Heart Disease Risk Predictor

Mohid Ali
Department of Computer Science
Lahore University of Management Sciences (LUMS)

December 30, 2025

Abstract

Heart disease is a major global health concern, and early detection plays a critical role in reducing mortality. This project explores the use of machine learning models to predict the presence of heart disease using patient clinical data. A decision tree classifier is used as a baseline model, while a neural network serves as the main model. The dataset is carefully preprocessed, including handling missing values, feature scaling, and conversion of the target variable into a binary format. Model performance is evaluated using accuracy, precision, recall, confusion matrices, and ROC-AUC. The results demonstrate that neural networks (main model) provide improved predictive performance over the baseline (decision tree) while highlighting important ethical considerations and limitations.

Contents

1 Introduction **4**

2 Dataset Description **4**
- 2.1 Features 4
- 2.2 Target Variable 5
- 2.3 Data Preprocessing 5

3 Methodology **5**
- 3.1 Baseline Model 5
- 3.2 Main Model 5
- 3.3 Training Procedure 6

4 Evaluation Metrics **6**

5 Results **7**
- 5.1 Baseline vs Neural Network 7
- 5.2 ROC Curve and AUC 7

6 Explainability **7**

7 Real-World Demonstration **8**

8 Ethics and Limitations **8**

9 Conclusion **8**

1 Introduction

Heart disease is one of the leading causes of death worldwide. Early identification of individuals at risk allows for early intervention and improved patient outcomes. Machine learning provides a data-driven approach for analyzing medical datasets and identifying complex patterns that may not be immediately apparent using traditional methods. In this project, supervised learning is used to predict heart disease as a binary classification task using clinical patient data.

2 Dataset Description

The dataset used in this project is the Heart Disease dataset obtained from the UCI Machine Learning Repository. It consists of 318 patient records with multiple clinical attributes collected during routine medical examinations.

2.1 Features

The dataset contains 13 input features describing patient characteristics, including:

- **age**: age of the patient in years
- **sex**: biological sex (0 = female, 1 = male)
- **cp**: chest pain type
  - **Value 1**: typical angina
  - **Value 2**: atypical angina
  - **Value 3**: non-anginal pain
  - **Value 4**: asymptomatic
- **trestbps**: resting blood pressure (mm Hg)
- **chol**: serum cholesterol level (mg/dl)
- **fbs**: fasting blood sugar (greater than 120 mg/dl (1 = true; 0 = false))
- **restecg**: resting electrocardiographic results
  - **Value 0**: normal
  - **Value 1**: having ST-T wave abnormality
  - **Value 2**: showing probable or definite left ventricular hypertrophy
- **thalach**: maximum heart rate achieved
- **exang**: exercise-induced angina (1 = true; 0 = false)
- **oldpeak**: ST depression induced by exercise
- **slope**: slope of the peak exercise ST segment
  - **Value 1**: upsloping

  - **Value 2**: flat
  - **Value 3**: downsloping

- **ca**: number of major vessels colored by fluoroscopy
- **thal**: thalassemia status
  - **Value 3**: normal
  - **Value 6**: fixed defect
  - **Value 7**: reversable defect

These features represent numerical clinical indicators commonly used in cardiac assessment.

2.2 Target Variable

The original target variable num in the dataset contained multiple values indicating different severity levels of heart disease. For this project, the target was converted into a binary format:

- 0: no heart disease
- 1: presence of heart disease

This transformation simplifies the task into a binary classification problem suitable for risk screening.

2.3 Data Preprocessing

Ca and thal columns contained missing values represented by "?". This was identified when their datatype was shown as object instead of float. To handle this, they were replaced with proper missing value indicators (Nan) and then replaced with mode . Feature scaling was applied using standardization to ensure all numerical features had comparable ranges. The dataset was then split into training, validation, and test sets using stratified sampling to preserve class balance.

3 Methodology

3.1 Baseline Model

A decision tree classifier was used as a baseline model due to its simplicity and interpretability. This model provides an easily understandable reference for comparison against our main model.

3.2 Main Model

The main model used in this project is a fully connected neural network. The network consists of 2 hidden layers with ReLU activation functions and dropout regularization. Dropout was applied to reduce overfitting, which is particularly important given the limited dataset size. Softmax was used in the output layer and model outputs the probability of the presence of heart disease.

3.3 Training Procedure

The neural network was trained using mini-batch gradient descent with the Adam optimizer to ensure stable and efficient learning. A cross-entropy loss function was used to update weights during back propagation. Model performance was also monitored at the end of each epoch.

Hyperparameters such as the number of hidden layers, dropout rates, and learning rates were tuned through experimentation to improve generalization and reduce overfitting. Loss and accuracy curves for both the training and validation sets were plotted to visually assess convergence and detect signs of overfitting. The model achieving the best validation performance was saved and used for final evaluation.

4 Evaluation Metrics

Model performance was evaluated using accuracy, precision, recall, and F1-score. While accuracy provides an overall measure of correctness, it can be misleading in medical screening tasks where class imbalance may exist. Therefore, recall was prioritized to minimize false negatives, as incorrectly classifying a patient with heart disease as healthy can delay diagnosis and treatment.

To better understand model errors, confusion matrices were analyzed, allowing explicit examination of false positive and false negative predictions. In addition, ROC curves and AUC scores were generated using predicted probabilities to evaluate model performance across different decision thresholds rather than relying on a single classification threshold.

Hyperparameters were tuned with a specific focus on improving recall, even when this required accepting a modest increase in false positives. This trade-off is appropriate in healthcare-related applications where early detection is more critical.

To ensure reproducibility and deterministic evaluation, random seeds were fixed, and deterministic training behavior was enforced. This ensured that model training, validation splits, and evaluation metrics remained consistent across multiple runs, allowing reliable comparison of results.

Baseline Tree evaluation on test set:
Accuracy:0.7708333333333334
Precision:0.8125
Recall:0.6190476190476191
F1 Score:0.7027027027027026

Test Classification Report:

	precision	recall	f1-score	support
0	0.75	0.89	0.81	27
1	0.81	0.62	0.70	21
accuracy			0.77	48
macro avg	0.78	0.75	0.76	48
weighted avg	0.78	0.77	0.77	48

(a) Baseline Model Evaluation

Neural Network evaluation on test set:
Accuracy:0.8333333333333334
Precision:0.8421052631578947
Recall:0.7619047619047619
F1 Score:0.8

Test Classification Report:

	precision	recall	f1-score	support
0	0.83	0.89	0.86	27
1	0.84	0.76	0.80	21
accuracy			0.83	48
macro avg	0.83	0.83	0.83	48
weighted avg	0.83	0.83	0.83	48

(b) Main Model Evaluation

Figure 1: Comparison of baseline and main model

5 Results

5.1 Baseline vs Neural Network

The baseline decision tree showed a higher number of false negatives, indicating missed disease cases. In contrast, the neural network demonstrated improved recall and overall predictive performance.

5.2 ROC Curve and AUC

ROC curves were constructed using predicted probabilities from both models. The neural network achieved a higher AUC value, indicating better classification between patients with and without heart disease.

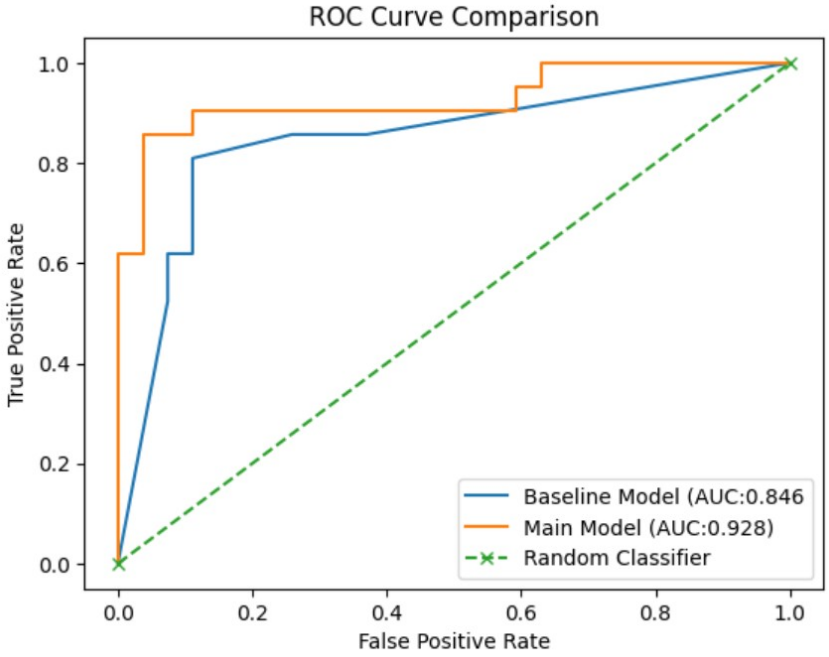

Figure 2: ROC Curve Comparison

6 Explainability

This was explored to better understand how and why the models make their predictions, particularly in a healthcare context where transparency is important. Identifying influential features helps build trust in the model and provides insight into which clinical attributes contribute most to predicted risk.

For the decision tree model, feature importance scores were used to highlight the clinical attributes that had the strongest influence on prediction outcomes.

For the neural network, direct interpretability is more challenging. Therefore, approximate feature influence was analyzed by examining the magnitude of weights in the input layer, which provides a general indication of which features have a stronger impact on the model's output.

7 Real-World Demonstration

A single patient from the test set was selected to demonstrate model usage. The main model outputs a predicted risk probability along with a final classification, simulating a real-world screening scenario.

8 Ethics and Limitations

This system is designed solely as a decision-support and learning tool and does not provide medical advice. It should not be used for diagnosis or treatment planning without the involvement of qualified medical professionals. The predictions generated by the model are based on historical data patterns and do not account for individual patient circumstances.

The dataset used in this study may also contain demographic, geographic, and sampling biases, as it represents a limited population collected under specific conditions. As a result, the model's performance may not generalize well to broader or more diverse patient groups. This highlights the need for caution when interpreting results outside the original data context.

Before any real-world deployment, the model would require extensive clinical validation, including testing on larger and more diverse datasets, evaluation by medical experts, and alignment with regulatory standards. In addition, strict data privacy and security measures would be essential to protect sensitive patient information and ensure compliance with healthcare data protection regulations.

9 Conclusion

This project demonstrates the potential of Machine learning algorithms for heart disease risk prediction using clinical data. Compared to a simple baseline, the neural network provides improved predictive performance while maintaining acceptable generalization. Future work includes using larger and more diverse datasets, improving model interpretability, and validating performance in clinical settings.

References

[1] UCI Machine Learning Repository, *Heart Disease Dataset*. Available at: https://archive.ics.uci.edu