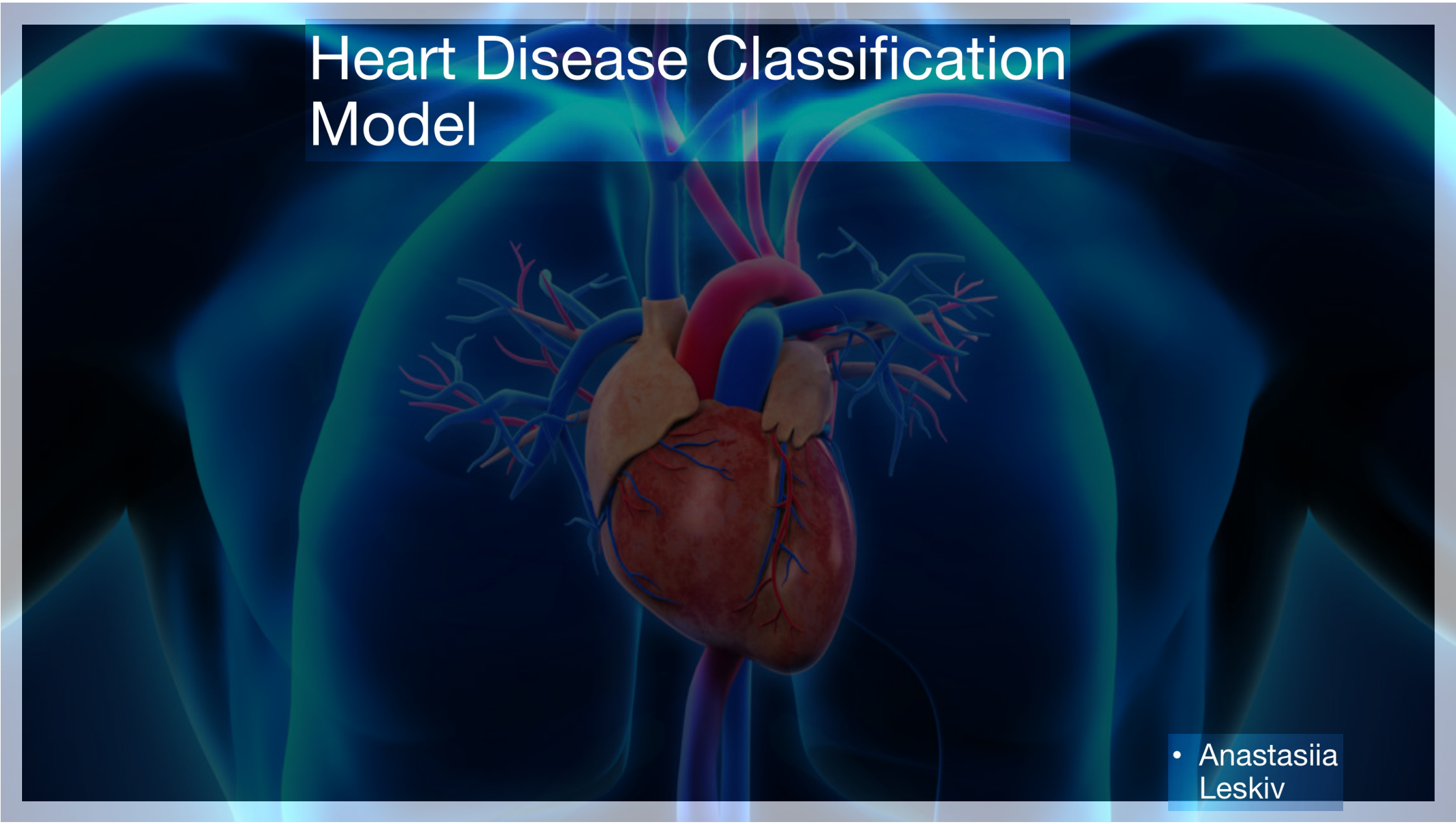
Heart Disease Classification Model



- Anastasiia Leskiv

HEART DISEASE

20%
Percentage of heart attacks that are silent

18.2M
Americans aged 20 and older with coronary heart disease

#1
Number one leading cause of death for men and women in the U.S.

647K
Number of Americans who die from heart disease each year

80%
Percentage of preventable cases of heart disease and stroke

healthcentral

Target:

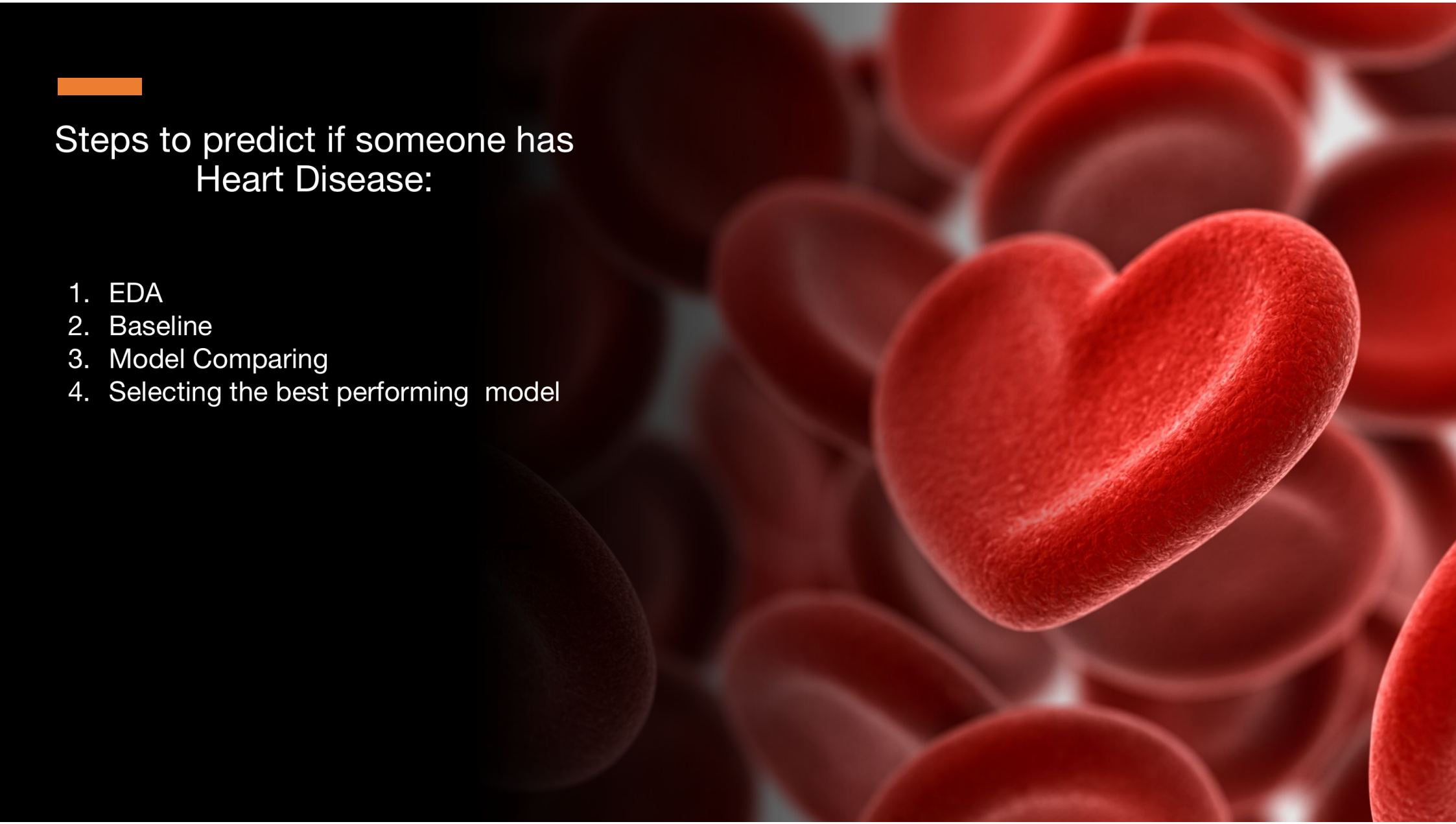
Build classification models to see whether or not a patient has heart disease.

Features

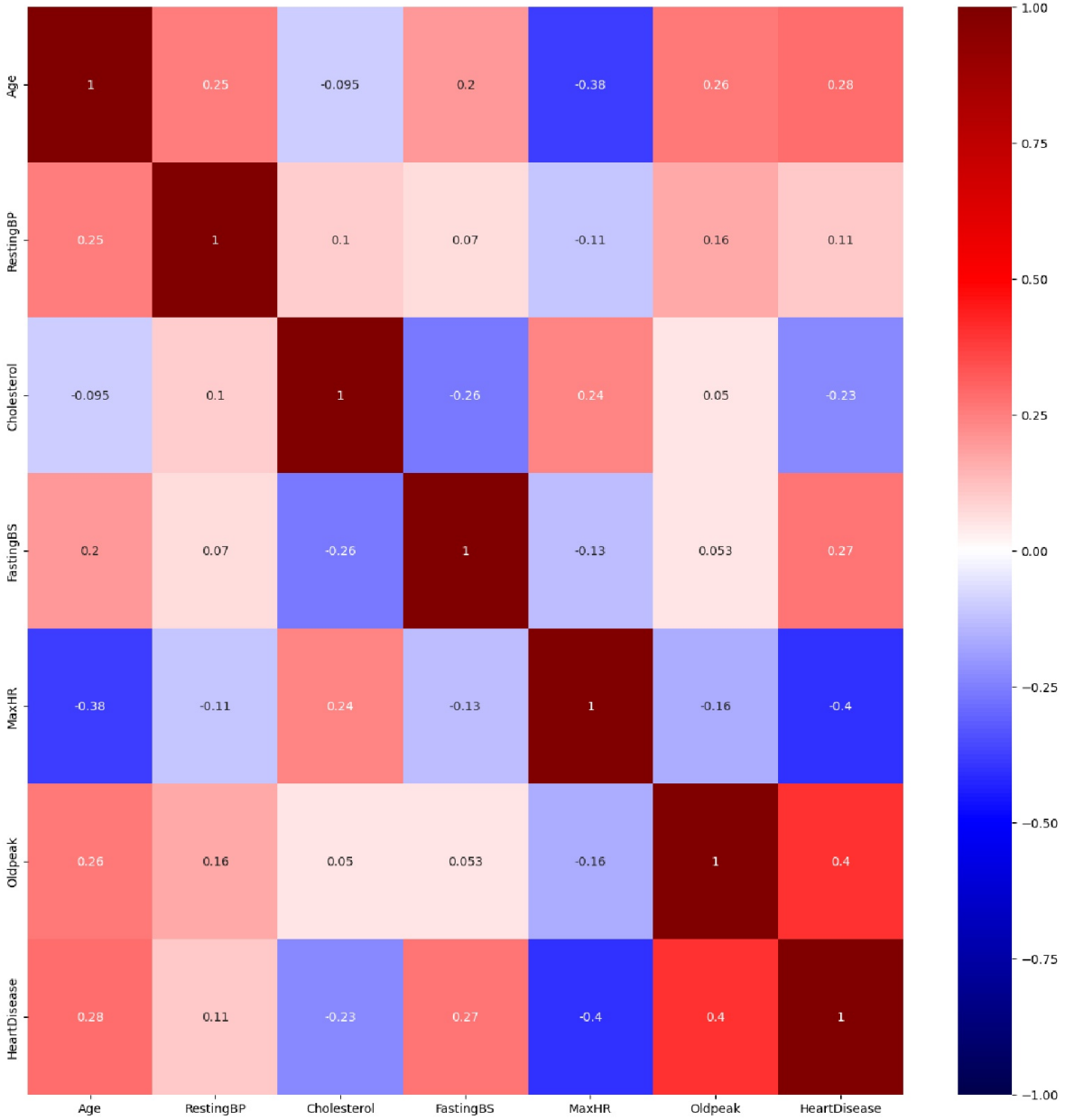
- Age
- Sex
- Resting Blood Pressure
- Fasting Blood sugar
- Resting ECG
- Cholesterol
- Chest Pain Type
- Max Heart Rate
- Exercise Angina
- Oldpeak
- ST_Slope
- HeartDisease

Steps to predict if someone has Heart Disease:

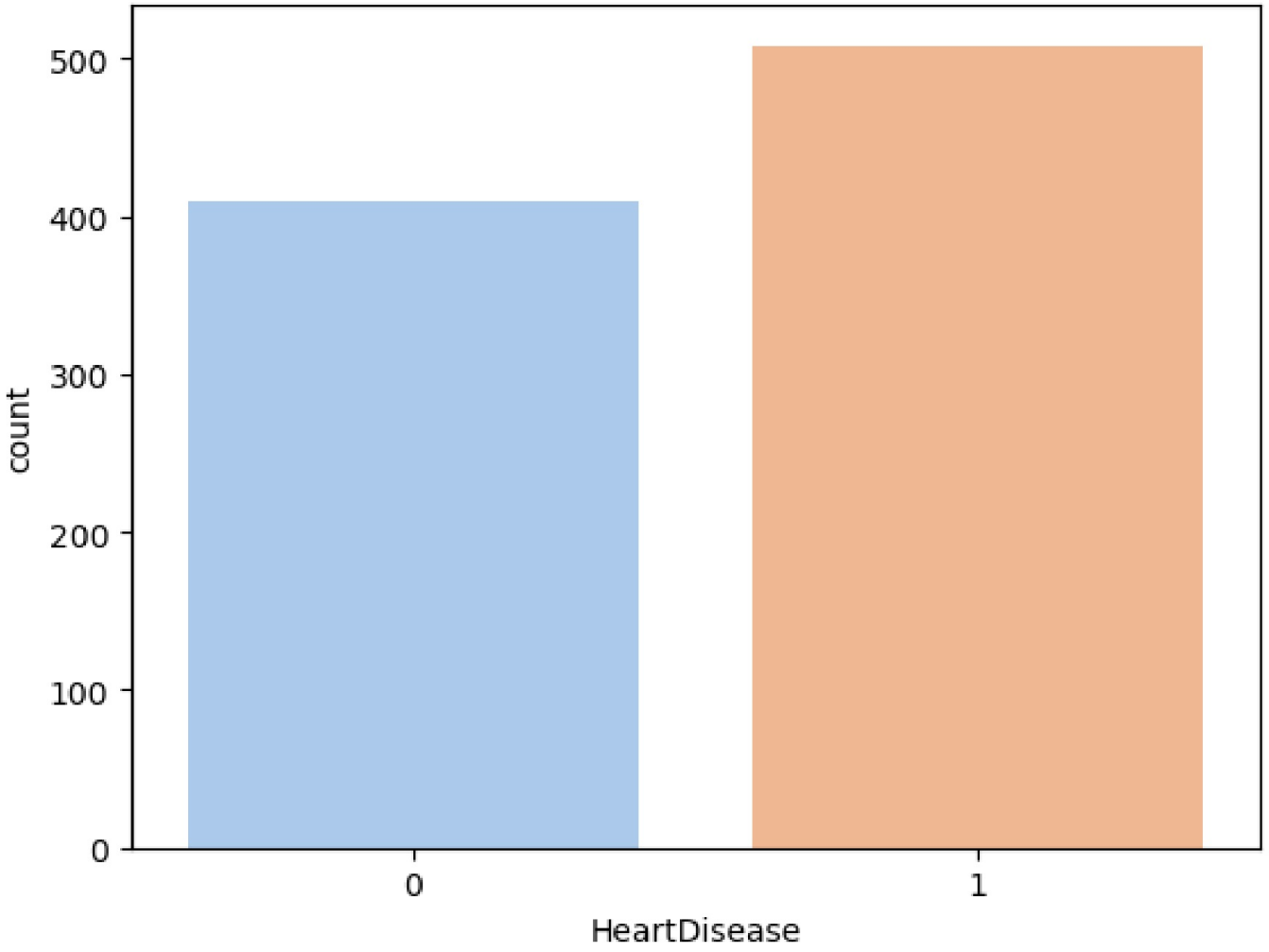
1. EDA
2. Baseline
3. Model Comparing
4. Selecting the best performing model



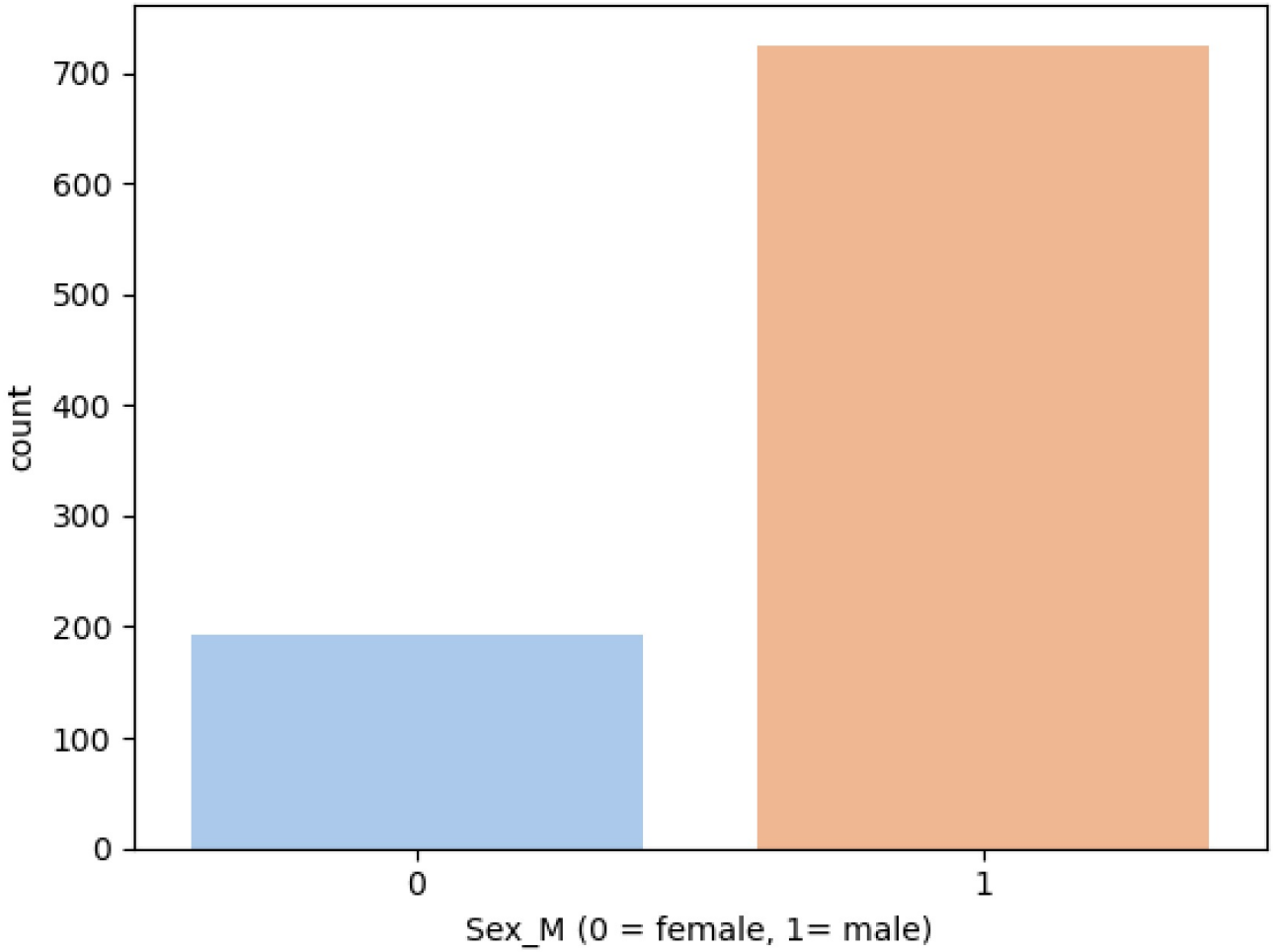
Correlation between the variables the larger the number the higher the correlation.

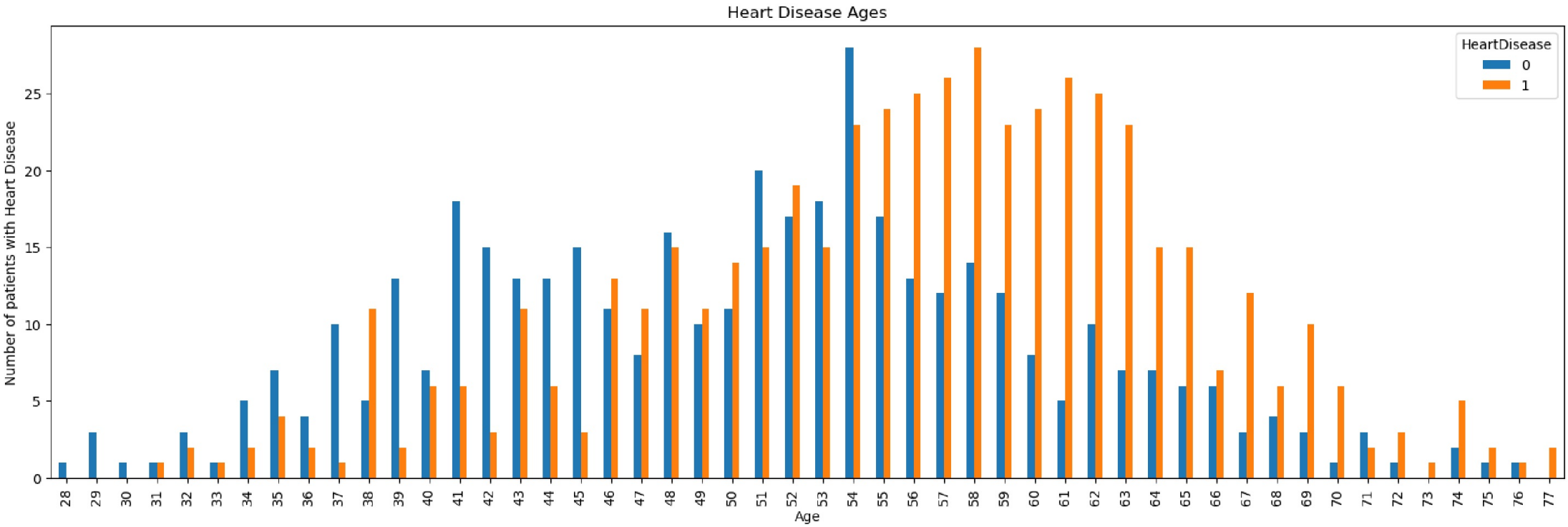


Patients with Heart Disease (1) VS Patients without Heart Disease (0)

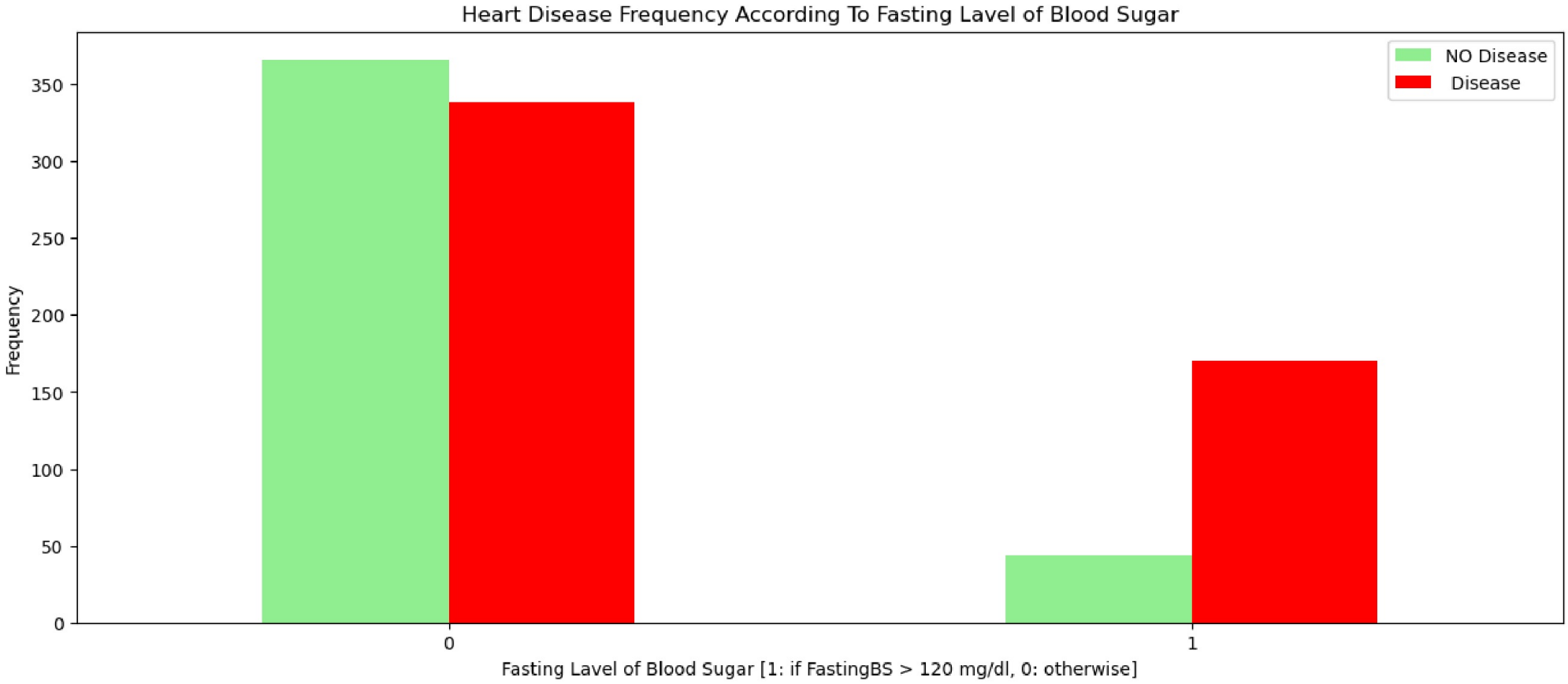


Who has a greater risk of getting a Heart Disease Male VS Female

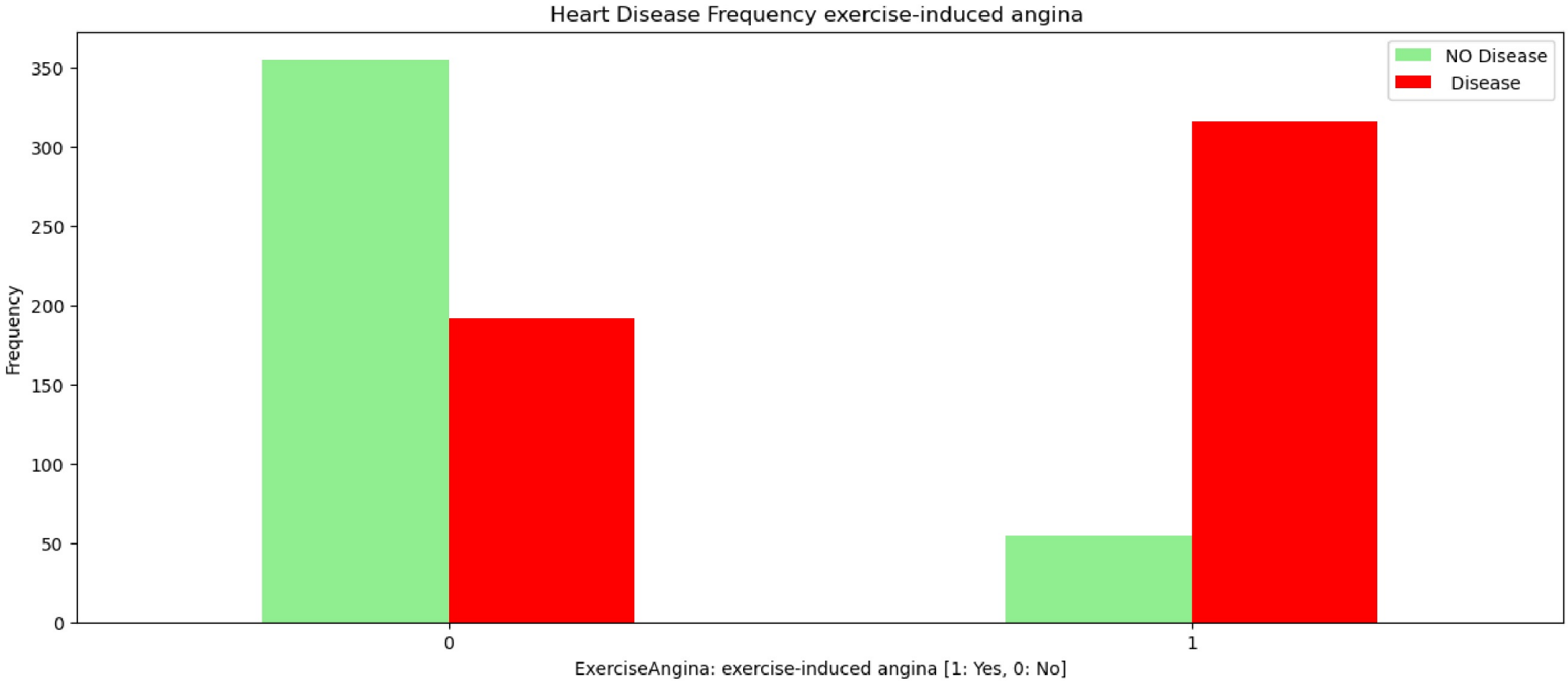




People at the age between the ages of 50-65 have the heist risk of getting a Heart Disease.



People with high levels of sugar have a higher risk of getting a Heart Disease.



People who experienced Angina have a higher risk of getting a Heart Disease.

Hyperparameter Tuning

First Model

Accuracy: 0.6141
Precision: 0.6636
Recall: 0.6822
F1: 0.6728
Cross Validation Accuracy:
Score: 0.6294

Second Model

Accuracy: 0.6033
Precision: 0.6771
Recall: 0.6075
F1: 0.6404
Cross Validation Accuracy:
Score: 0.6280

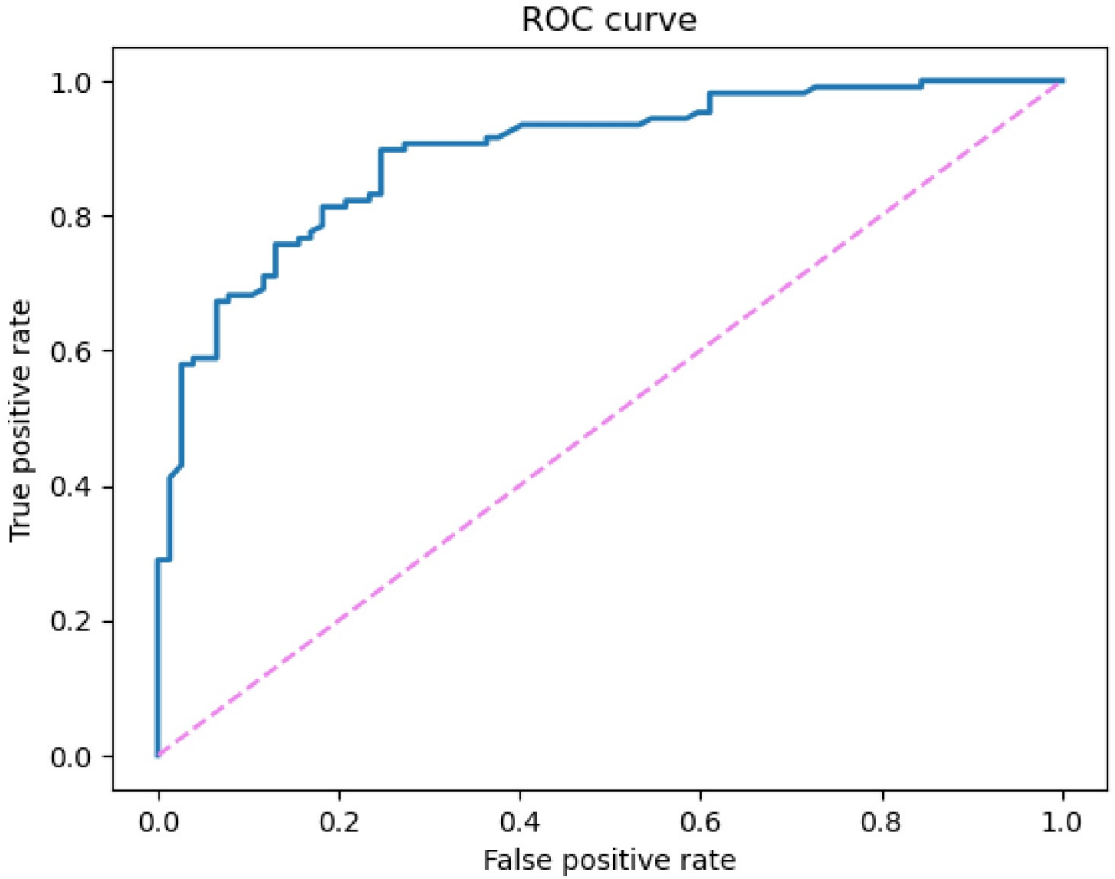
Third Model

Accuracy: 0.6087
Precision: 0.6842
Recall: 0.6075
F1: 0.6436
Cross Validation Accuracy:
Score: 0.6239

Best Model

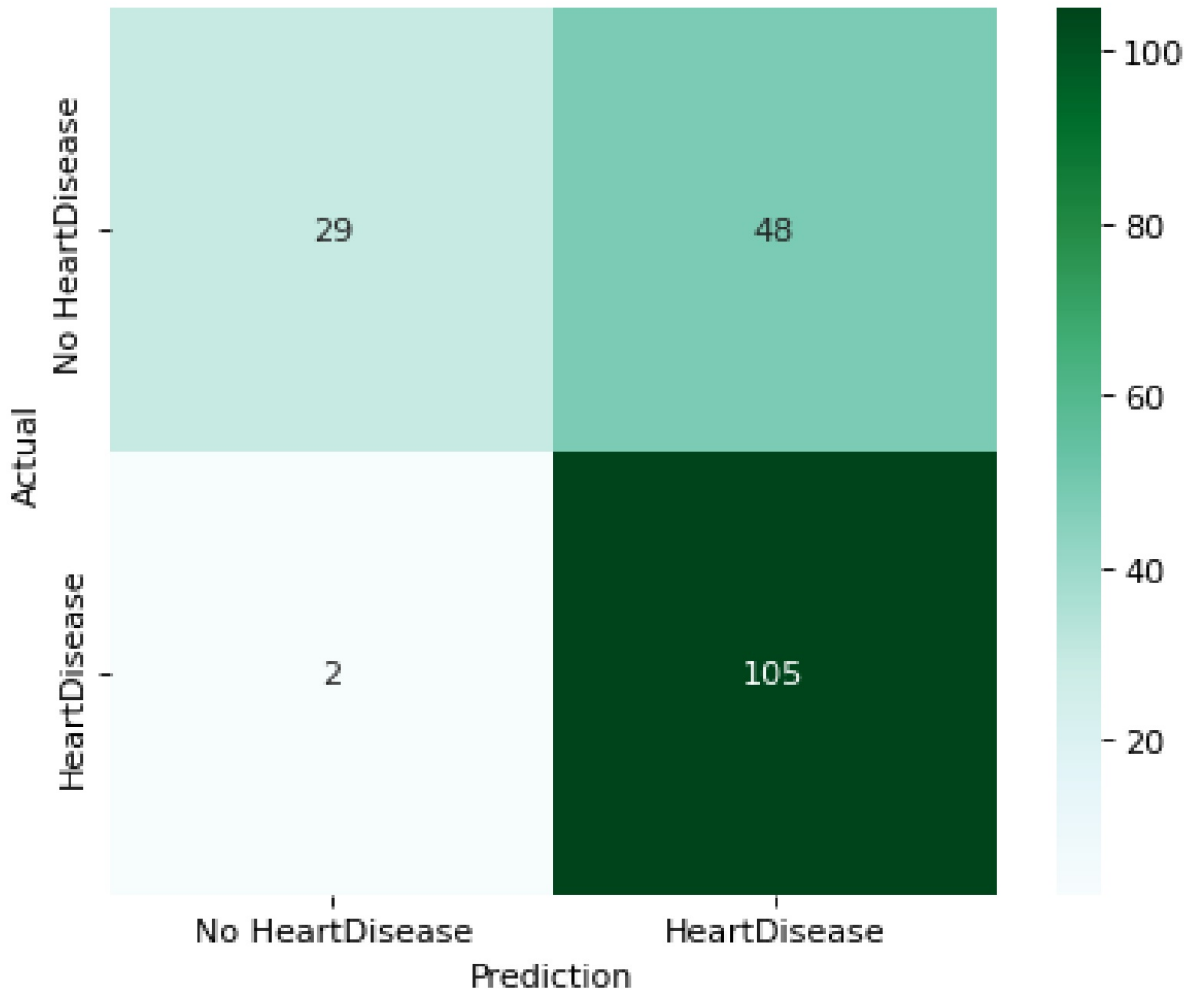
Accuracy: 0.5978
Precision: 0.6051
Recall: 0.8879
F1: 0.7197
Cross Validation Accuracy:
Score: 0.5954

Best model
Random
Forest



Area under the curve **ROC** Score : 0.89
means the classifier can almost perfectly distinguish between all the Positive and the Negative class points.

Best Performance Model Random Forest

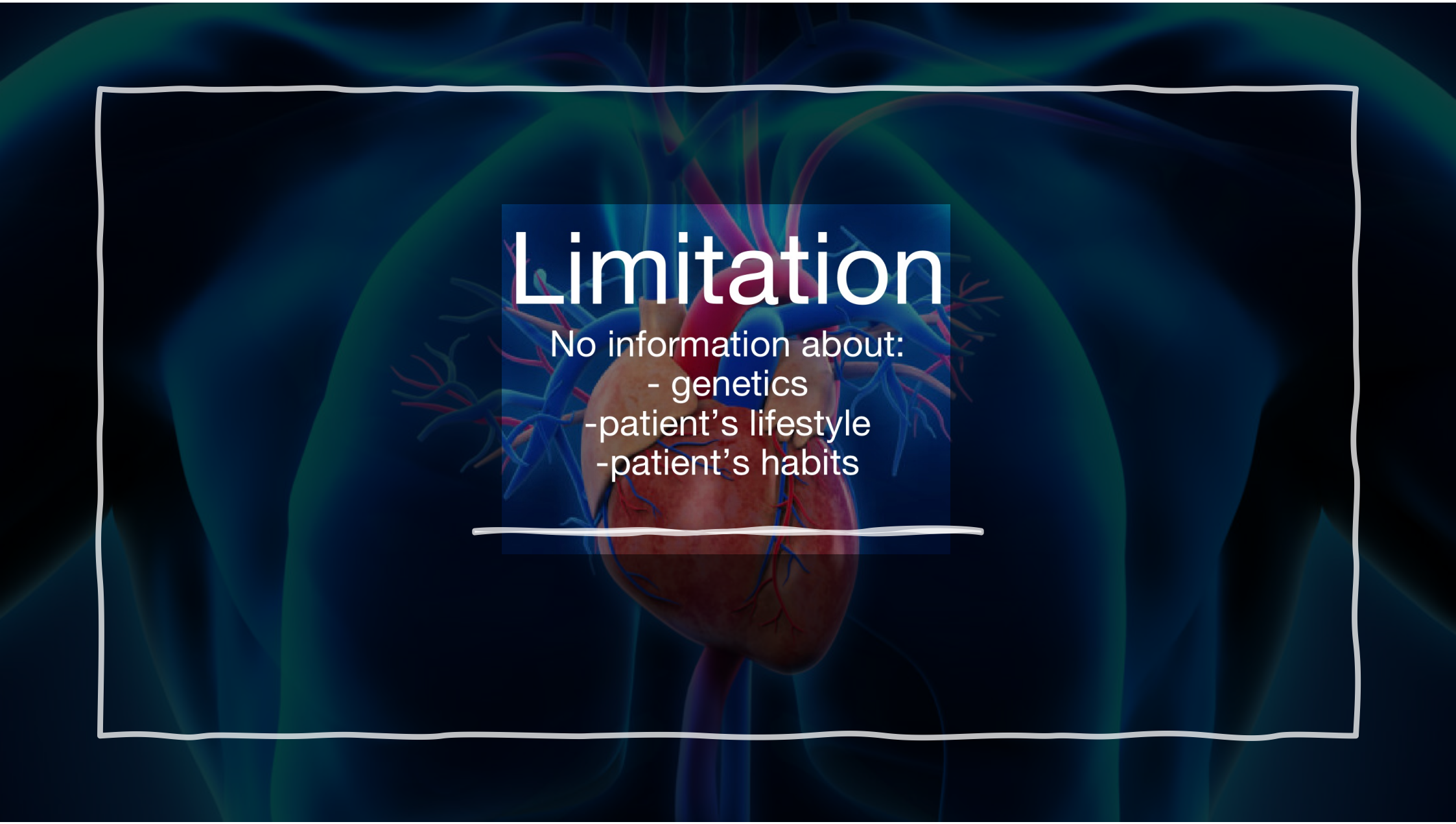


Conclusion

In conclusion, to diagnose if a patient has a Heart Disease, we must pay careful attention to factors that increase the chances of getting heart disease such as:

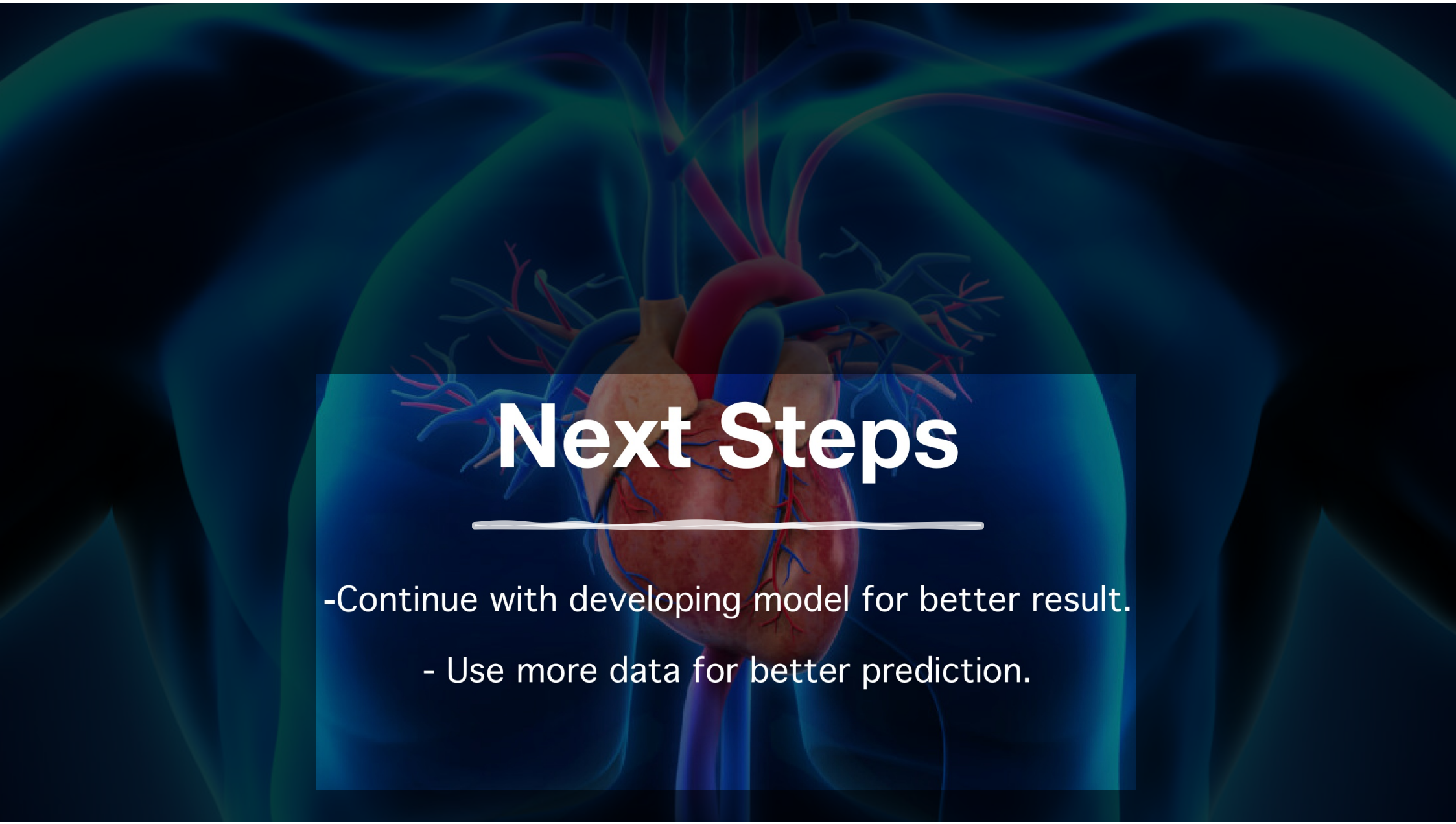
- Age
- Gender
- Sugar level
- Angina
- Blood pressure

My model predicted with
The Area under the ROC curve is 0.89


Limitation
No information about:
- genetics
-patient’s lifestyle
-patient’s habits

Recommendation

I would recommend to my client to pay more attention to:
-Gender
- Sugar level
-Blood pressure
-Age of the patient
-If patient experienced Angina
-.


Next Steps
-Continue with developing model for better result.
- Use more data for better prediction.

Anastasiia Leskiv
GitHub:https://github.com/anastasiialeskiv
Linkedin:https://www.linkedin.com/in/anastasiia-leskiv-