A

Major Project Presentation

on

Enhancing Heart Failure Prediction using Ensemble Learning Technique

in

Partial Fulfillment of Final Year Computer Engineering Course-Seminar and Technical Communication

(A. Y. 2024-25)

Department of Computer Engineering (Regional Language)

PCET's Pimpri Chinchwad College of Engineering

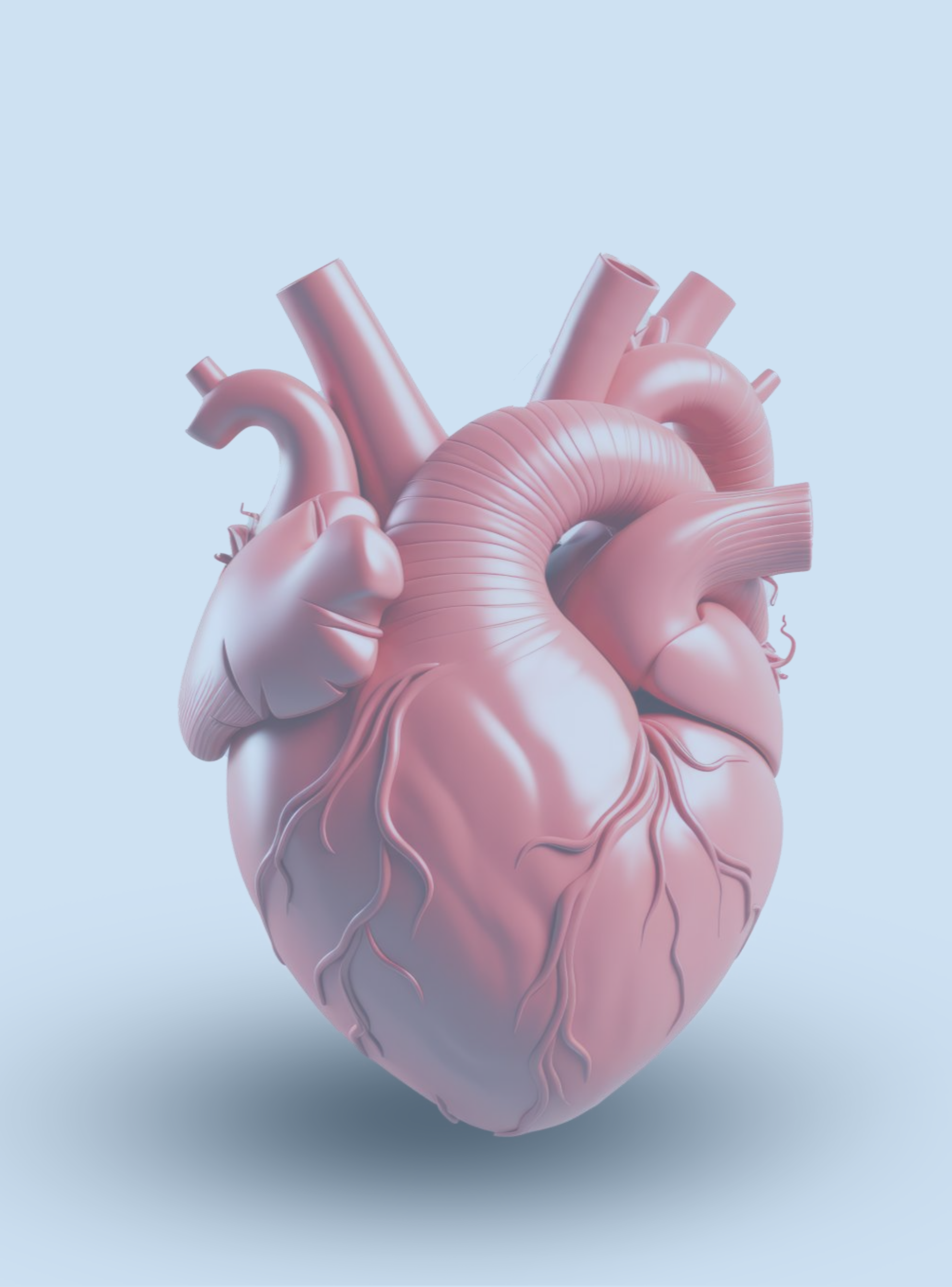
Enhancing Heart Failure Prediction

using Ensemble Learning Technique

Team Members

Abhijit Motekar : 121B1D051
Gayatri Joshi : 121B1D047
Aditya Jaiswal : 121B1D013
Sanket Deshmukh : 121B1D002

Guided By Prof. Rohini Sarode



Contents

1. Introduction
2. Motivation
3. Objectives
4. Literature Survey
5. Algorithm Used
6. Project Requierments
7. Dataset (Theory)
8. UML Diagrams
9. Software Testing
10. Implementation
11. Results
12. Contribution to Sustainable Development Goals
13. Future Scope
14. Conclusion
15. Details of Paper Publication
16. References

Introduction

- Heart disease is one of the leading causes of death worldwide.
- Traditional methods for diagnosis are already in place but are often slow and inaccurate.
- This project uses ensemble machine learning techniques to improve prediction accuracy.
- Additionally, a chatbot interface is integrated to make the system more interactive, user-friendly, and accessible.

Motivation

- **Our advanced approach for doctors**: Traditional heart disease prediction methods can miss crucial signs, delaying treatment for those in need.

- **Ensemble learning to the rescue**: We combine multiple machine learning models to create a more powerful and accurate prediction system.

- **More accurate predictions, faster action**: With improved accuracy, doctors can detect heart disease earlier and start treatment sooner, leading to better patient outcomes.

Objectives

- To enhance the heart failure prediction through ensemble learning model.
- To implement and optimize ensemble learning algorithms, bagging ,stacking, AdaBoost utilizing appropriate hyperparameter tuning techniques such as grid search to maximize predictive accuracy.
- To compare the predictive performance of bagging against traditional models such as random forest, decision tree, SVM ,XGBOOST through statistical analysis.
- To validate the model’s performance using various metrics, such as accuracy, precision, recall, F1-score
- To integrate an AI-powered chatbot

Literature survey

SR.NO	TITLE	YEAR	METHODOLOGY	STRENGTHS	WEAKNESS	DATASET
1	Cardiovascular Disease Detection using Ensemble Learning [1]	2022	Employed ensemble learning techniques (RF, KNN, DT, XGB) on Cardiovascular Disease dataset from Kaggle. Utilized feature extraction and various ML and DL classifiers.	Successfully utilized ensemble learning for cardiovascular disease prediction. Effective use of feature extraction and multiple classifiers.	Limited discussion on the scalability and generalizability of the proposed approach. Lack of detailed analysis on the impact of individual classifiers within the ensemble.	Cardiovascular Disease dataset (2019)
2	Early heart disease prediction using ensemble learning techniques [2]	2022	Proposed ensemble classifier model for heart disease prediction. Utilized sampling strategies and feature selection methods.	Achieved high detection rates (e.g., 99% with random oversampling). Successful application of ensemble learning on unbalanced datasets.	Limited evaluation on only two datasets. Future work needed to assess performance on diverse datasets and real-world applications.	Heart Disease (2020)
3	Enhancing Heart Disease Prediction Through Ensemble Learning and Feature Selection [3]	2023	Applied data preprocessing, feature selection, data balancing, and ensemble methods (stacking, majority voting, bagging) on UCI dataset.	Improved prediction accuracy using ensemble methods and feature selection techniques. Successful application of data balancing and ensemble learning.	Limited discussion on the impact of individual ensemble methods on model performance. Lack of external validation on additional datasets.	Heart Disease (2020)
4	Heart disease risk prediction using deep learning techniques with feature augmentation [4]	2023	Combined sparse autoencoder and convolutional classifier for feature augmentation and deep learning. Employed ensemble learning techniques on a dataset from five independent centers.	Outperformed other state-of-the-art methods with a 4.4% improvement in precision. Successful integration of deep learning and ensemble learning for heart disease prediction.	Computationally expensive compared to some techniques like stacking. Limited dataset size may restrict generalization to larger populations.	Heart Failure Prediction Dataset (2021)
5	Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning [5]	2021	Used a combination of machine learning and deep learning on Public Health Dataset. Applied preprocessing techniques (e.g., outlier detection, data normalization).	Achieved promising results with improved accuracy using deep learning and feature preprocessing. Identified importance of dataset normalization and preprocessing for better model performance.	Limited discussion on the scalability and generalizability of the proposed approach. Lack of external validation on additional datasets	Heart Disease (2020)

Algorithms Used

Individual Models

1. Random Forest : constructs multiple decision trees and combines their predictions
2. Decision Tree : A tree-like model where decisions are made at each node based
3. SVM : supervised learning algorithm for classification and regression tasks
4. XGBoost : Uses decision trees as base learners

Ensemble Technique

1. Bagging : A technique that builds multiple models using subsets of the training data and combines their predictions

Project Requirements

S/W Requirements:

- Code Editor: VS Code, Google Colab.
- Languages & Libraries: Python (ML models), HTML, CSS, JavaScript.
- Frameworks & Tools: Flask (for backend), Scikit-learn, Pandas, NumPy

H/W Requirements:

- Opreating System : Above(4 GB RAM , i3 Processor)
- Browser : Chrome, Edge, Brave etc.

Functional Requirements:

- Data preprocessing and feature selection.
- Ensemble learning implementation (Bagging, Stacking, AdaBoost) .
- AI-powered chatbot for health recommendations.

Non-Functional Requirements:

- User-friendly web interface
- Fast and optimized model inference

Dataset

Features:

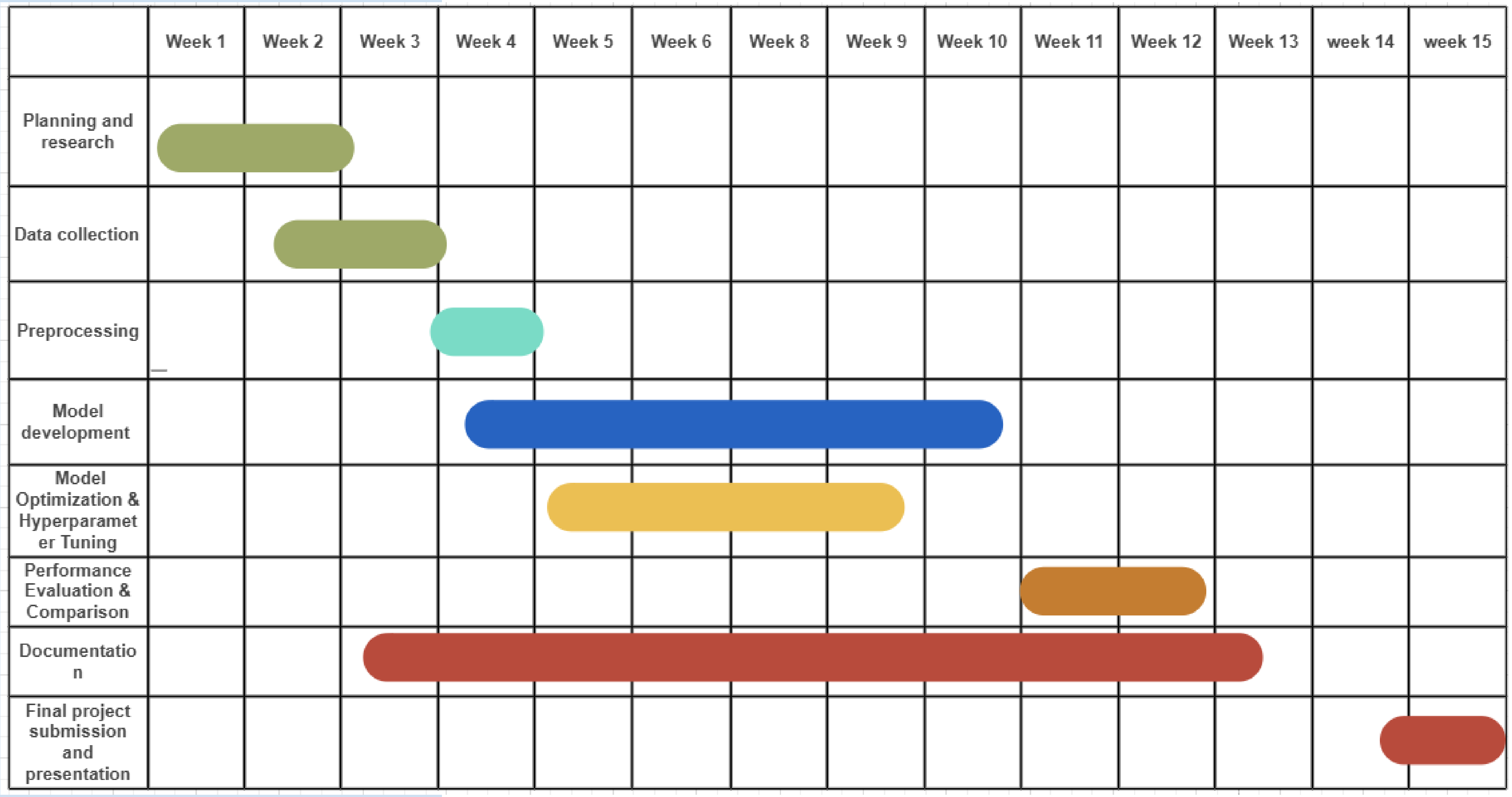
1. Age
2. Sex
3. Chest Pain Type
4. Resting Blood Pressure (RestingBP)
5. Cholesterol
6. Fasting Blood Sugar (FastingBS)
7. Resting Electrocardiographic Results (RestingECG)
8. Maximum Heart Rate Achieved (MaxHR)
9. Exercise-Induced Angina (ExerciseAngina)
10. ST Depression Induced by Exercise Relative to Rest (Oldpeak)
11. Slope of the Peak Exercise ST Segment (ST_Slope)
12. Target Variable: Heart Disease

Total Instances: 920
Total Features: 12
[LINK]

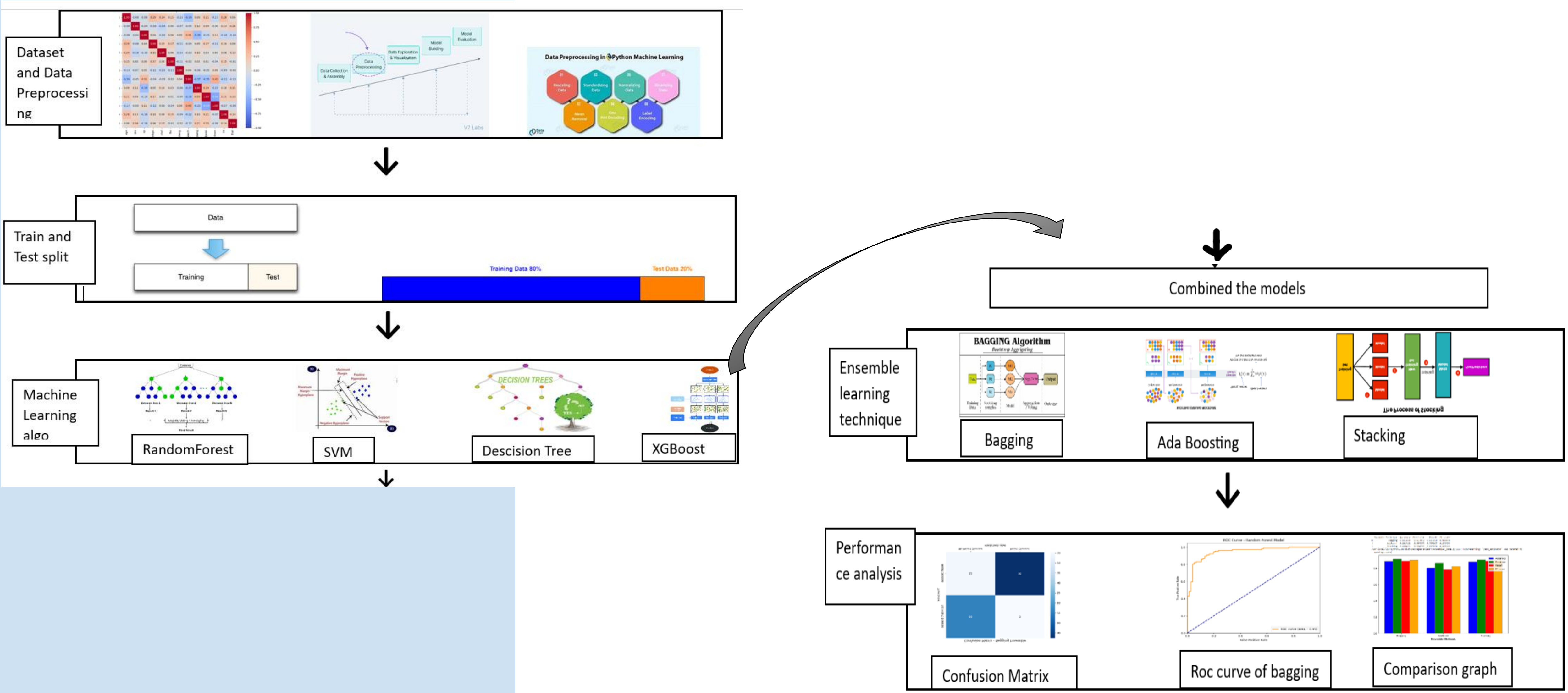
Dataset Design

Feature Name	Data Type	Description	Possible values
Age	Integer	Patient’s age in years	30,45,60…
Sex	categorical	Gender of Patient	M, F..
ChestPain	Categorical	Type of chest pain experienced by the patient·	0: Typical Angina 1: Atypical Angina 2: Non-Anginal Pain 3: Asymptomatic
RestingBP	Integer	Resting blood pressure (in mm Hg).	140,170…
Cholesterol	Integer	Cholesterol level (mg/dL)	180,200,150…
FastingBS	Boolean	Fasting blood sugar level·	1: ≥ 120 mg/dL (High) 0: < 120 mg/dL (Normal).
RestingECG	Categorical	Resting electrocardiogram results·	0: Normal 1: ST-T wave abnormality 2: Left ventricular hypertrophy
MaxHR	Integer	Max heart rate during stress test	120,150..
ExerciseAngina	Boolean	Chest pain during exercise	N,Y
Oldpeak	Float	Depression induced by exercise relative to rest	0,1.5,1…
ST_Slope	categorical	Slope of peak exercise ST segment	UP, Flat
Heart Disease (target)	Boolean	prediction target (whether patient has heart disease)	0 or 1

Project Planning : Timeline Diagram



System Architecture



Software Testing

1. Unit Testing

Testing of individual units (e.g., prediction algorithm, chatbot response module).

Verifies that every function behaves as desired (input → output validation).

Example: Verification of the prediction model with example patient data.

2. Integration Testing

Verifies how various modules (Prediction Model + Chatbot + UI) integrate together. Example: After prediction, user is able to switch to Chatbot without fail.

Software Testing

3. Functional Testing

   Verifies all the features (input form, prediction, chatbot conversation).

   Example: Does the system accurately forecast heart failure risk from inputs?

4. Performance Testing

   Verifies system responsiveness and speed under load.

   Example: How quickly does the model forecast when several users access it?

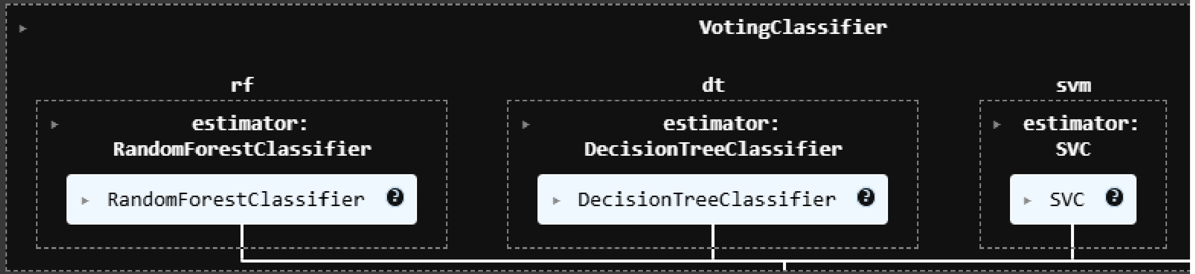
5. Usability Testing

   Verifies the interface is easy to use for non-technical users (doctors/patients).

   Example: Is navigation intuitive? Are buttons labeled clearly?

Implementation

```
# Training the Bagging ensemble
bagging_classifier.fit(X_train_prep, y_train)
```




```
# Predictions on training set
y_train_pred_bagging = bagging_classifier.predict(X_train_prep)
train_accuracy_bagging = accuracy_score(y_train, y_train_pred_bagging)
print("Bagging Ensemble - Train Accuracy:", train_accuracy_bagging)

# Predictions on testing set
y_test_pred_bagging = bagging_classifier.predict(X_test_prep)
test_accuracy_bagging = accuracy_score(y_test, y_test_pred_bagging)
print("Bagging Ensemble - Test Accuracy:", test_accuracy_bagging)


# from google.colab import files
# import joblib
# joblib.dump(bagging_classifier, 'trained_model.pkl')
# files.download('trained_model.pkl')
```

```
Bagging Ensemble - Train Accuracy: 0.9182561307901907
Bagging Ensemble - Test Accuracy: 0.8858695652173914
```

Bagging Ensemble (Random Forest, Decision Tree, SVM, and XGBoost)

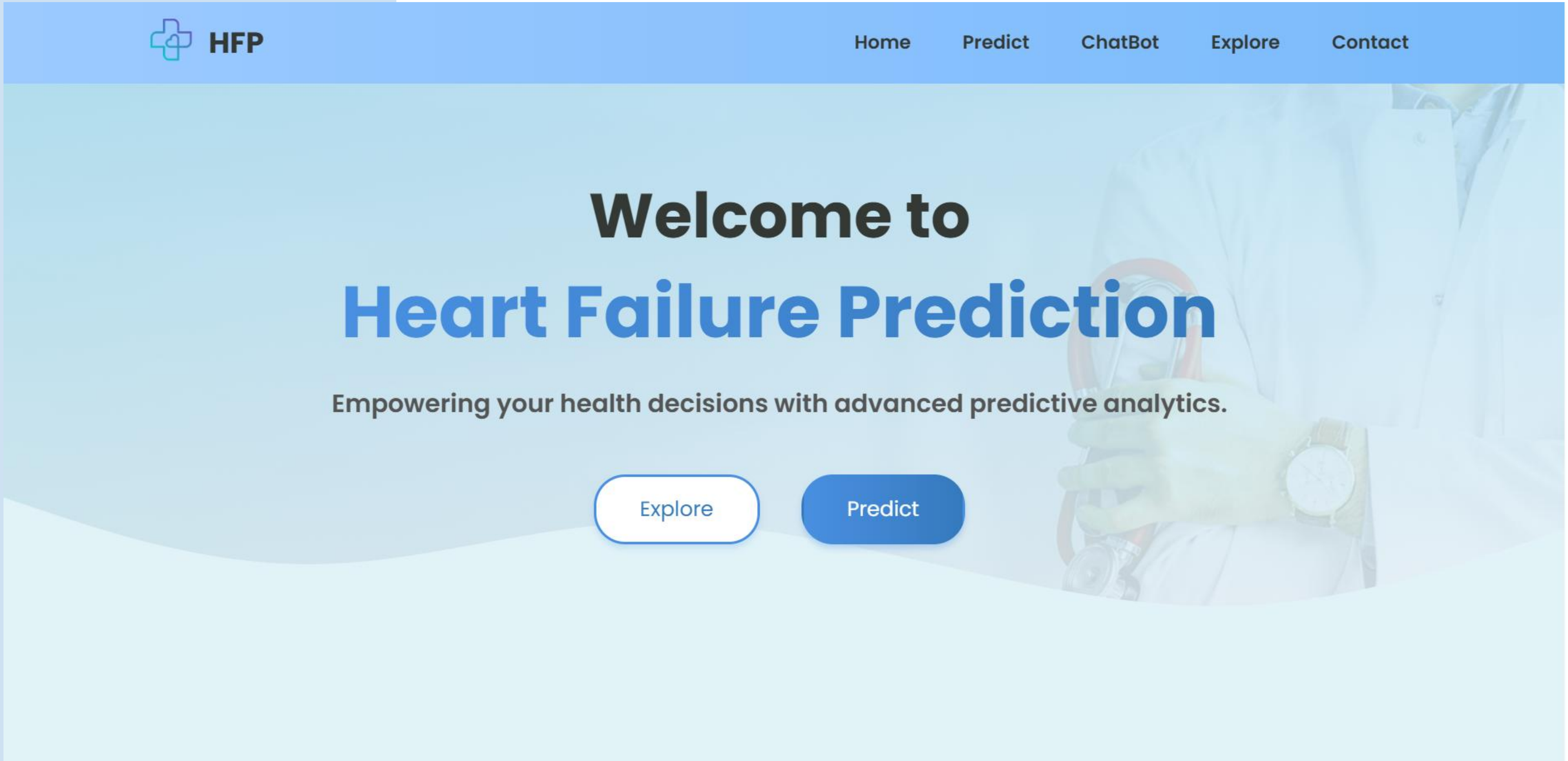
```
from sklearn.ensemble import BaggingClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC
from xgboost import XGBClassifier
```

```
# Initialize individual models
rf_classifier = RandomForestClassifier(**rf_grid_search.best_params_, random_state=42)
dt_classifier = DecisionTreeClassifier(**dt_grid_search.best_params_, random_state=42)
svm_classifier = SVC(**svm_grid_search.best_params_, random_state=42)
xgb_classifier = XGBClassifier(**xgb_grid_search.best_params_, random_state=42)
```

```
# Initialize Bagging ensemble for each individual model
bagging_rf = BaggingClassifier(estimator=rf_classifier, n_estimators=10, random_state=42)
bagging_dt = BaggingClassifier(estimator=dt_classifier, n_estimators=10, random_state=42)
bagging_svm = BaggingClassifier(estimator=svm_classifier, n_estimators=10, random_state=42)
bagging_xgb = BaggingClassifier(estimator=xgb_classifier, n_estimators=10, random_state=42)
```

```
# Combine Bagging ensembles of individual models
bagging_classifier = VotingClassifier(estimators=[
    ('rf', bagging_rf),
    ('dt', bagging_dt),
    ('svm', bagging_svm),
    ('xgb', bagging_xgb)
])
```

Result (Frontend Design)



Result (Frontend Design)

About and Analytics

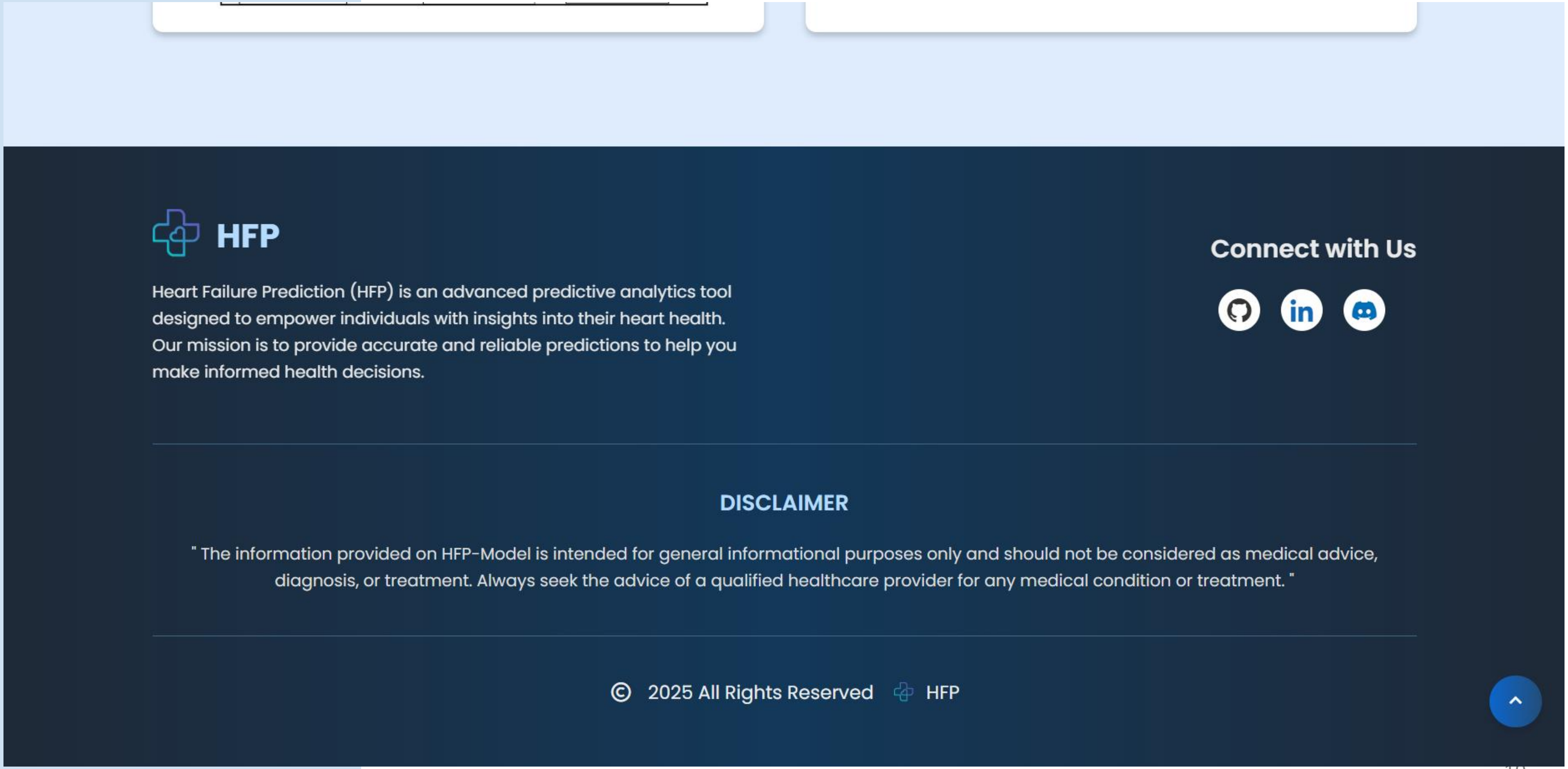
‹ Dataset Information and Design ›

The Heart Failure Prediction model is built using a dataset from Kaggle (2021). This dataset contains 920 instances and 12 features, including age, sex, chest pain type, and more. The target variable is the presence of heart disease.

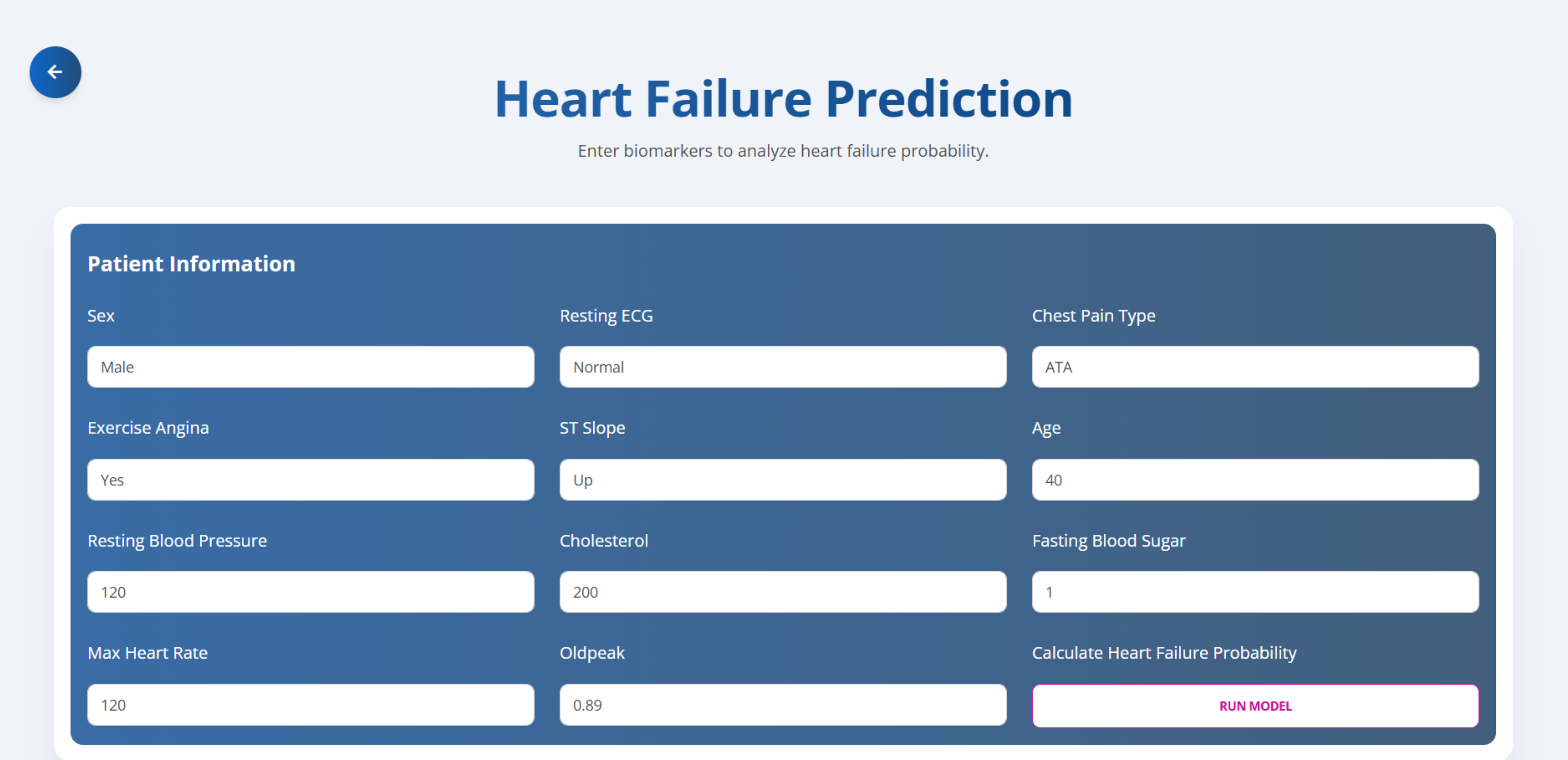
1. **Age:** Age of the patient.
2. **Sex:** Gender of the patient.
3. **Chest Pain Type:** Type of chest pain experienced.
4. **Resting Blood Pressure:** Blood pressure at rest.
5. **Cholesterol:** Serum cholesterol level.
6. **Fasting Blood Sugar:** Blood sugar level after fasting.
7. **Resting ECG:** Resting electrocardiographic results.
8. **Max Heart Rate:** Maximum heart rate achieved.
9. **Exercise Angina:** Presence of exercise-induced angina.

Feature Name	Data Type	Description	Possible values
Age	Integer	Patient's age in years	30,45,60...
Sex	categorical	Gender of Patient	M, F..
ChestPain	Categorical	Type of chest pain experienced by the patient:	0: Typical Angina 1: Atypical Angina 2: Non-Anginal Pain 3: Asymptomatic
RestingBP	Integer	Resting blood pressure (in mm Hg).	140,170...
Cholesterol	Integer	Cholesterol level (mg/dL)	180,200,150...
FastingBS	Boolean	Fasting blood sugar level:	1: ≥ 120 mg/dL (High) 0: < 120 mg/dL (Normal).
RestingECG	Categorical	Resting electrocardiogram results:	0: Normal 1: ST-T wave abnormality 2: Left ventricular hypertrophy
MaxHR	Integer	Max heart rate during stress test	120,150..
ExerciseAngina	Boolean	Chest pain during exercise	N,Y
Oldpeak	Float	Depression induced by exercise relative to rest	0,1.5,1...

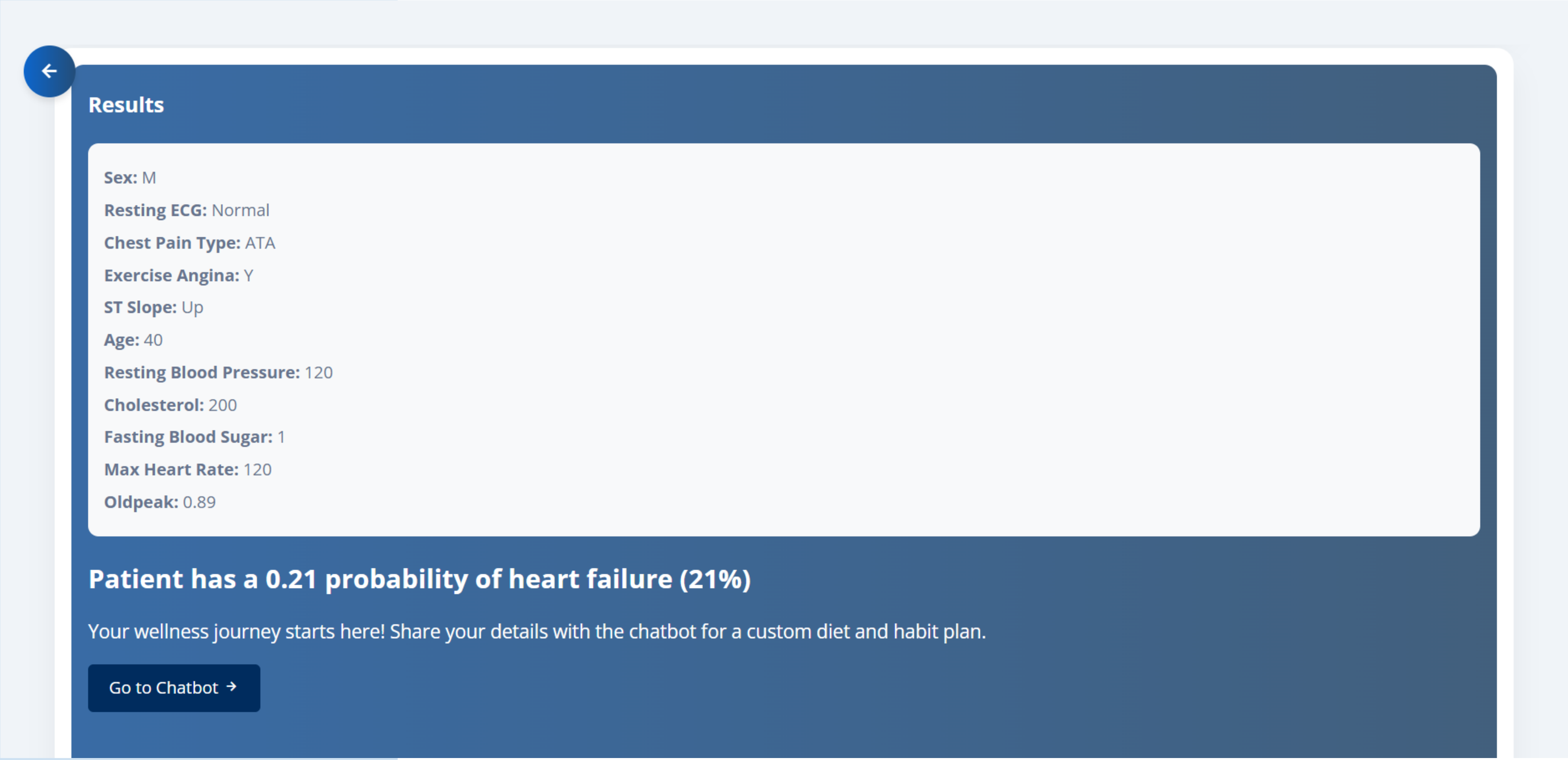
Result (Frontend Design)



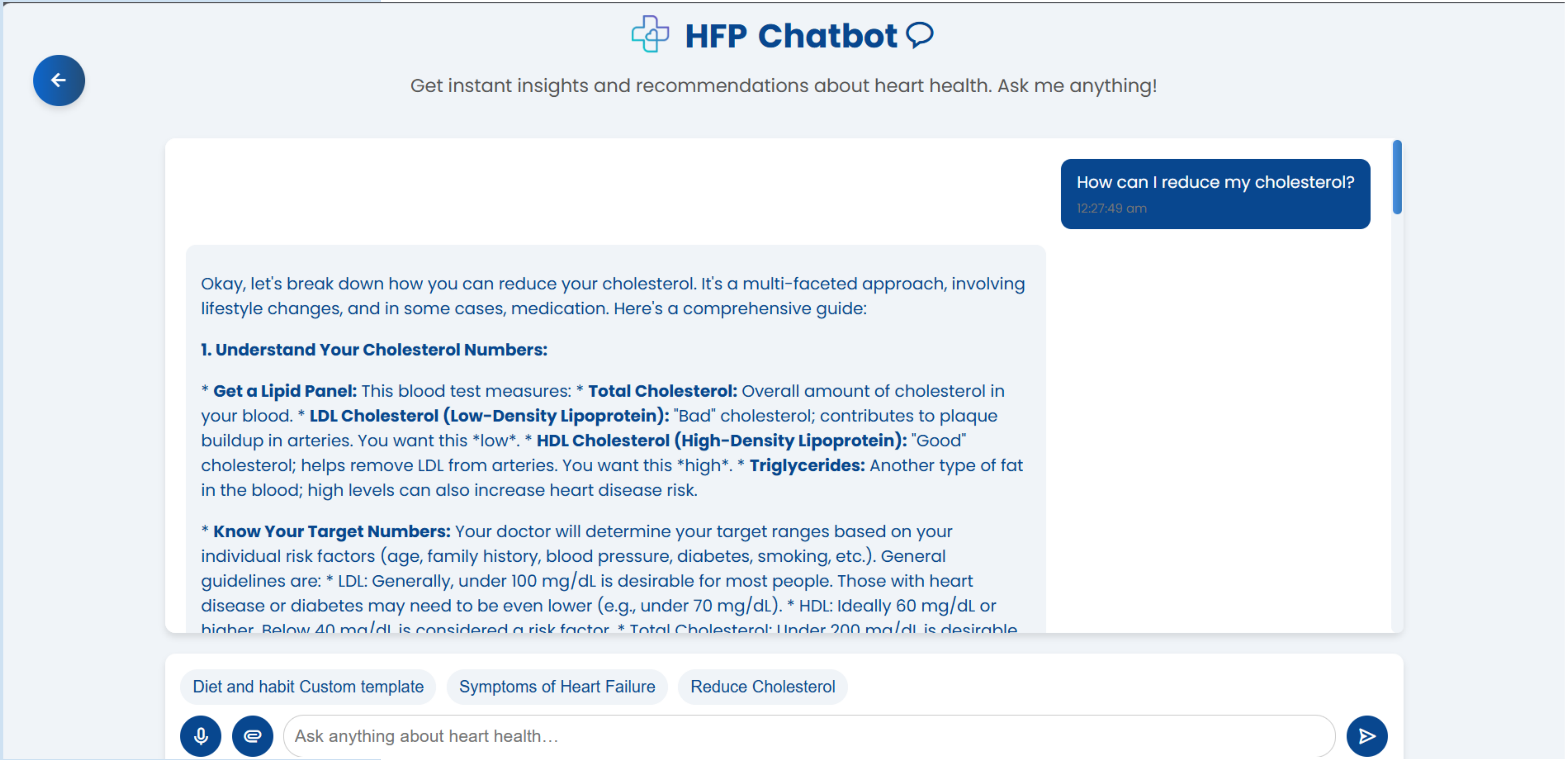
Result (Frontend Design)



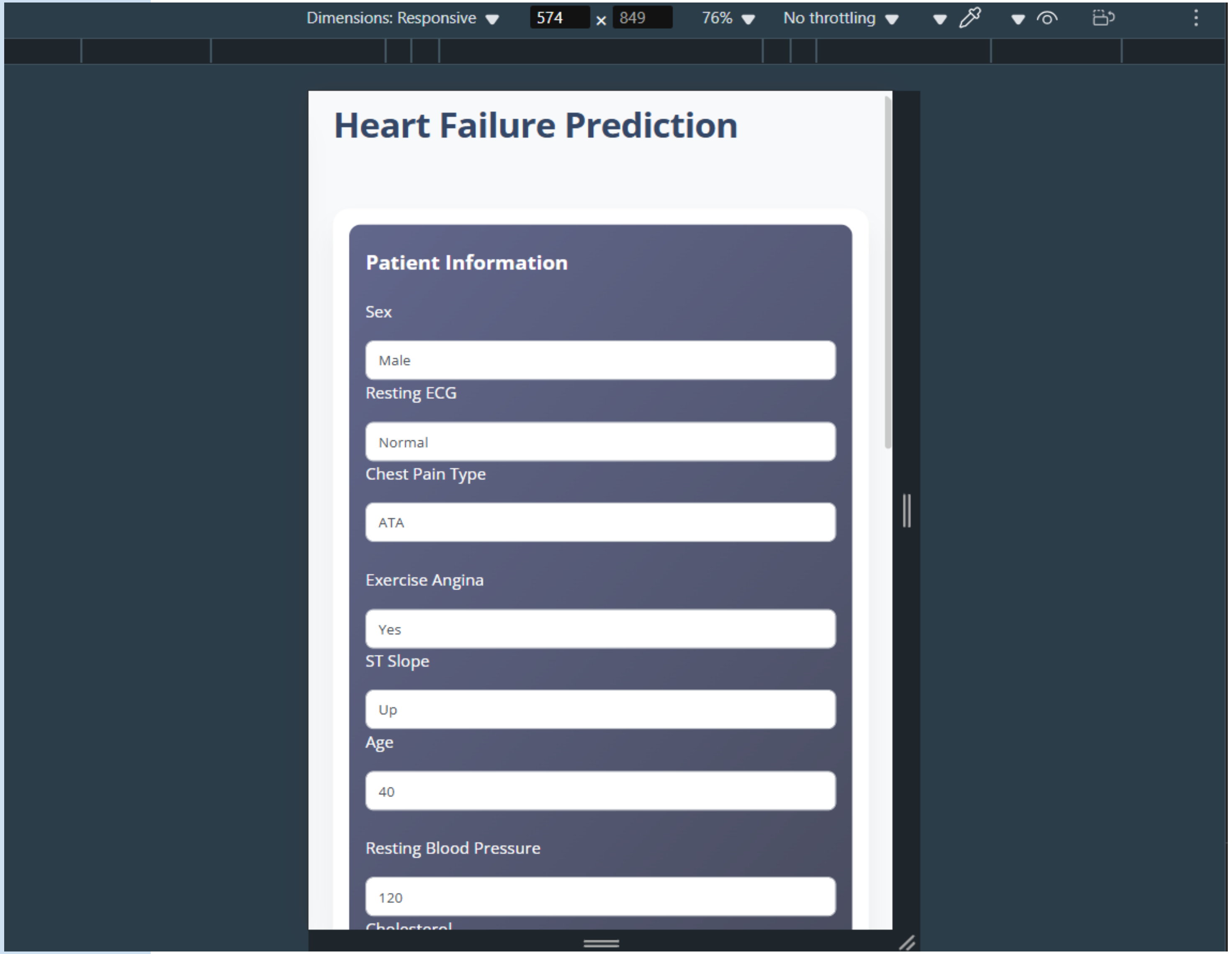
Result (Frontend Design)



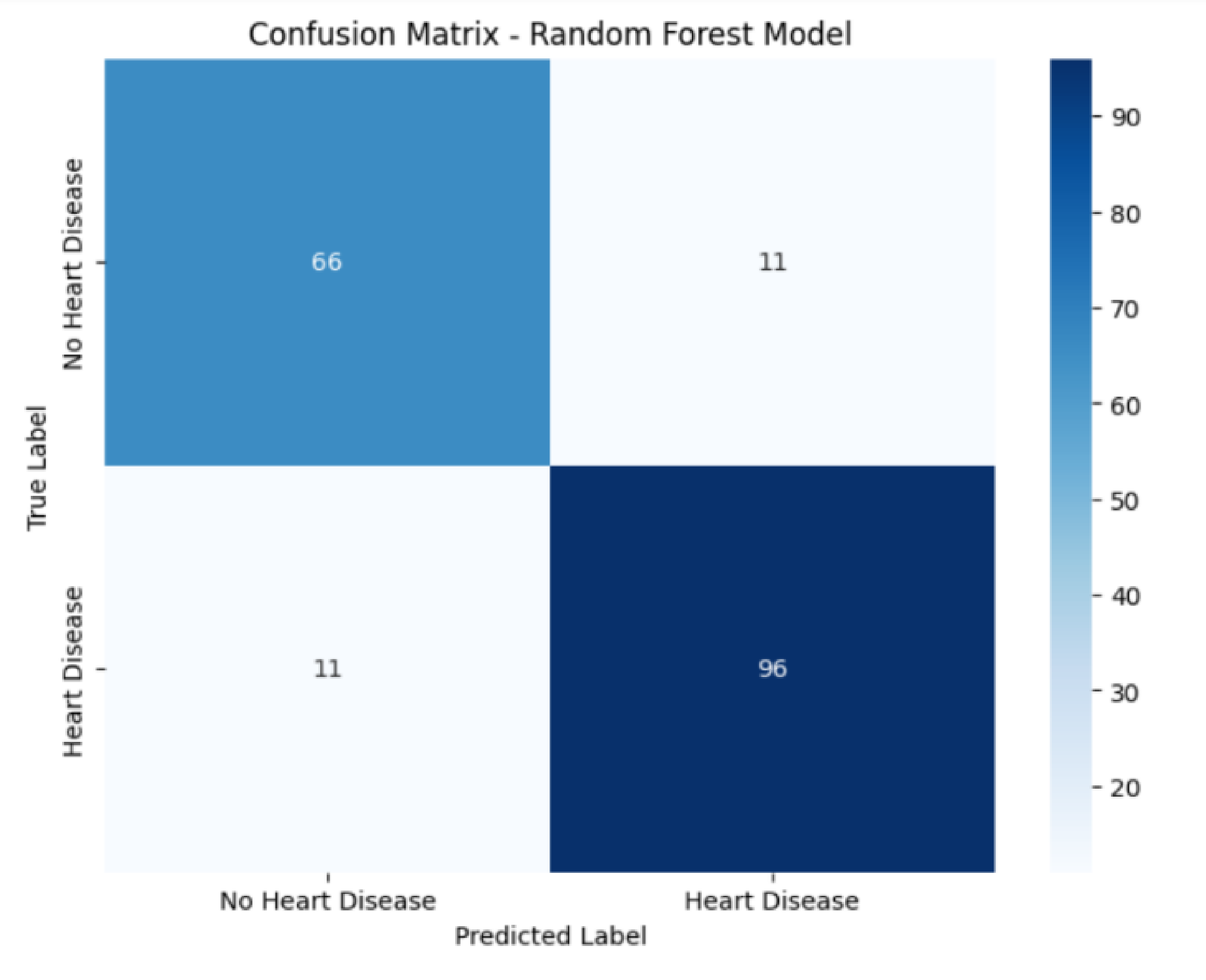
Result (Frontend Design)



Responsive Design



Result (Confusion Matrix)



Random Forest :

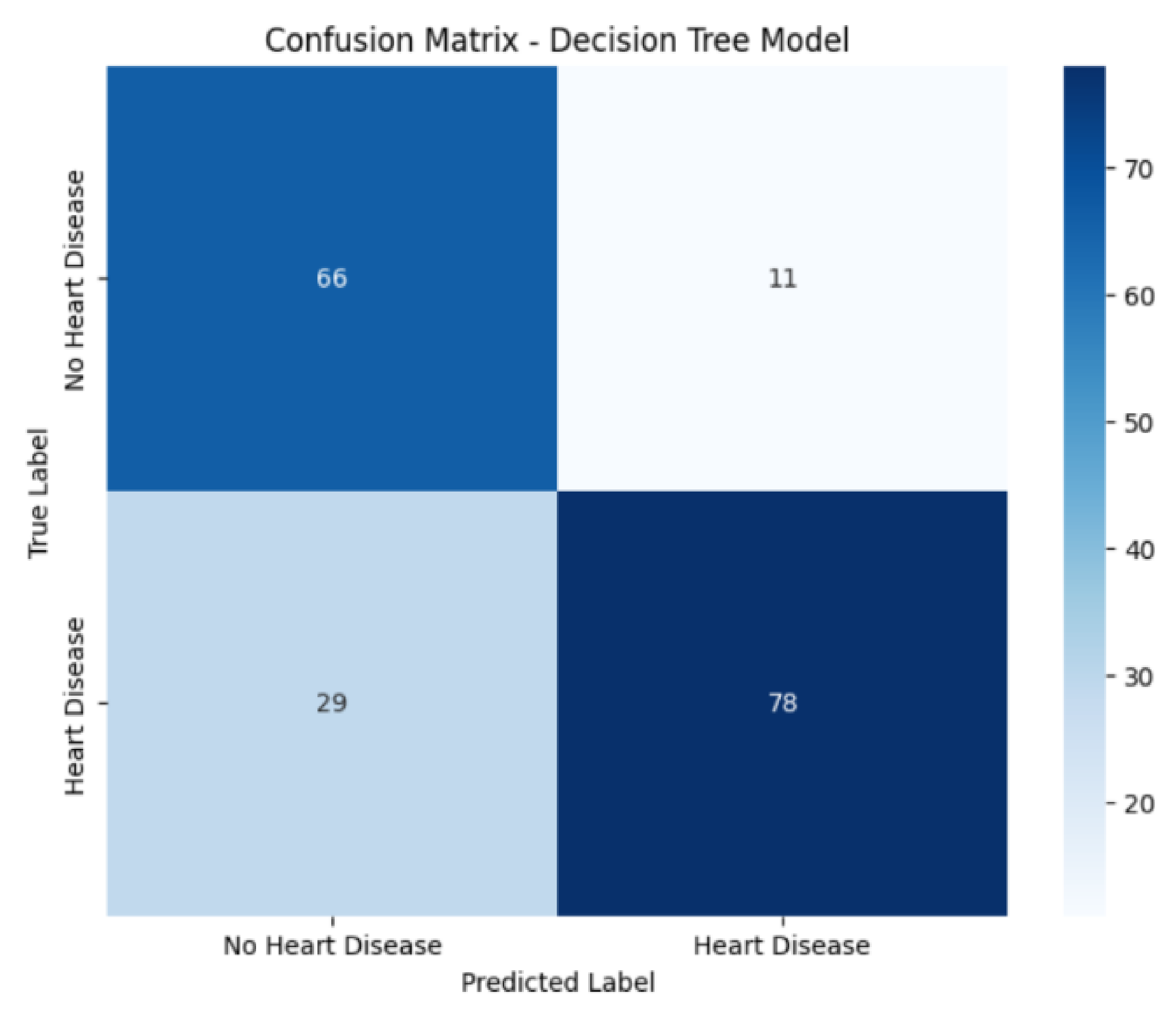
66 (True Negative - TN) 35.87%

11 (False Negative - FN) 5.98%

11 (False Positive - FP) 5.98%

96 (True Positive - TP) 52.17%

Result (Confusion Matrix)



Decision Tree :

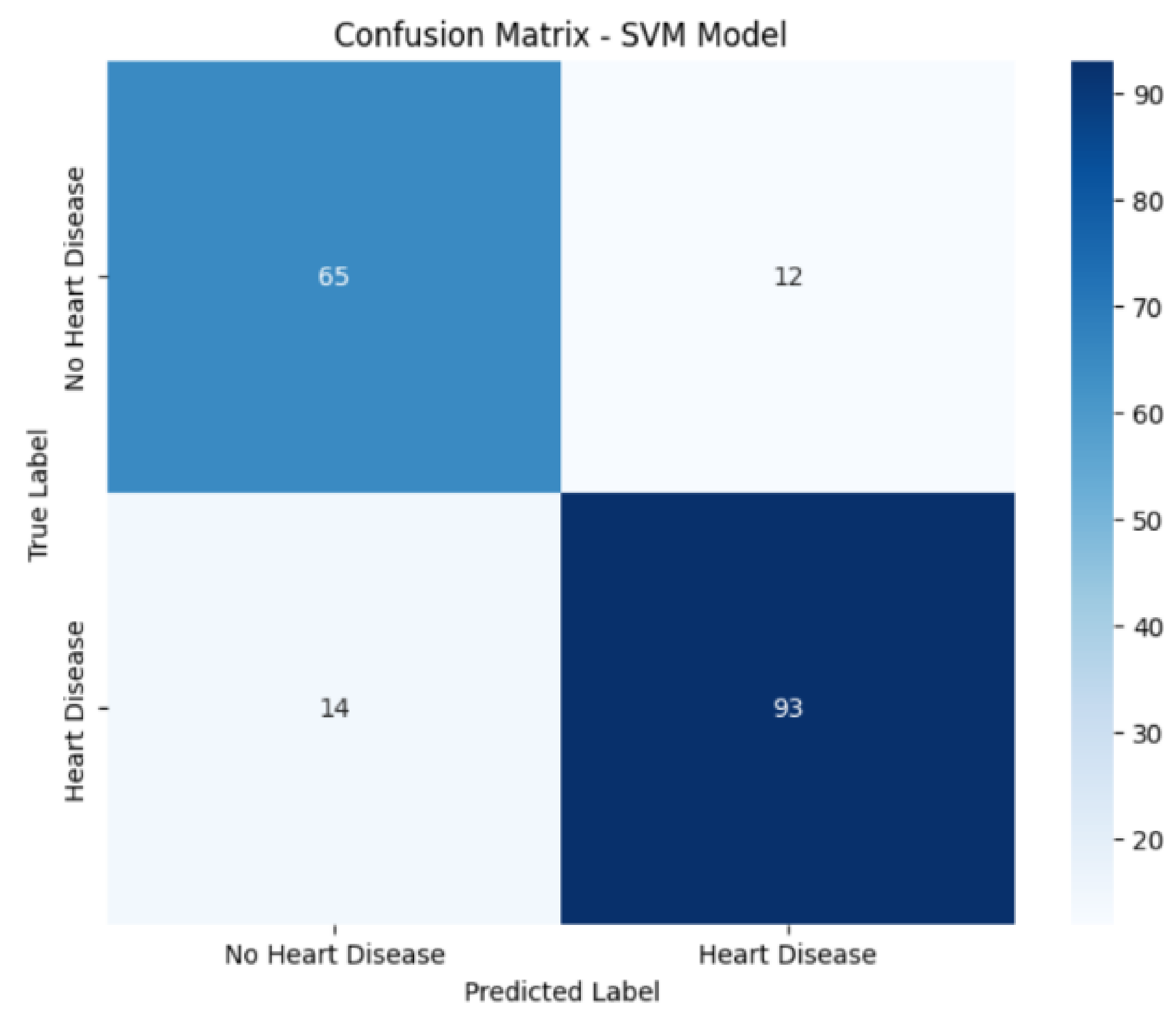
66 (True Negative - TN) 35.87%

29 (False Negative - FN) 15.76%

11 (False Positive - FP) 5.98%

78 (True Positive - TP) 42.39%

Result (Confusion Matrix)



SVM :

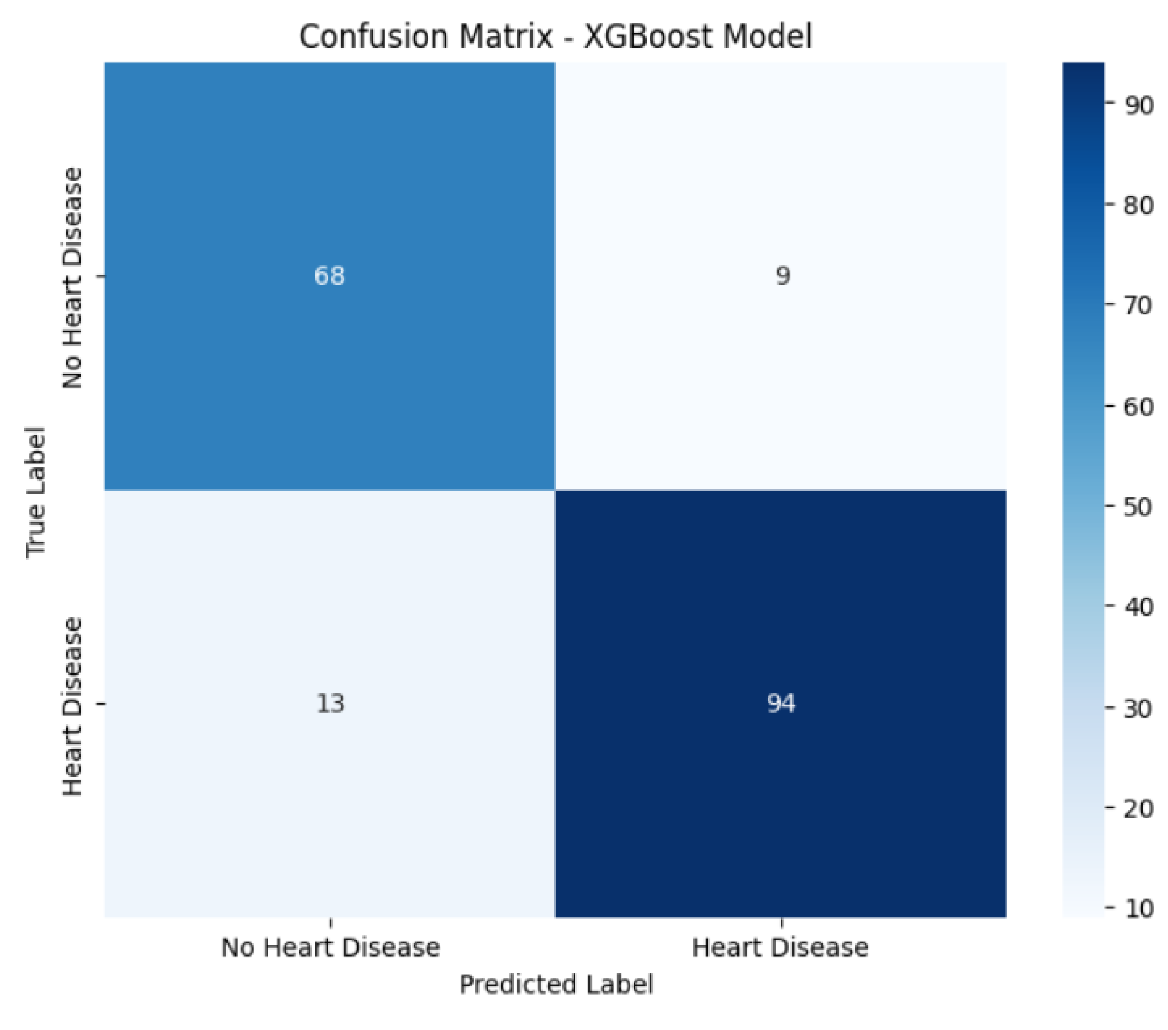
65 (True Negative - TN) 34.39%

14 (False Negative - FN) 7.41%

12 (False Positive - FP) 6.35%

98 (True Positive - TP) 51.85%

Result (Confusion Matrix)



XGBoost :

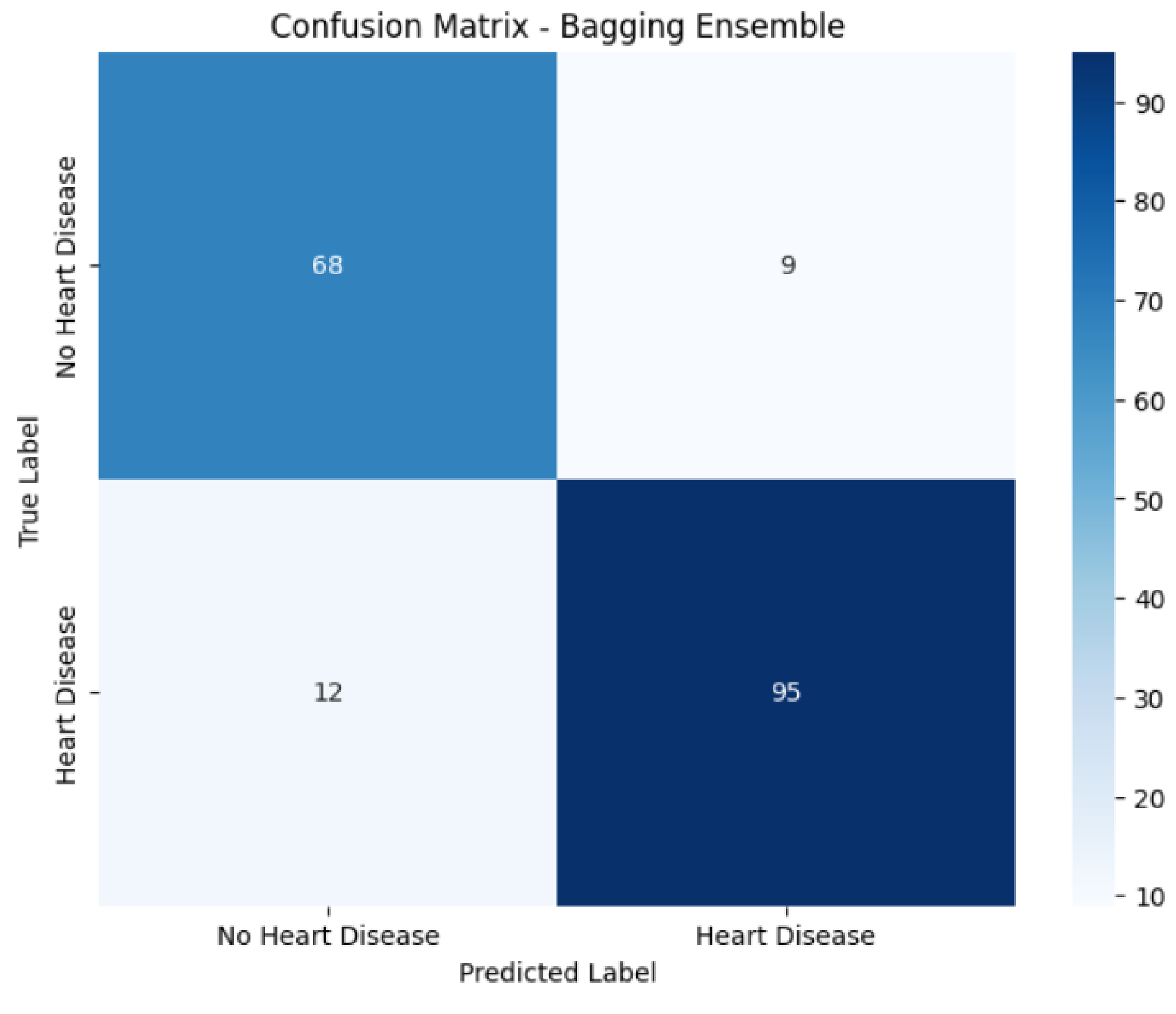
68 (True Negative - TN) 35.05%

23 (False Negative - FN) 11.86%

9 (False Positive - FP) 4.64%

94 (True Positive - TP) 48.45%

Result (Confusion Matrix)



Bagging :

68 (True Negative - TN) 36.96%

12 (False Negative - FN) 6.52%

9 (False Positive - FP) 4.89%

95 (True Positive - TP) 51.63%

Results

Methods	Precision	Recall	F1 Score	Accuracy
Random Forest	0.86	0.86	0.86	0.88
Decision Tree	0.69	0.86	0.77	0.78
SVM	0.82	0.84	0.83	0.86
XGBoost	0.84	0.88	0.86	0.88

Methods	Precision	Recall	F1 Score	Accuracy
Bagging Ensemble:	0.91	0.88	0.90	0.89
Stacking	0.90	0.88	0.89	0.88
AdaBoost	0.86	0.78	0.82	0.80

🌍 Contribution to Sustainable Development Goals

Goal Alignment

•Supports **SDG 3**: Ensuring healthy lives and promoting well-being for all.

•Targets **early detection** and **reduction of premature mortality** from cardiovascular diseases.

Project Contributions

- Early Risk Prediction of heart failure using ensemble ML models (Random Forest, XGBoost, SVM).
- AI Chatbot Integration for real-time health guidance and preventive tips.
- Accessible Web Interface for both patients and healthcare providers.

Impact

•Reduces false predictions, ensuring reliable diagnosis.

•Encourages preventive care and self-awareness.

•Low-cost, scalable solution improving access to heart health services

Future Scope

1. **Enhanced Accuracy**: Collect more data through user surveys or input forms to train the model on a broader set of cases.

2. **Database Integration**: Store user health profiles and predictions using a connected database system for future tracking, analysis, and personalized recommendations.

Conclusion

After careful examination of the Random Forest, Decision Tree, SVM, and XGBoost algorithms, as well as the integration of ensemble learning strategies like boosting, stacking, and bagging :

- The analysis revealed that bagging emerged as the most effective method in terms of accuracy.

- Bagging achieved the highest accuracy of 88.6% and among all the algorithms XGBoost had a highest Precision of 90.4% on the [Heart Failure Prediction Dataset](2021) .

- This underscores the significance of employing ensemble learning methods to enhance predictive performance in machine learning tasks.

Details of paper publication

Paper	Title	Conference/journal	Status
1.	Optimizing heart failure prediction with ensemble learning technique	9th International Conference on Control Communication, Computing and Automation	Submitted

References

[1] Shital Patil and Surendra Bhosale. Enhancing heart disease prediction through ensemble learning and feature selection. Computer Integrated Manufacturing Systems, 29(4):288–296, 2023.

[2] Muhammad Affan Alim, Shamsheela Habib, Yumna Farooq, and Abdul Rafay. Robust heart disease prediction: a novel approach based on sig nificant feature and ensemble learning model. In 2020 3rd international conference on computing, mathematics and engineering technologies (iCoMET), pages 1–5. IEEE, 2020.

[3] Salam Ismaeel, Ali Miri, and Dharmendra Chourishi. Using the extreme learning machine (elm) technique for heart disease diagnosis. In 2015 IEEE Canada international humanitarian technology conference (IHTC2015), pages 1–3. IEEE, 2015.

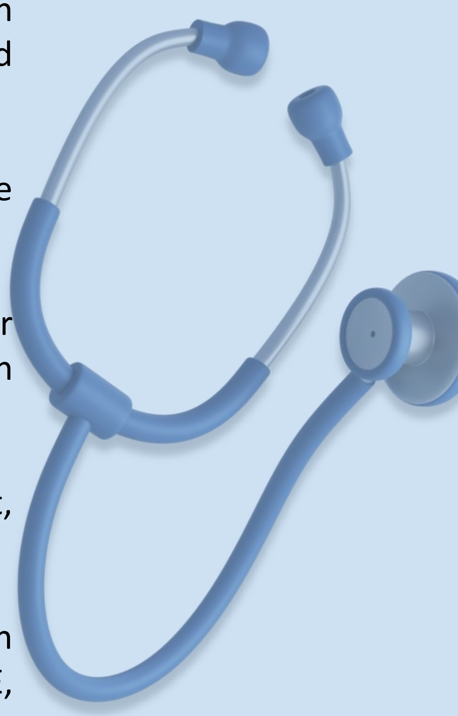
[4] Md Razu Ahmed, SM Hasan Mahmud, Md Altab Hossin, Hosney Jahan, and Sheak Rashed Haider Noori. A cloud based four-tier architecture for early detection of heart disease with machine learning algorithms. In 2018 IEEE 4th International Conference on Computer and Communications (ICCC), pages 1951–1955. IEEE, 2018.

[5] A Lakshmanarao, Y Swathi, and P Sri Sai Sundareswar. Machine learning techniques for heart disease prediction. Forest, 95(99):97, 2019.

[6] Aditi Gavhane, Gouthami Kokkula, Isha Pandya, and Kailas Devadkar. Prediction of heart disease using machine learning. In 2018 second international conference on electronics, communication and aerospace technology (ICECA), pages 1275–1278. IEEE, 2018.

[7] Devansh Shah, Samir Patel, and Santosh Kumar Bharti. Heart disease prediction using machine learning techniques. SN Computer Science, 1(6):345, 2020.

[8] K Rohit Chowdary, P Bhargav, N Nikhil, K Varun, and D Jayanthi. Early heart disease prediction using ensemble learning techniques. In Journal of Physics: Conference Series, volume 2325, page 012051. IOP Publishing, 2022.



THANK YOU!