

Heart Disease Detection using Machine Learning

Shravani Bande

Abstract

Heart disease remains one of the leading causes of mortality worldwide. Early detection can significantly improve outcomes. This project aims to build a machine learning system that predicts whether a patient is likely to have heart disease based on clinical attributes. We explore the dataset, visualize key relationships, train multiple models, and deploy a CLI-based prediction tool.

1. Objective

The goal of this project is to:

- Analyze patient medical data
- Train a predictive model to detect heart disease
- Evaluate model performance
- Build a user-friendly CLI tool for predictions

2. Dataset Description

The dataset contains patient-level data with 11 features and 1 target variable:

- **age**: Age in years
- **sex**: 0 = female, 1 = male
- **chest pain type**: 1-4 (angina/asymptomatic types)
- **resting bp s**: Resting blood pressure (mm Hg)
- **cholesterol**: Serum cholesterol (mg/dl)
- **fasting blood sugar**: 0 = ≤ 120 mg/dL, 1 = > 120 mg/dL
- **resting ecg**: 0 = normal, 1 = abnormal, 2 = LV hypertrophy
- **max heart rate**: Maximum heart rate achieved

- **exercise angina:** 0 = No, 1 = Yes
- **oldpeak:** ST depression induced by exercise
- **ST slope:** 1 = up, 2 = flat, 3 = down
- **target:** 0 = Normal, 1 = Heart Disease

3. Exploratory Data Analysis

We performed initial exploration using `pandas`, `seaborn`, and `matplotlib`. Key steps included:

- Checking dataset shape and column types
- Confirming no missing values
- Visualizing class distribution
- Exploring relationships (e.g., chest pain vs. heart disease)
- Heatmap to examine feature correlation with target

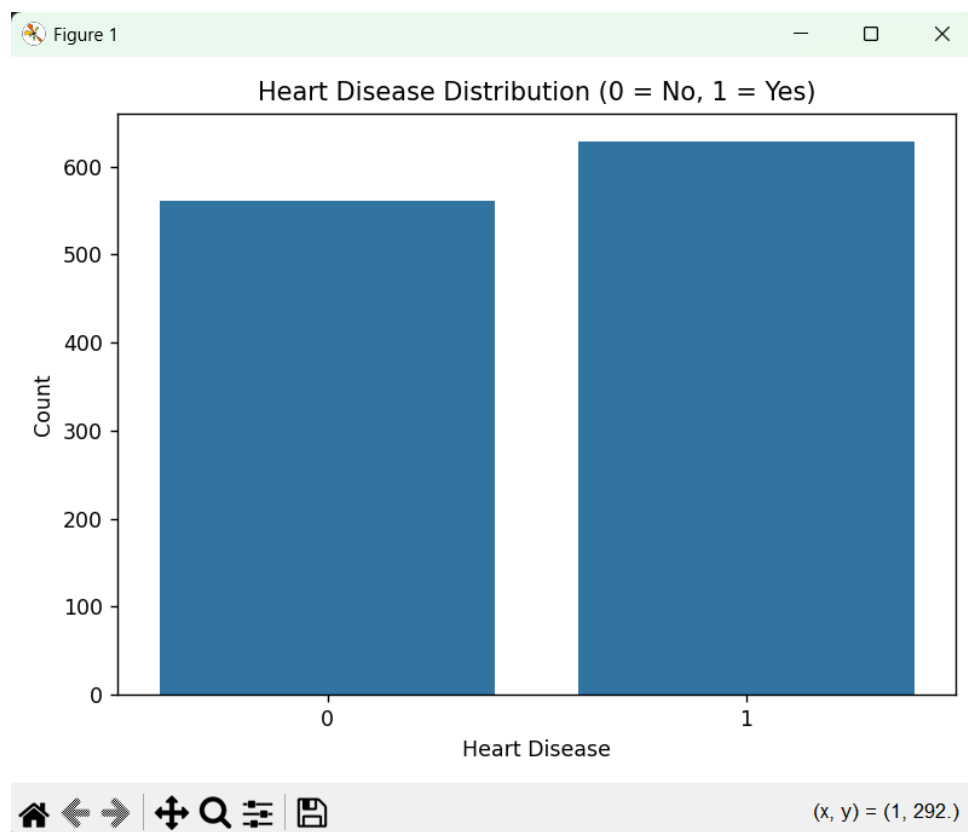


Figure 1: Target Distribution: 0 = No Heart Disease, 1 = Heart Disease

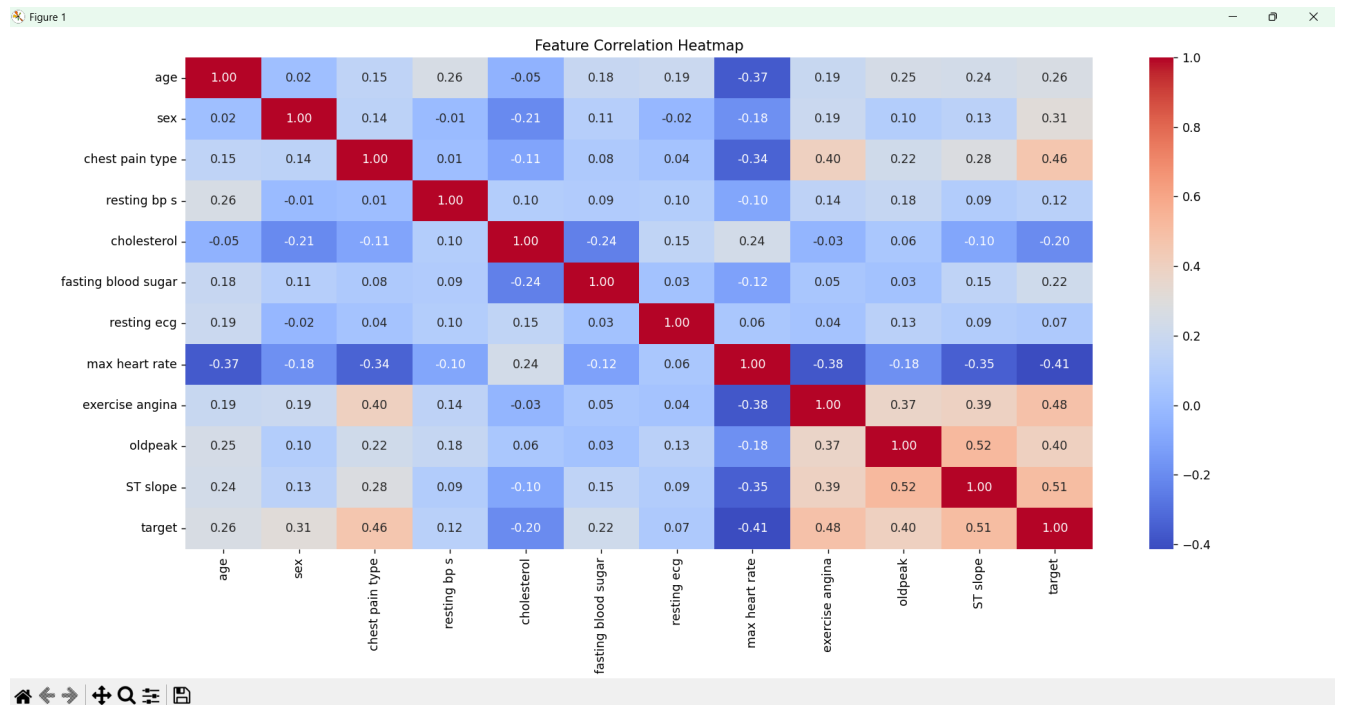


Figure 2: Correlation Heatmap of Features

4. Model Building

Two models were trained using `scikit-learn`:

- **Logistic Regression:** Baseline model with 84% accuracy
- **Random Forest Classifier:** Achieved 92% accuracy and better F1-score

The dataset was split into 80% training and 20% testing. The Random Forest model was saved using `joblib` for deployment.

5. Evaluation Metrics

Model performance was evaluated using:

- **Accuracy**
- **Precision, Recall, F1-score**
- **Confusion Matrix**

Random Forest Results (on test set):

- Accuracy: **92.4%**
- F1-score: **0.93**

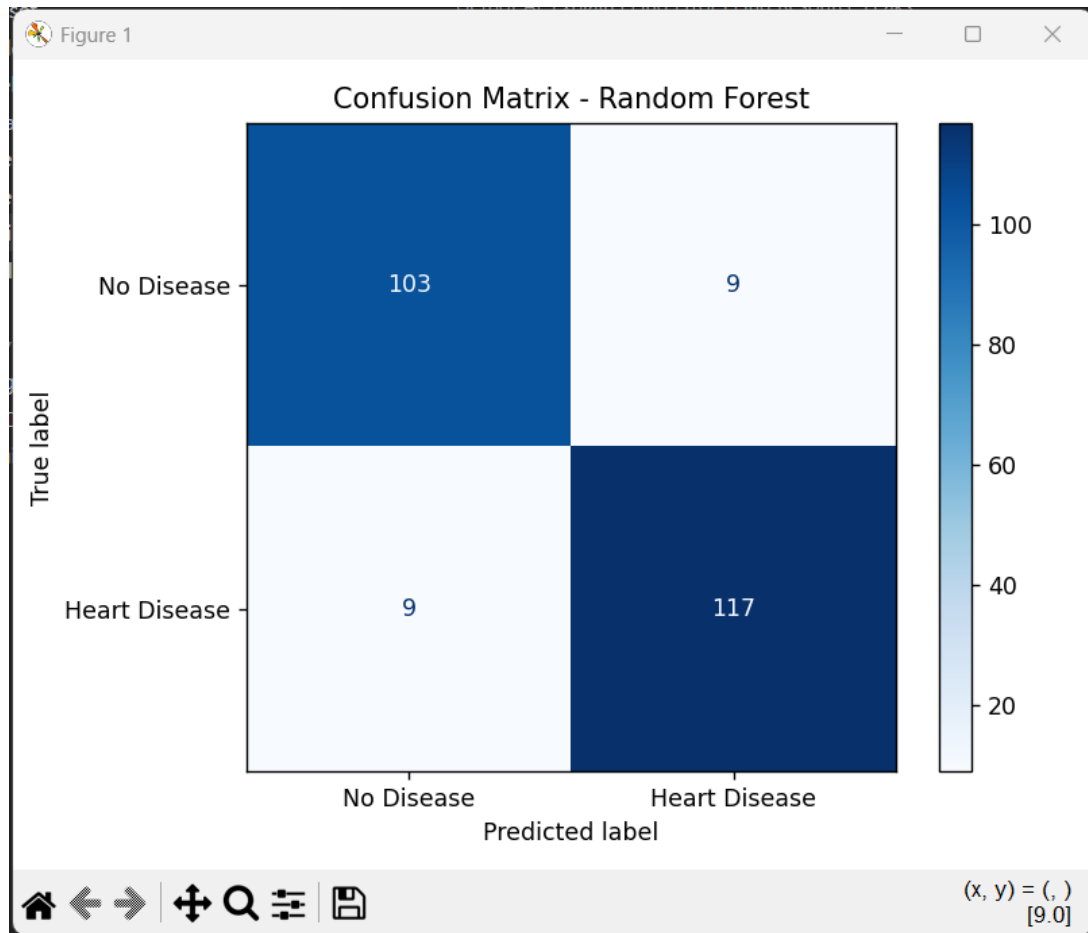


Figure 3: Confusion Matrix - Random Forest Classifier

6. CLI-based Prediction Tool

A Python script `predict_cli.py` was developed for command-line prediction:

- Asks user to enter patient details with input validation
- Passes values to the trained model
- Displays prediction result (Healthy or Heart Disease)

7. Conclusion

The Random Forest model was successful in predicting heart disease with high accuracy and low false positives. The CLI tool makes this model usable in a real-world context. This project demonstrates the potential of machine learning in healthcare screening.

8. Future Work

- Build a web-based UI using Streamlit

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS QUERY RESULTS SIXTH JIRA GPT4
• (venv) PS C:\Users\shrav\Desktop\Heart_Disease_Detection> python notebooks/predict_cli.py

🔗 Heart Disease Prediction - Enter Patient Info Below

Age: 40
Sex (0 = Female, 1 = Male): 0
Chest Pain Type (1=typical angina, 2=atypical angina, 3=non-anginal pain, 4=asymptomatic): 2
Resting BP (mm Hg): 150
Cholesterol (mg/dl): 60
❌ Please enter a value between 100 and 600
Cholesterol (mg/dl): 101
Fasting Blood Sugar (0 = sugar<120mg/dL, 1 = sugar>120mg/dL): 1
Resting ECG (0=normal, 1=abnormal, 2=LV hypertrophy): 0
Max Heart Rate Achieved: 110
Exercise Induced Angina (0 = No, 1 = Yes): 0
Oldpeak (ST depression): 5
ST Slope (1 = upward, 2 = flat, 3 = downward): 2

Predicted class: 1 (0 = Normal, 1 = Heart Disease)

💔 Patient is likely to have heart disease.
• (venv) PS C:\Users\shrav\Desktop\Heart_Disease_Detection> |
```

Figure 4: CLI Prediction Example

- Integrate additional datasets for robustness
- Deploy the app using Streamlit Cloud or Hugging Face Spaces

GitHub Repository

The complete project and source code is available at: https://github.com/savi-08/Heart_Disease_Detection