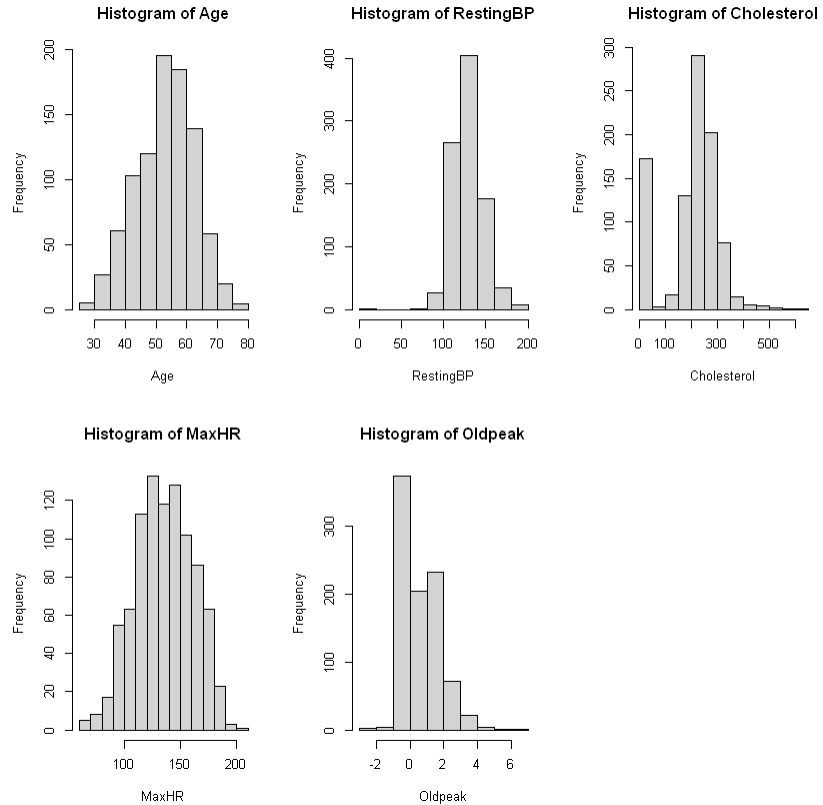
- Up: upsloping
- Flat: flat
- Down: downsloping

HeartDisease: output class

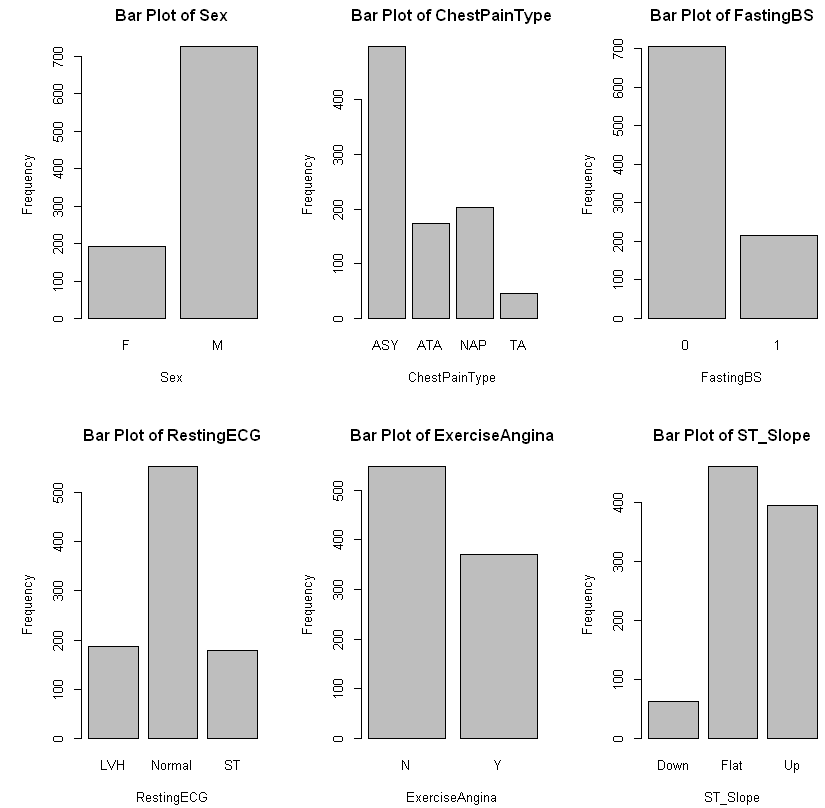
- 1: heart disease
- 0: normal

Distribution of variables

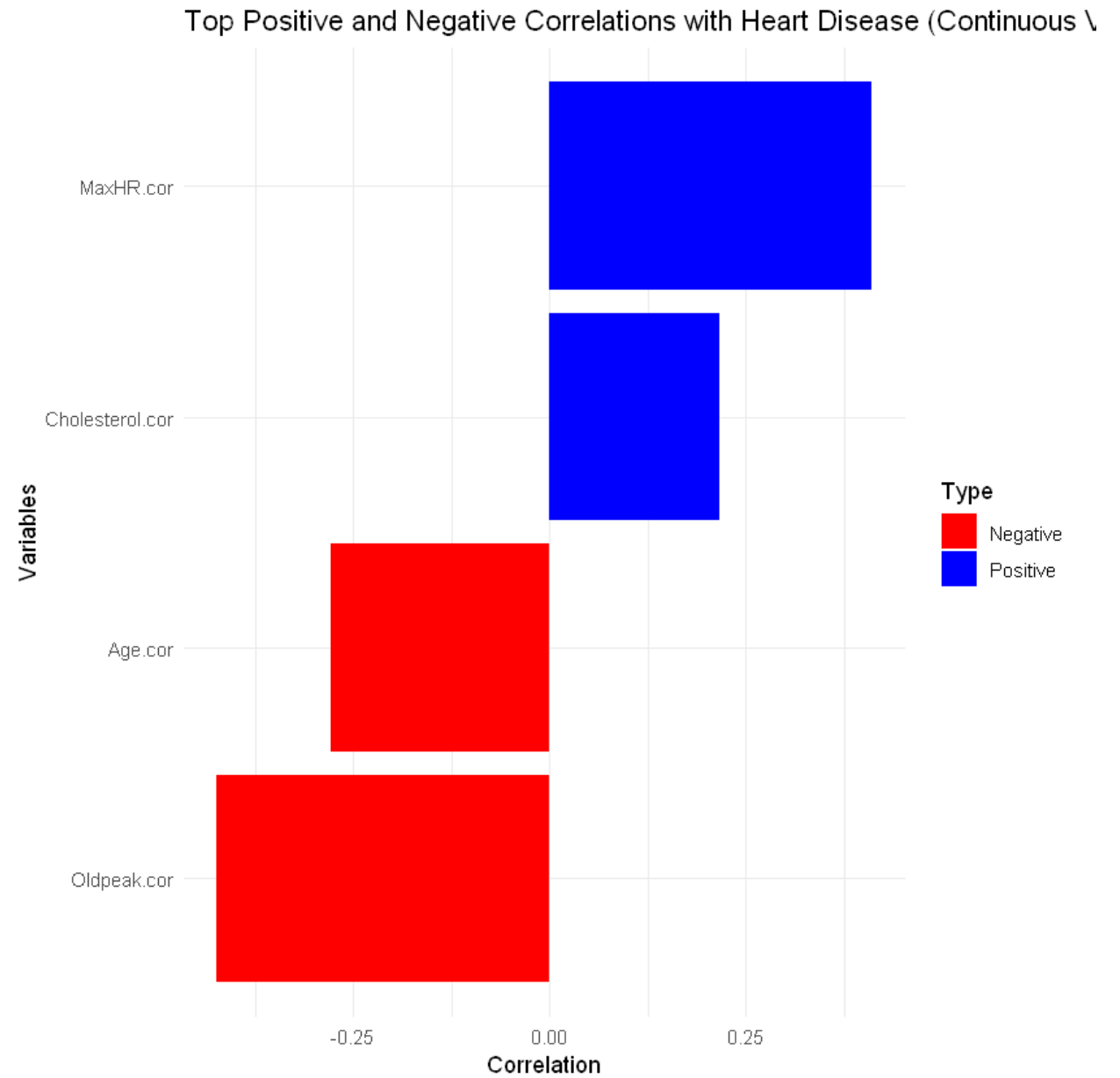
Continuous variables



Categorical variables

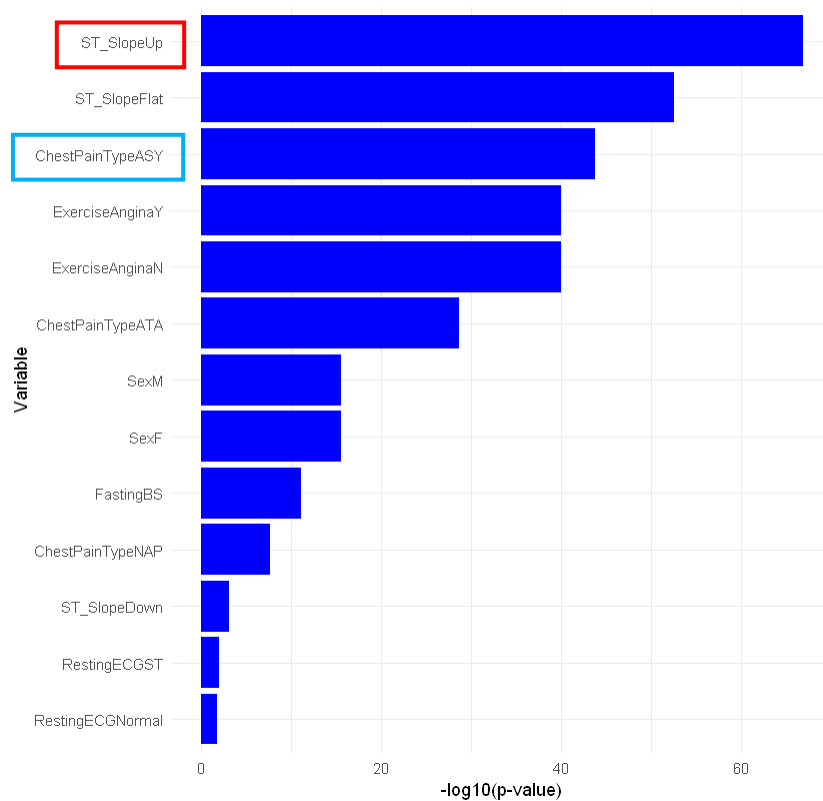


Correlation coefficient for continuous variables

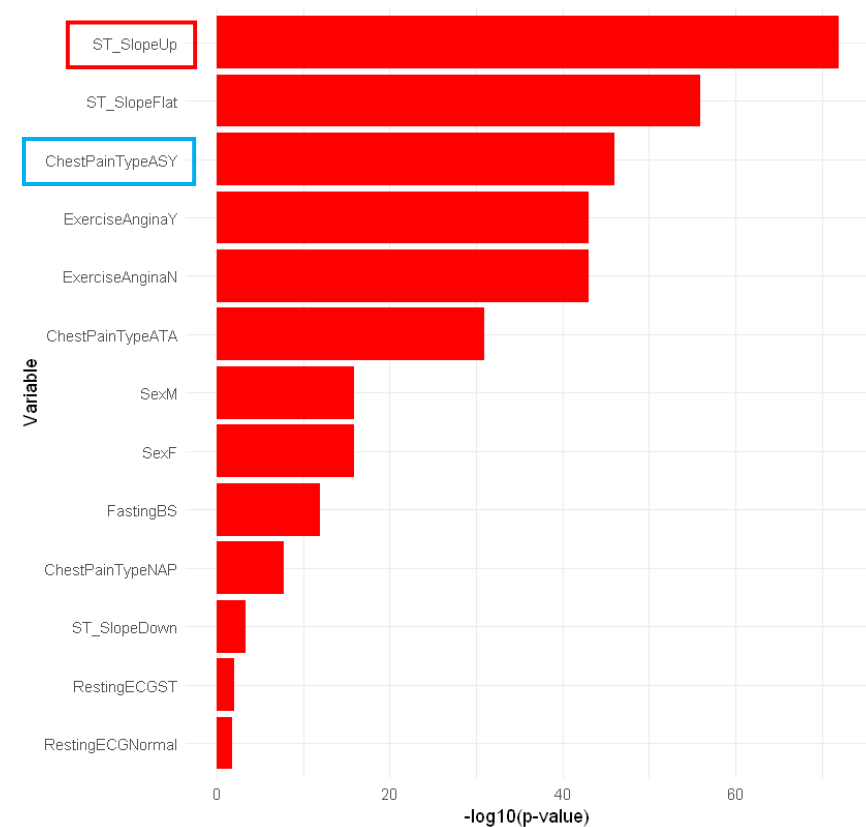


Chi-squared and Fisher's exact tests for categorical variables

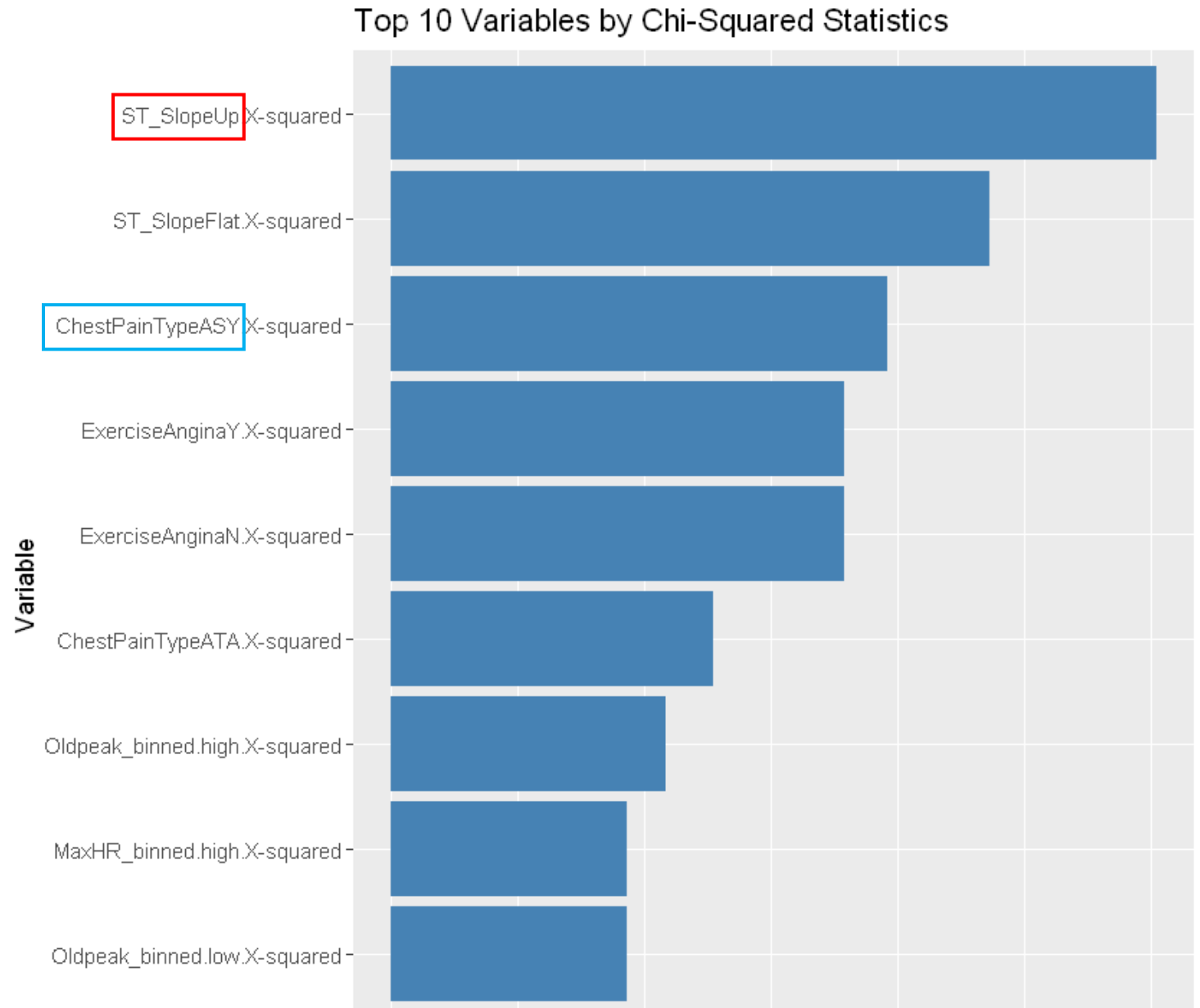
Significant Chi-Squared Test Results



Significant Fisher's Exact Test Results



Chi-square Test with All Variables Categorized



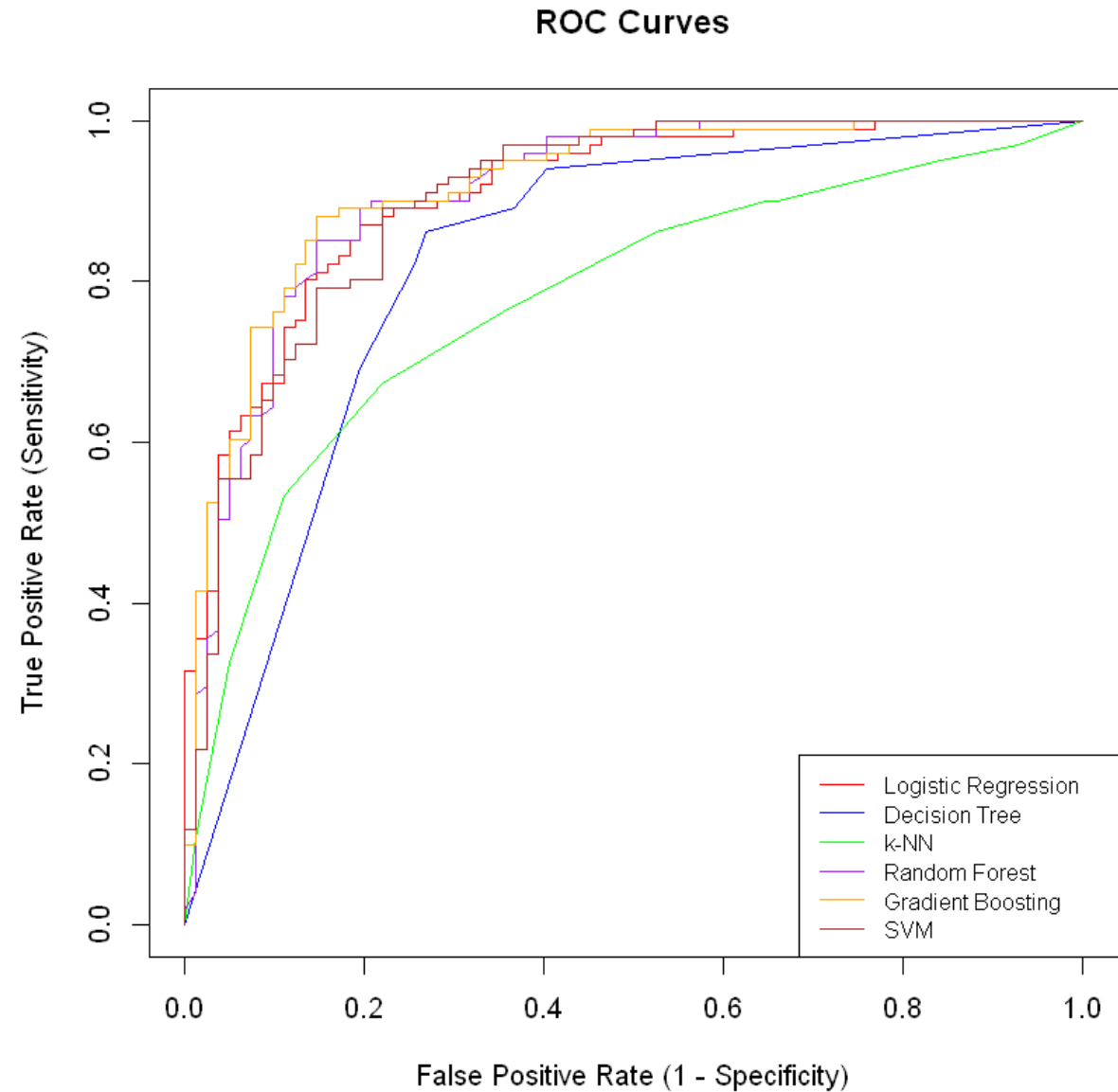
Model Performance Evaluation and Ranking of Variable Importance by Model

- Logistic Regression
- Decision Tree
- K-NN
- Random Forest
- Gradient Boosting
- SVM

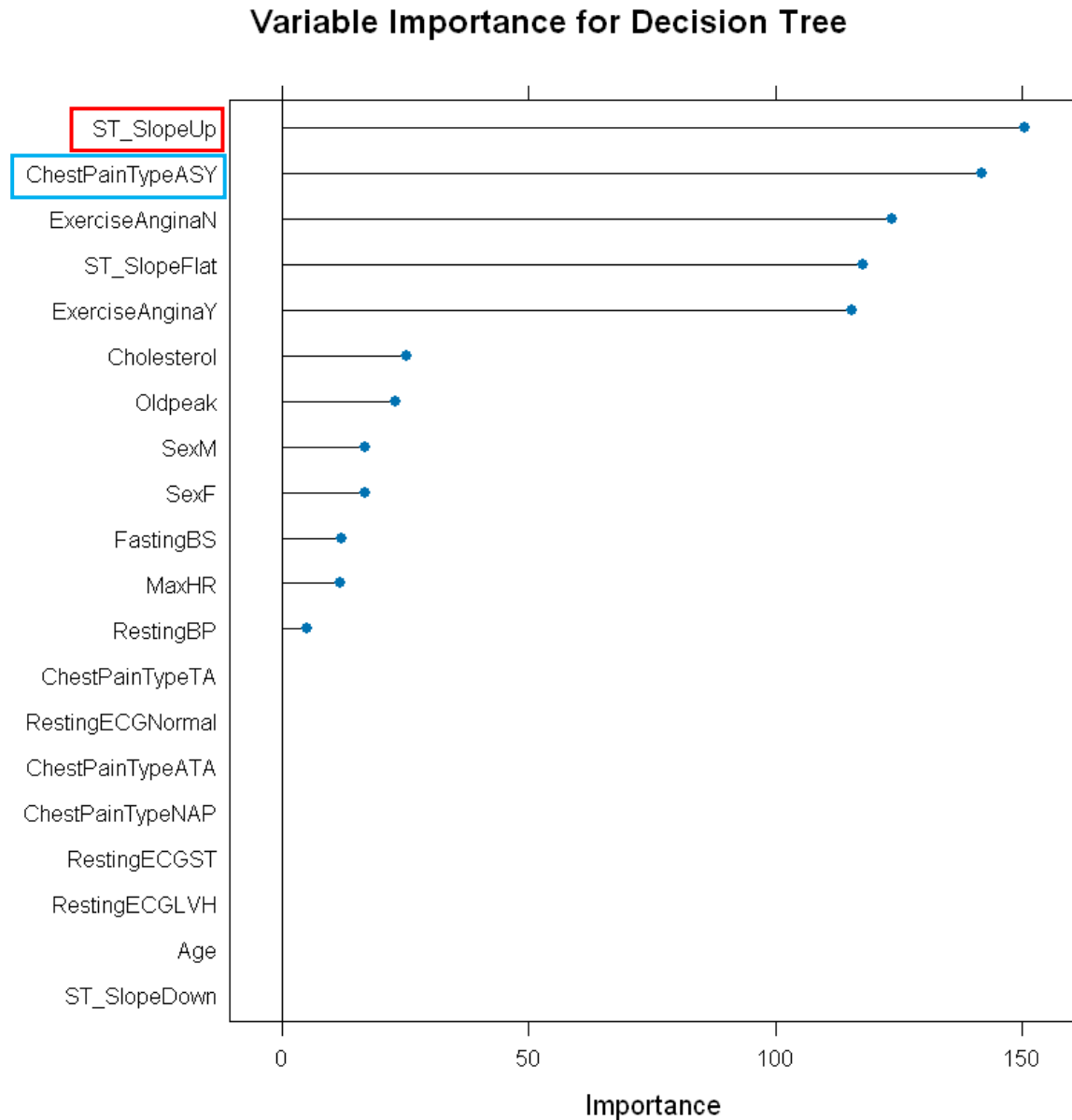
Model Performance

Model	Precision	Recall	F1 Score	AUC
Gradient Boosting	0.87	0.89	0.88	0.92
Random Forest	0.85	0.89	0.87	0.91
Logistic Regression	0.85	0.87	0.86	0.91
SVM	0.83	0.86	0.84	0.90
Decision Tree	0.80	0.86	0.83	0.82
k-NN	0.73	0.76	0.74	0.78

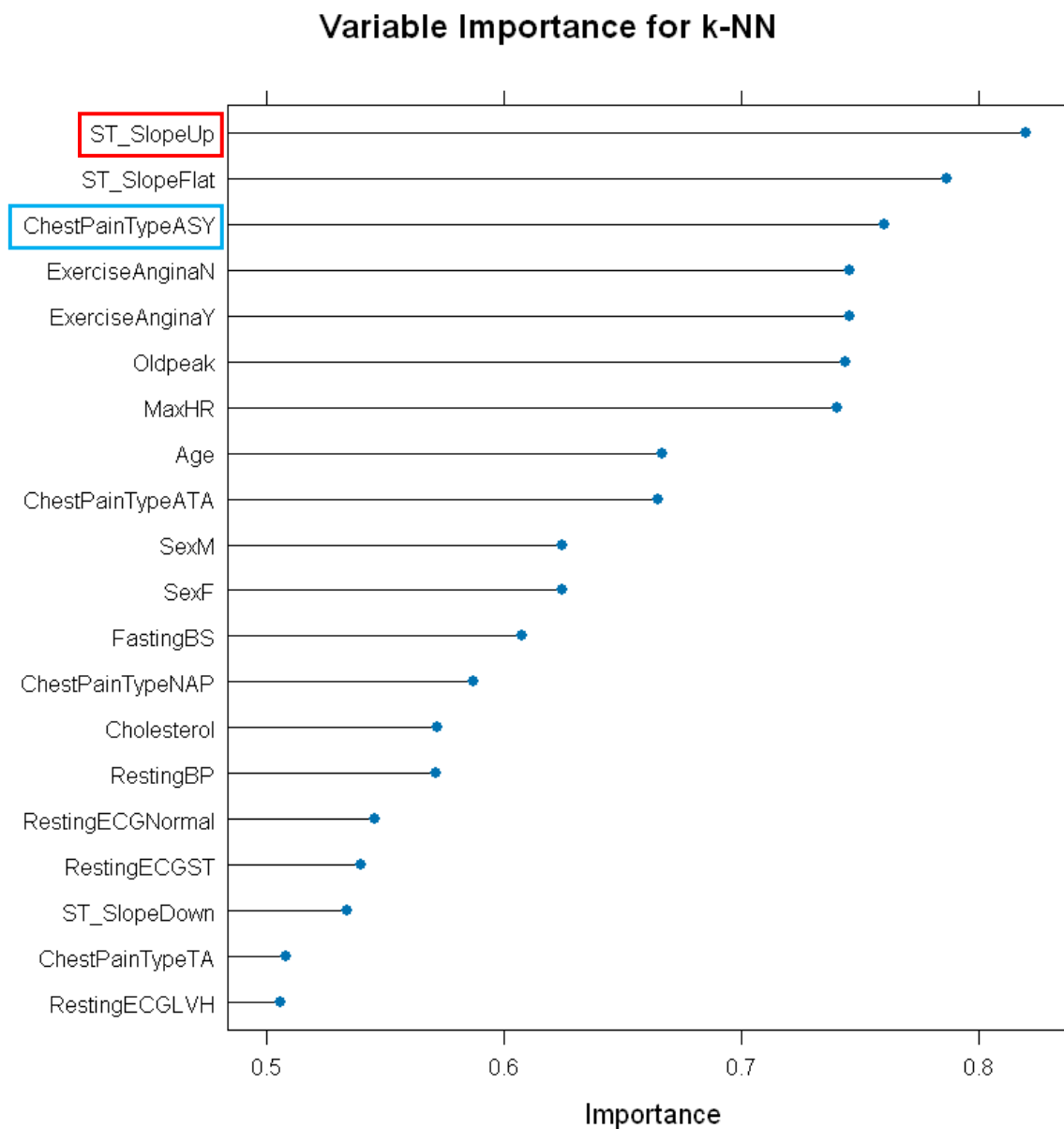
Model Performance Evaluation



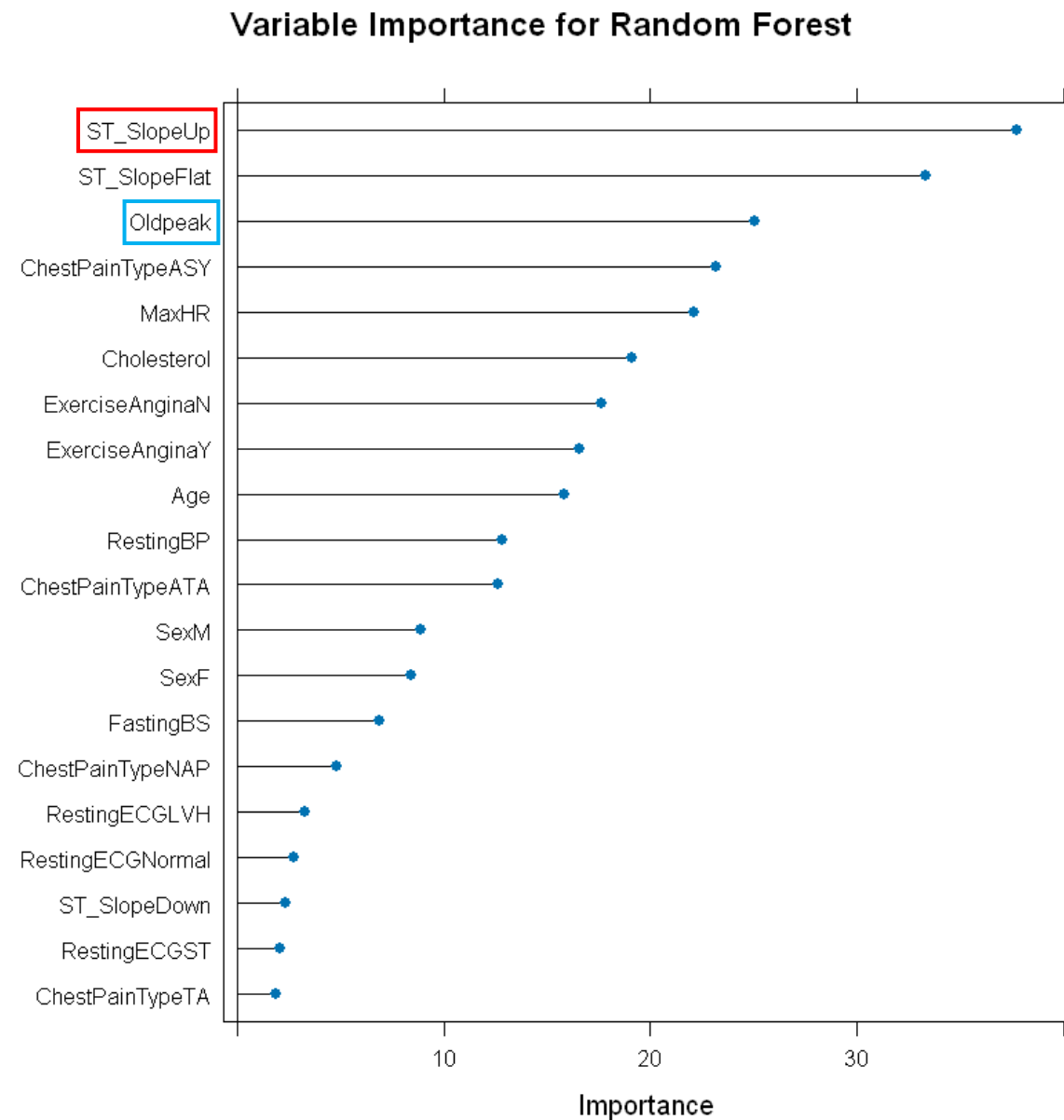
Variable importance Decision Tree



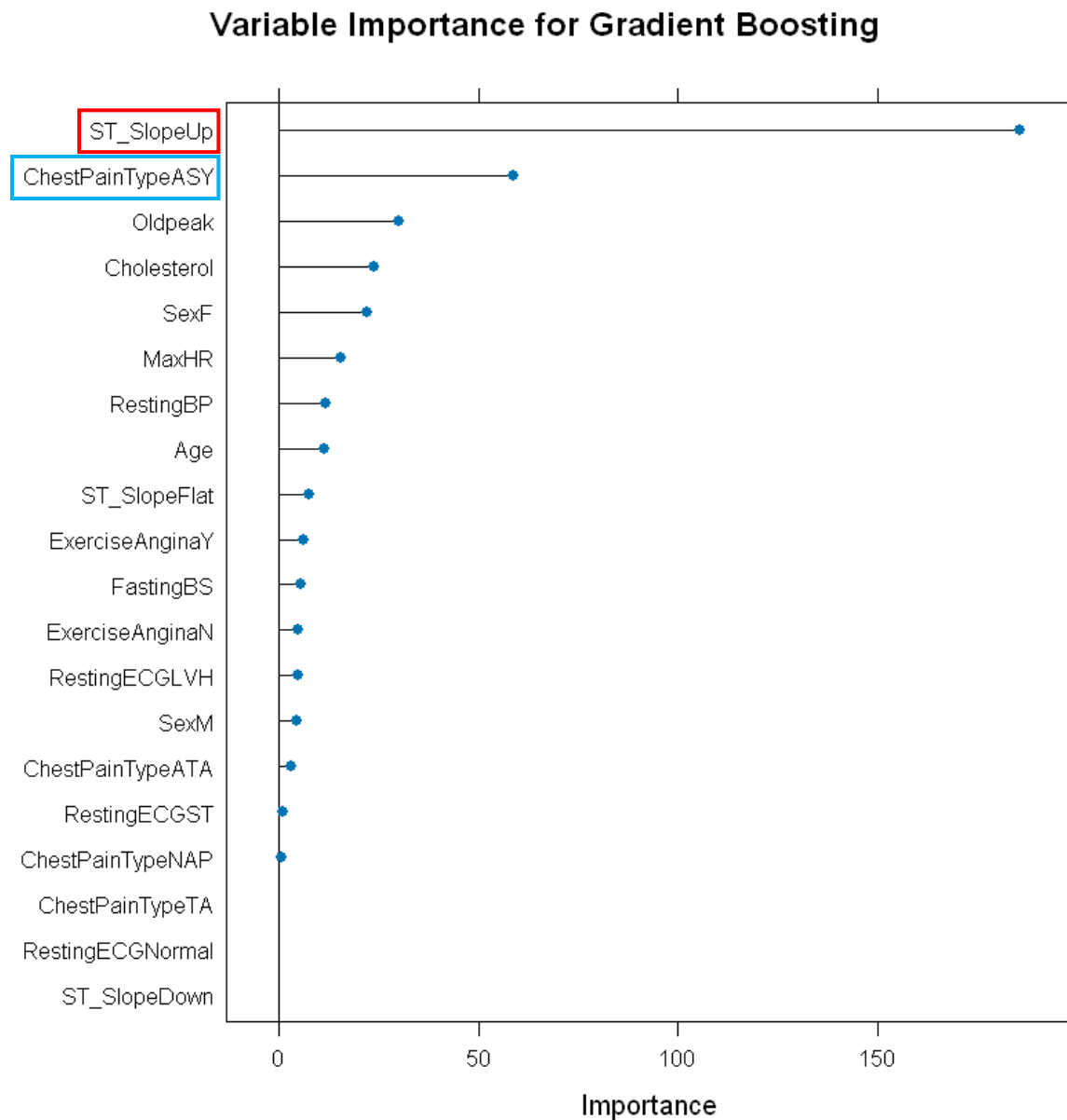
Variable importance k-NN



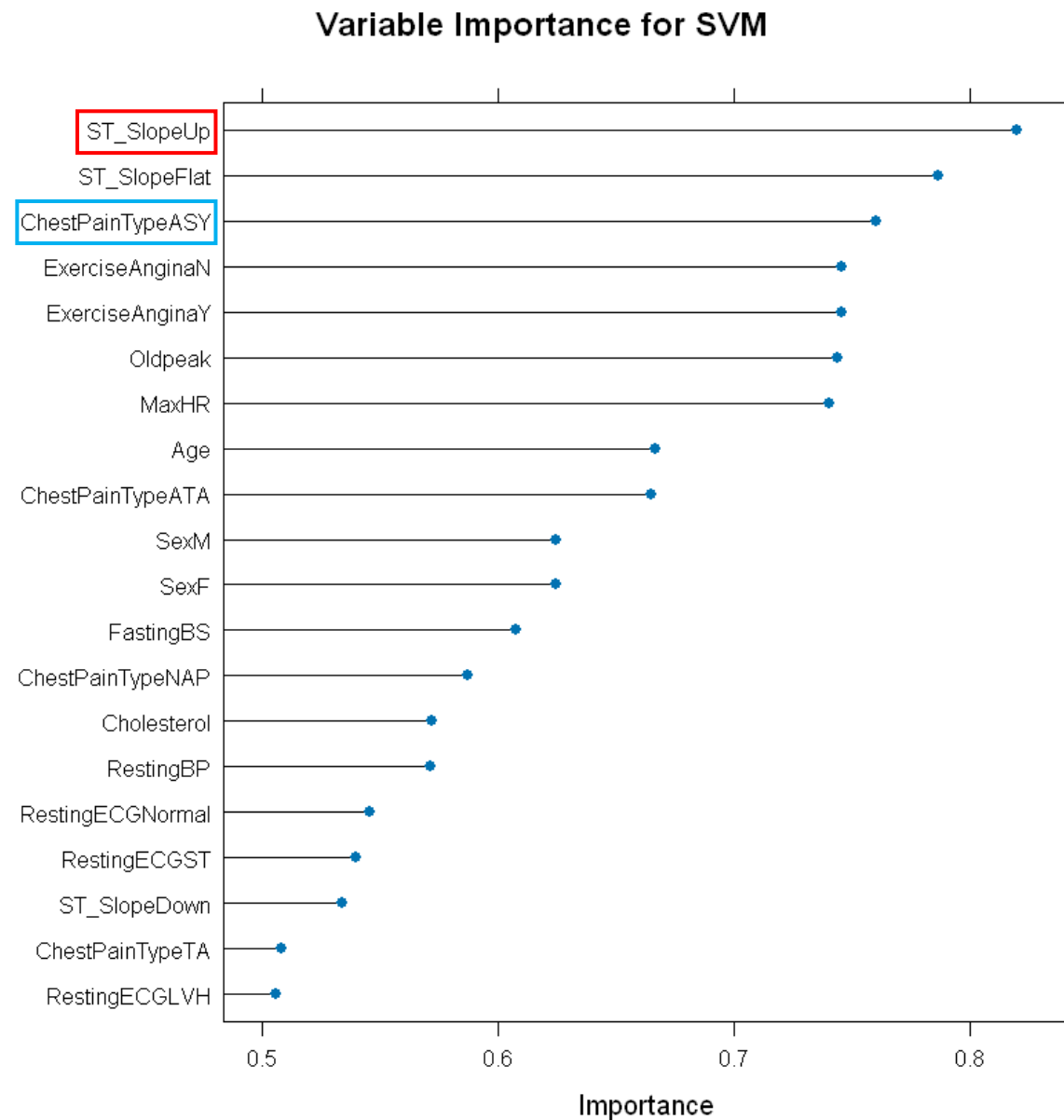
Variable importance Random Forest



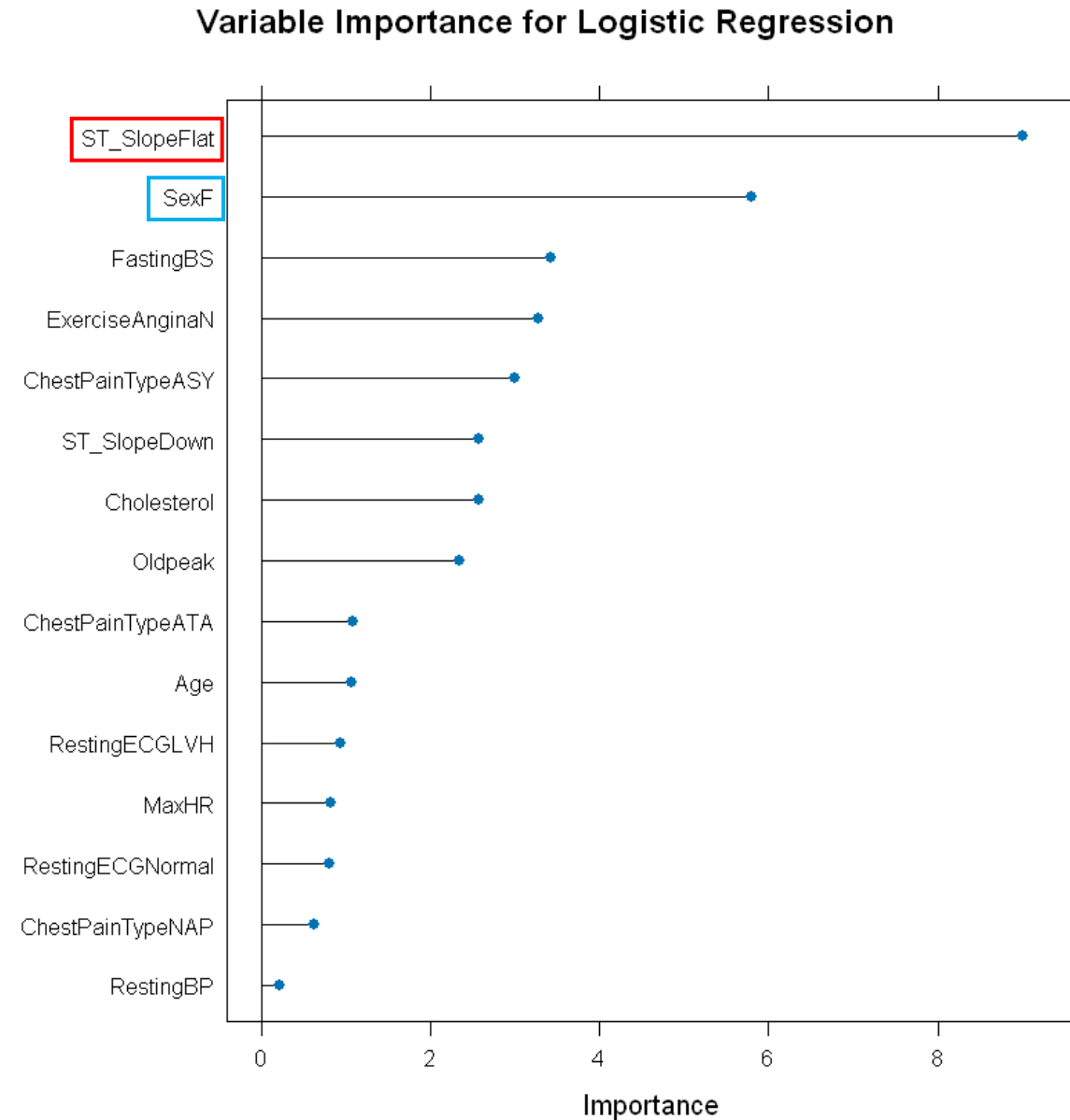
Variable importance Gradient Boosting



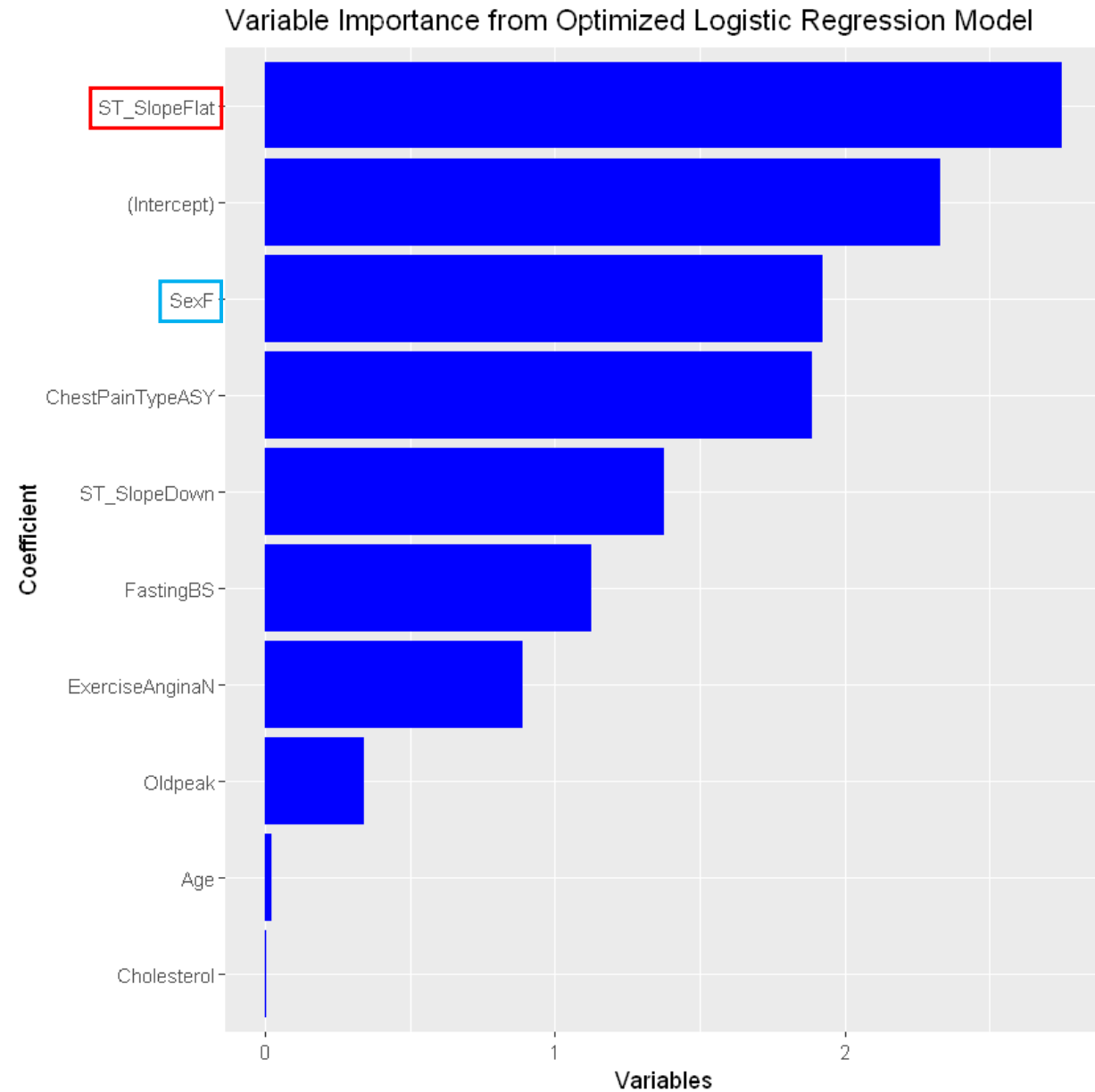
Variable importance SVM



Variable importance Logistic Regression

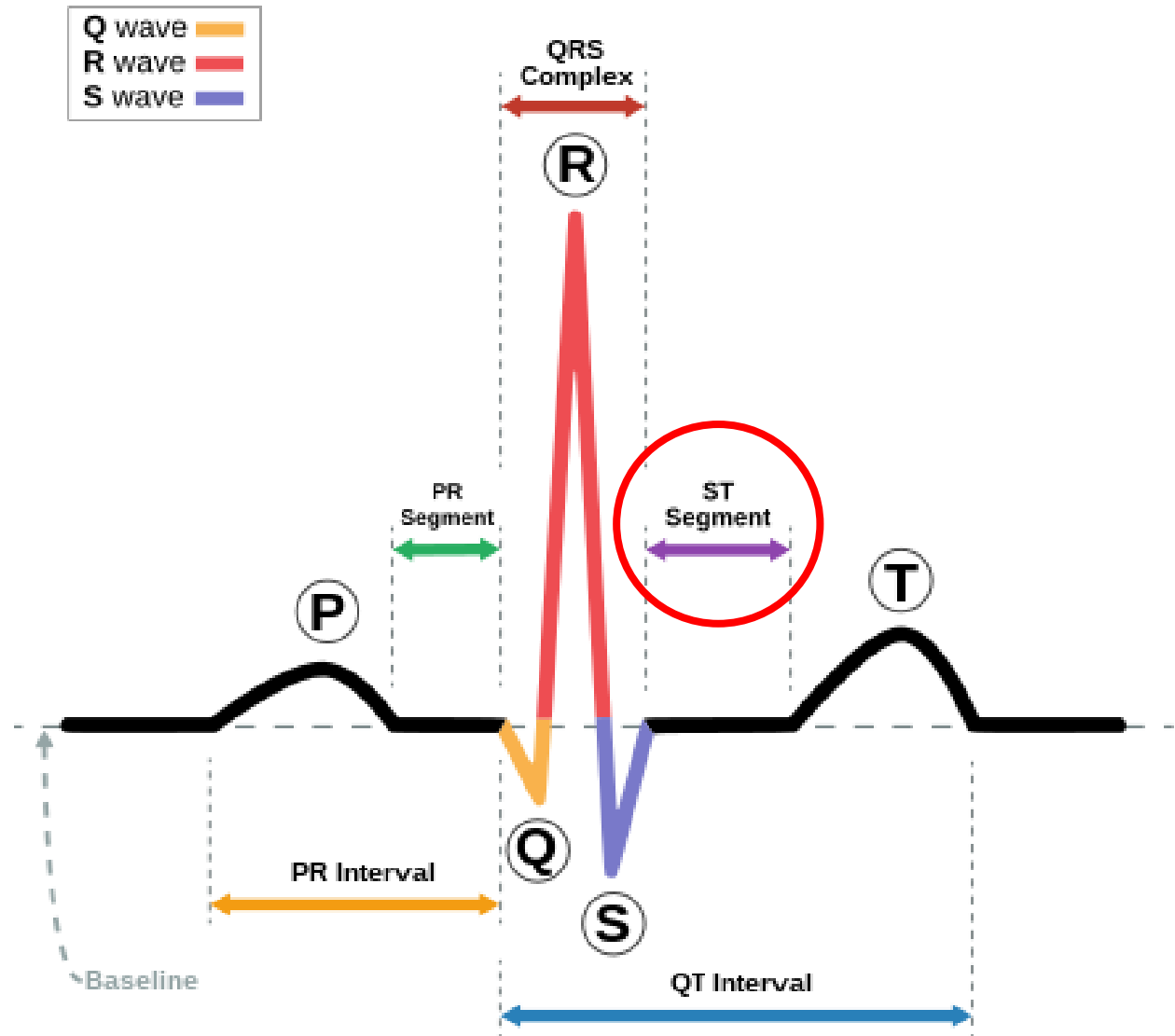


Variable importance Logistic Regression (Optimized)



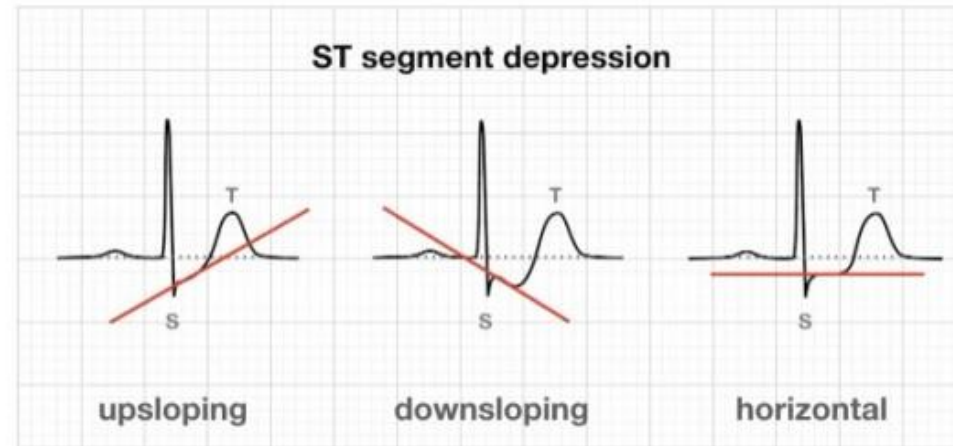
ST Segment:

Connects the QRS complex and the T Wave



Exercise stress electrocardiogram

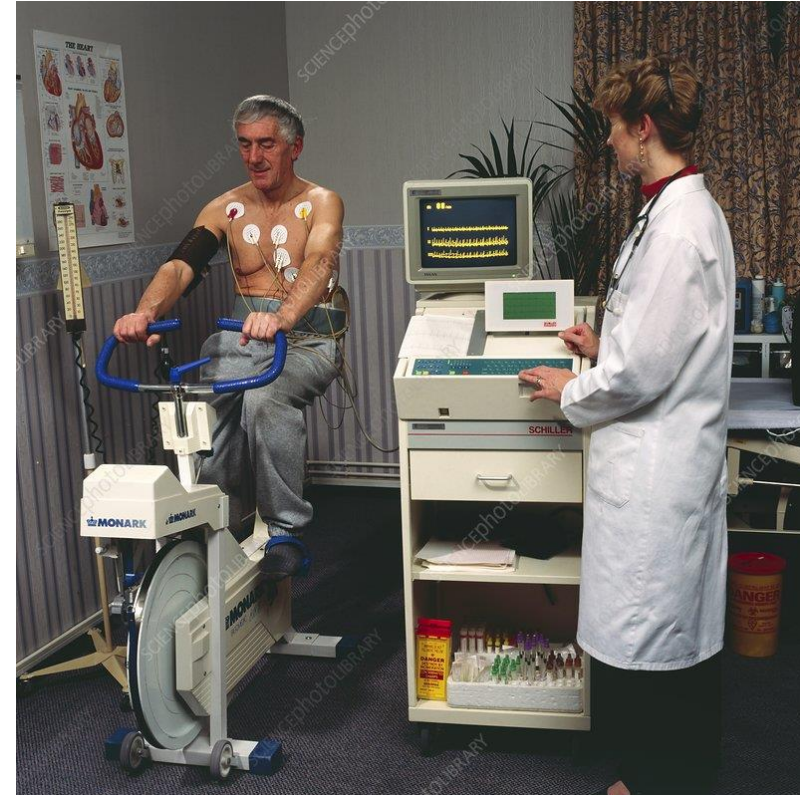
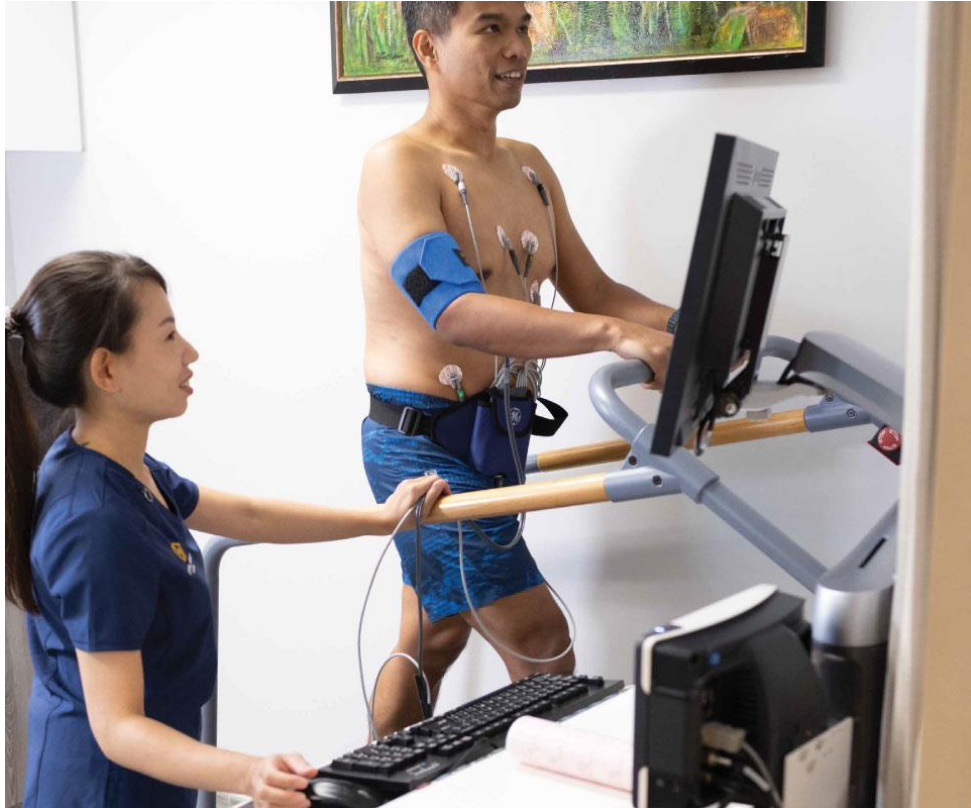
ST Segment depression



- Observe changes in the ST segment during an exercise stress test

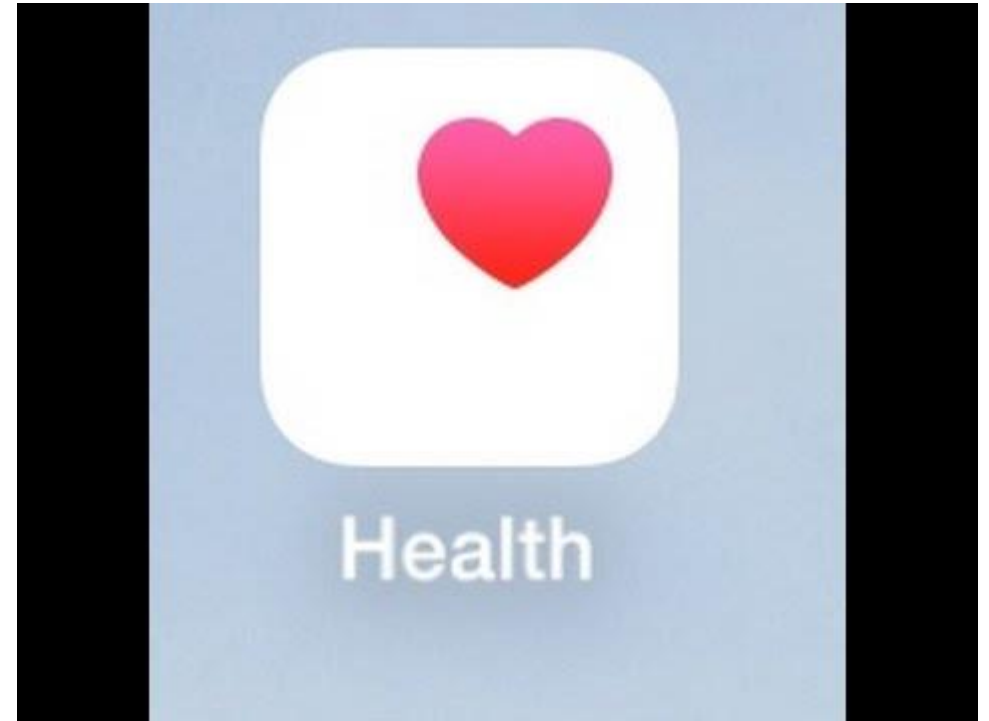
ST Slope	Heart Disease
Up	19.70%
Down	77.80%
Flat	82.80%

Hurdle



- 15 minutes of treadmill or bicycle ergometer

Suggestion



- Self diagnosis with easy ECG and App

Summary

- The ST Slope variable ranks first in both chi-square tests and machine learning feature importance evaluations.
- This means that the ST Slope is the most important predictive variable for heart disease.
- The ST Slope is measured during an exercise stress ECG and is classified as Up, Down, or Flat.
- An ST Slope Up indicates normal health, while Down and Flat suggest signs of heart disease.
- Since the exercise stress ECG is taken during approximately 15 minutes of exercise, a simple check is challenging.
- The development of an easy-to-use wireless ECG testing device is necessary. Additionally, there is a need for the development of an AI-based app that can accurately predict heart disease based on a precise interpretation of ECG and an individual's profile.
- Routine checks using the above device can contribute to the prevention of heart disease.

Thank you

