

**Lab Assignment 2**  
**Explainable AI**  
**E.pavani**  
**Roll No: 2303A52445**  
**Batch - 39**  
**Date: 22-08-2025**

## **Introduction**

This assignment aims at developing predictive models to identify cases of heart disease and study their decision-making process based on the Explainable AI (XAI) method. The predictions by machine learning models can be hard to understand since they tend to behave like black boxes. In healthcare, it is paramount to have explainability as the latter holds because doctors and patients should be capable of trusting and understanding behind model decisions.

This paper trained two models on Heart Disease data- Random Forest and XGBoost, and explained the predictions of the models through SHAP (SHapley Additive exPlanations) to interpret global feature importance and local prediction.

## **Dataset Description**

Source: UCI Heart Disease dataset (Kaggle version – heart.csv)

Size: 1025 samples, 14 columns (13 features + 1 target)

Features:

age – Age of the patient

sex – Gender (1 = male, 0 = female)

cp – Chest pain typ

trestbps – Resting blood pressure

chol – Serum cholesterol level

fbs – Fasting blood sugar > 120 mg/dl

restecg – Resting electrocardiographic results

thalach – Maximum heart rate achieved

exang – Exercise-induced angina

oldpeak – ST depression induced by exercise

slope – Slope of the ST segment

ca – Number of major vessels colored by fluoroscopy

thal – Thalassemia

Target Variable:  
target (0 = No heart disease, 1 = Heart disease present)

## **Preprocessing Steps**

Handling Missing Values

The dataset had no missing values.

For safety, median imputation was applied to any numeric columns if needed.

Feature/Target Split

Independent variables: 13 clinical features

Target variable: target

Train-Test Split

Data split into 80% training (820 samples) and 20% testing (205 samples) using stratification to maintain class balance.

### **Model & Performance**

Two machine learning models were trained and evaluated:

1. Random Forest Classifier

Parameters: n\_estimators = 200, random\_state = 42

Results:

Accuracy: 100%

Precision: 100%

Recall: 100%

F1-score: 100%

ROC-AUC: 1.0

2. XGBoost Classifier

Parameters: n\_estimators = 200, eval\_metric = logloss, random\_state = 42

Results:

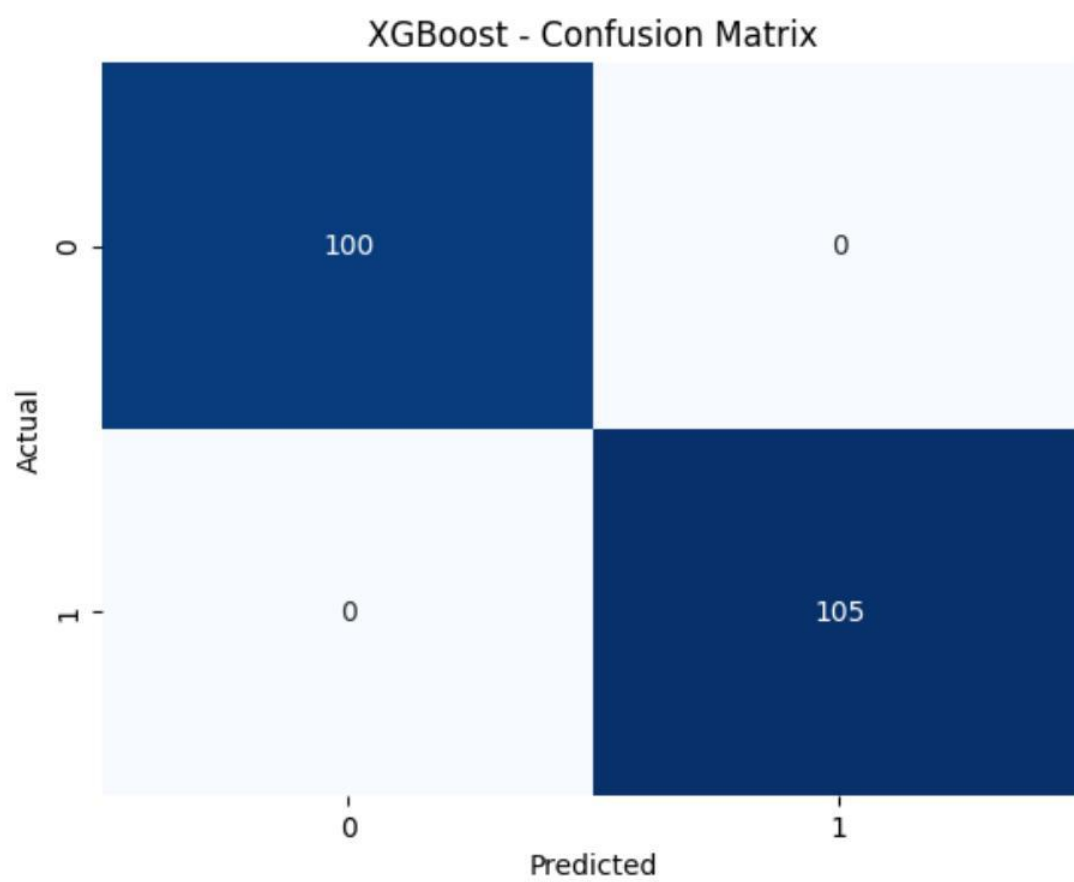
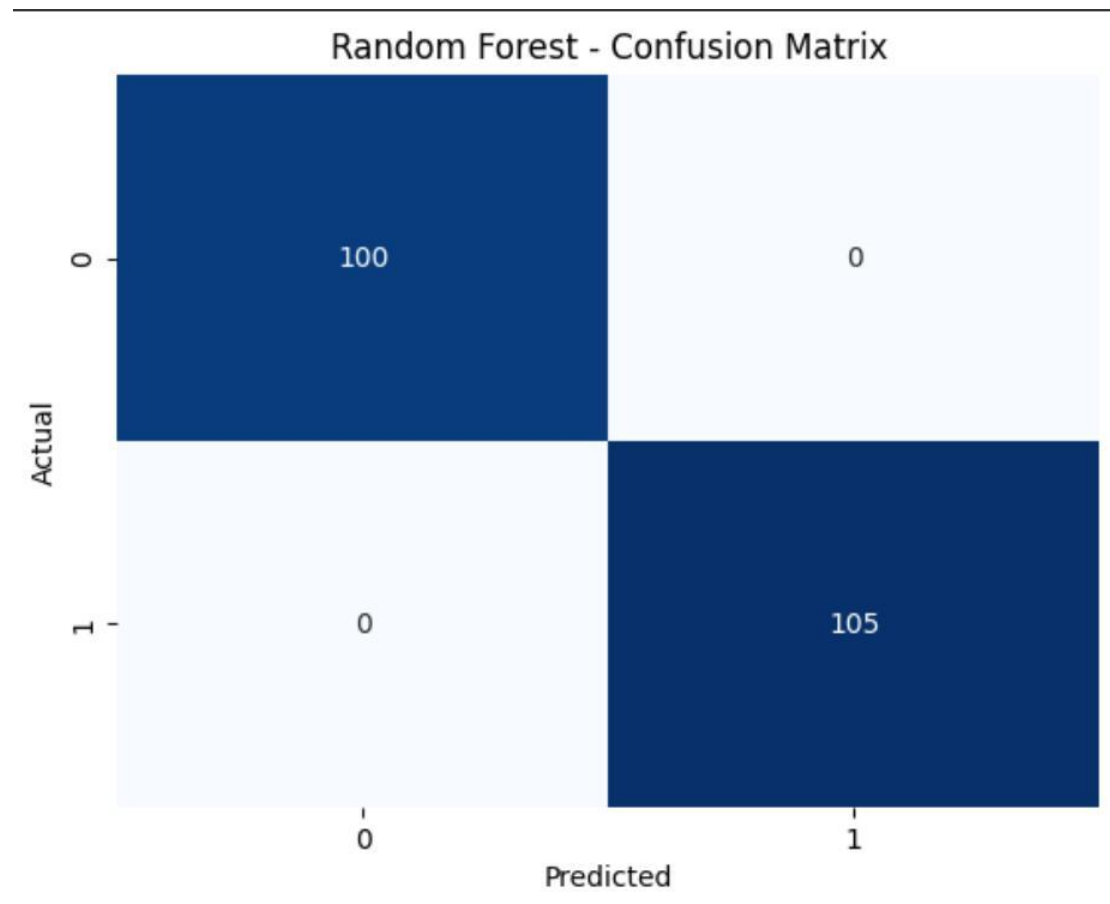
Accuracy: 100%

Precision: 100%

Recall: 100%

F1-score: 100% ROC-AUC: 1.0

Results



**SHAP Analysis**

SHAP was applied to interpret the XGBoost model:

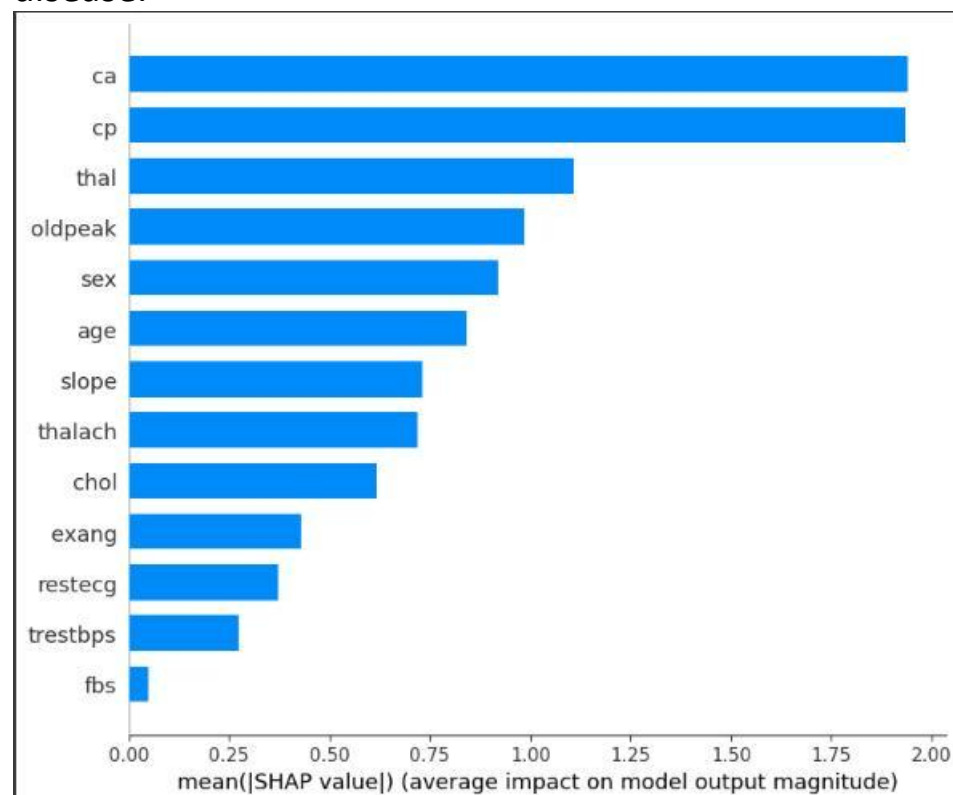
Global Explanations (Feature Importance):

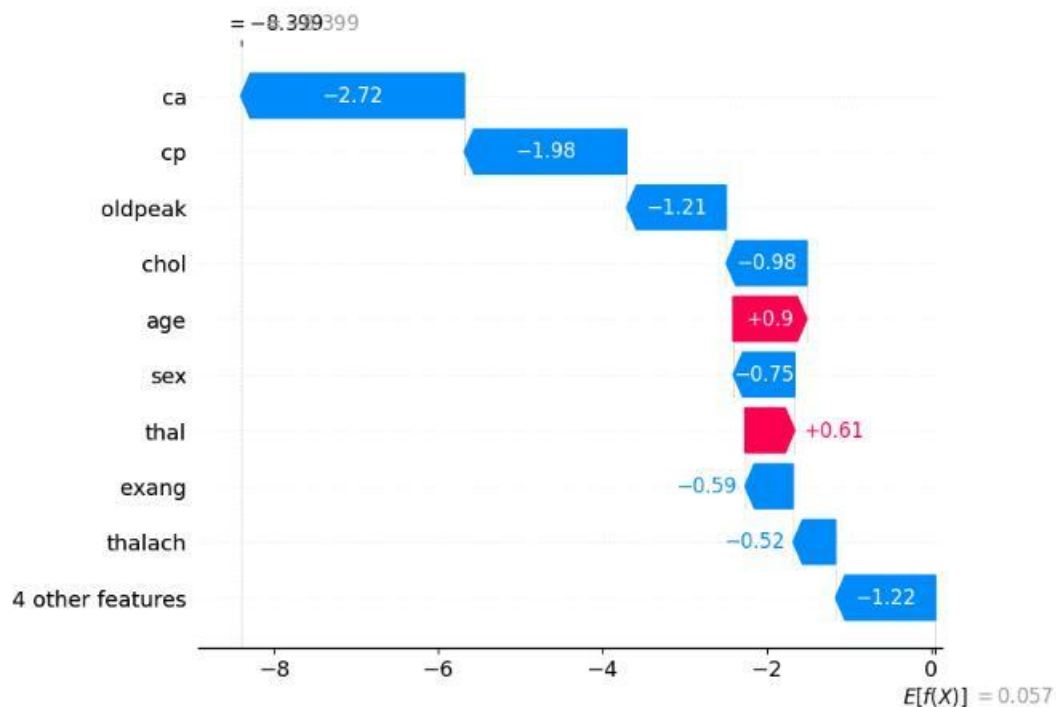
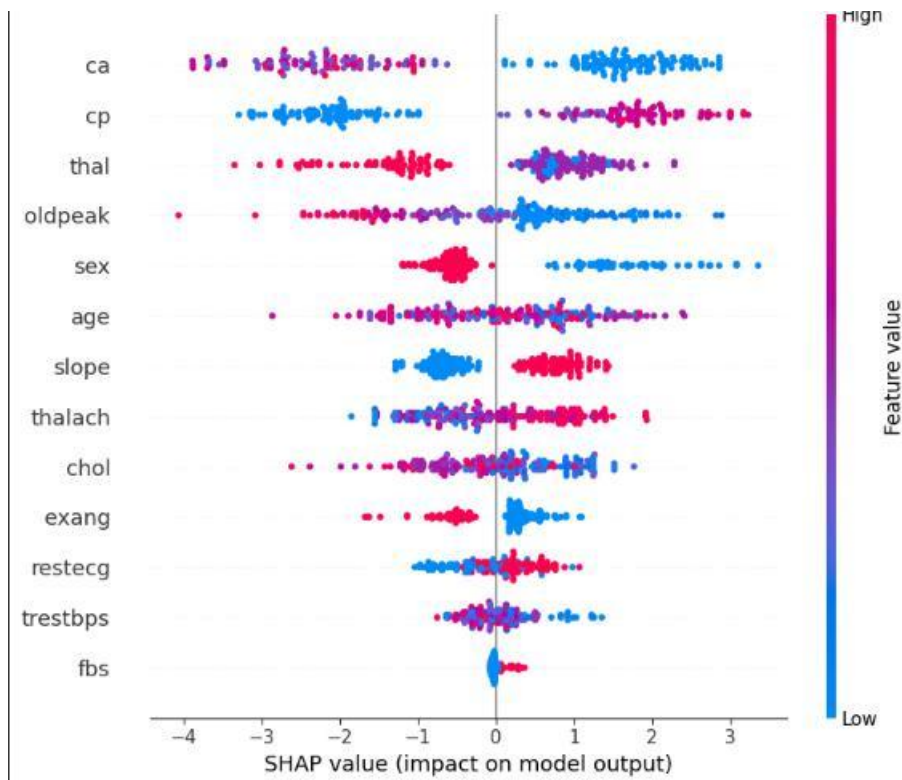
Features such as cp (chest pain type), thalach (max heart rate), oldpeak, ca (vessels), and thal strongly influenced predictions. SHAP summary plots ranked the most important features for predicting heart disease.

Local Explanations (Individual Predictions):

For a test sample (Actual label: 0), SHAP explained why the model predicted very low probability ( $\approx 0.0002$ ) of heart disease.

Force plots and waterfall plots showed how features like oldpeak and thalach pushed the prediction toward “no heart disease.”





## Conclusion

This assignment demonstrated how explainability methods like SHAP can enhance trust in machine learning models applied to healthcare.

Key Insights: Chest pain type (cp), maximum heart rate (thalach), oldpeak, and ca were critical predictors.

SHAP enabled both global (feature ranking) and local (patient-level) explanations.

Limitations:

Perfect accuracy may not generalize well—real-world data is noisier.

Possible model overfitting needs further investigation.

Future Improvements:

Use cross-validation to validate model robustness.

Test on external datasets for generalization.

Compare with other XAI methods (e.g., LIME, Anchors).

Incorporate domain expertise to interpret medical relevance of features.