Comparing Classical Machine Learning and Deep Learning Approaches for Heart Disease Prediction

Nuredin Muhumed, Mahad Kashif, Abdirahman Hassan

University of Minnesota

Abstract

Heart disease remains one of the leading causes of death worldwide. This makes accurate early prediction an important part of preventive healthcare. We compare a classical machine learning model (logistic regression) with a fully connected deep neural network trained in PyTorch using the UCI Cleveland Heart Disease dataset. After we applied standardized preprocessing, hyperparameter tuning, and evaluation, logistic regression achieved significantly better results, especially in accuracy, F1-score, and ROC-AUC. Our findings show that classical models remain highly effective for small medical datasets and often generalize better than deep learning methods.

Introduction

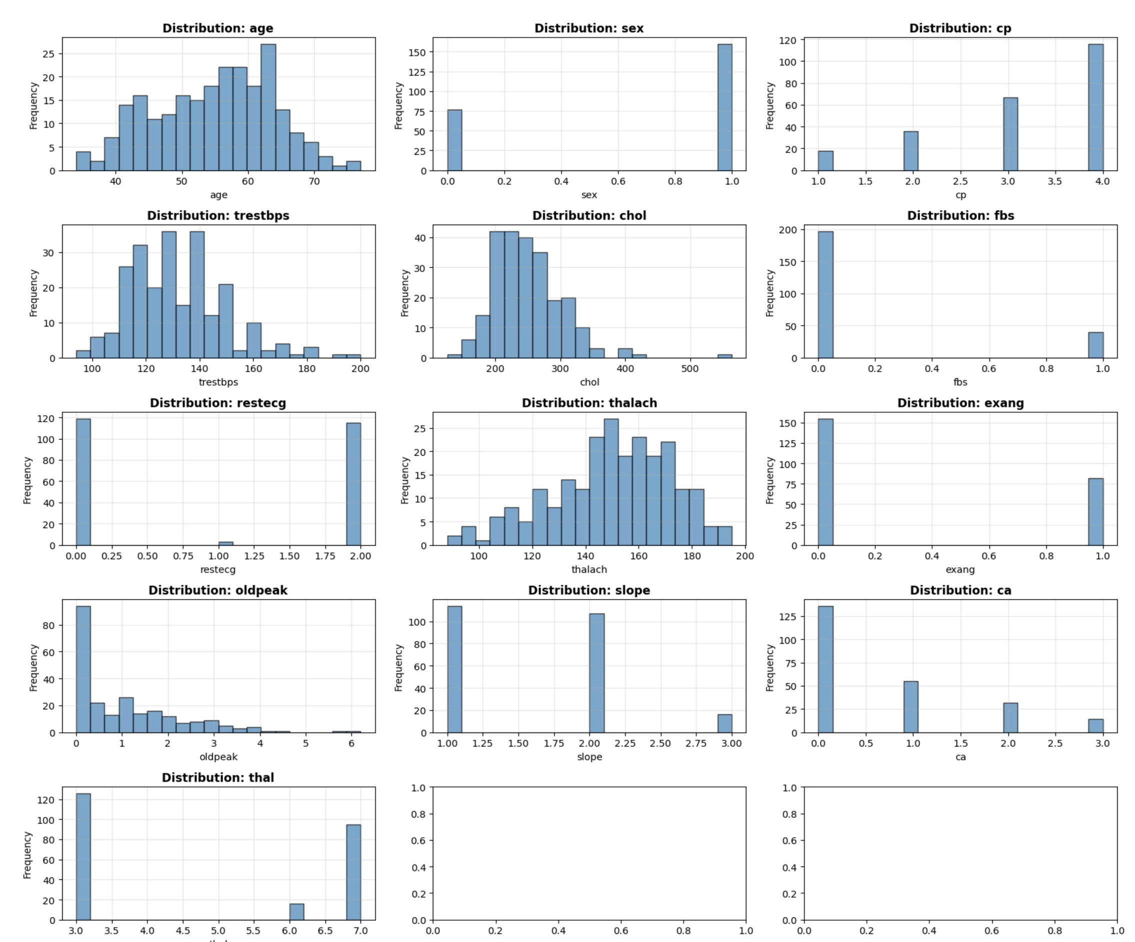
Machine learning has become an important tool for analyzing structured clinical data, especially as healthcare systems work to improve early detection of heart disease. The UCI Cleveland Heart Disease dataset is a widely used benchmark in this area because it contains well-defined clinical features and has a long history of strong performance from classical models. This makes it a suitable dataset for exploring how different modeling approaches behave when working with limited, tabular medical data.

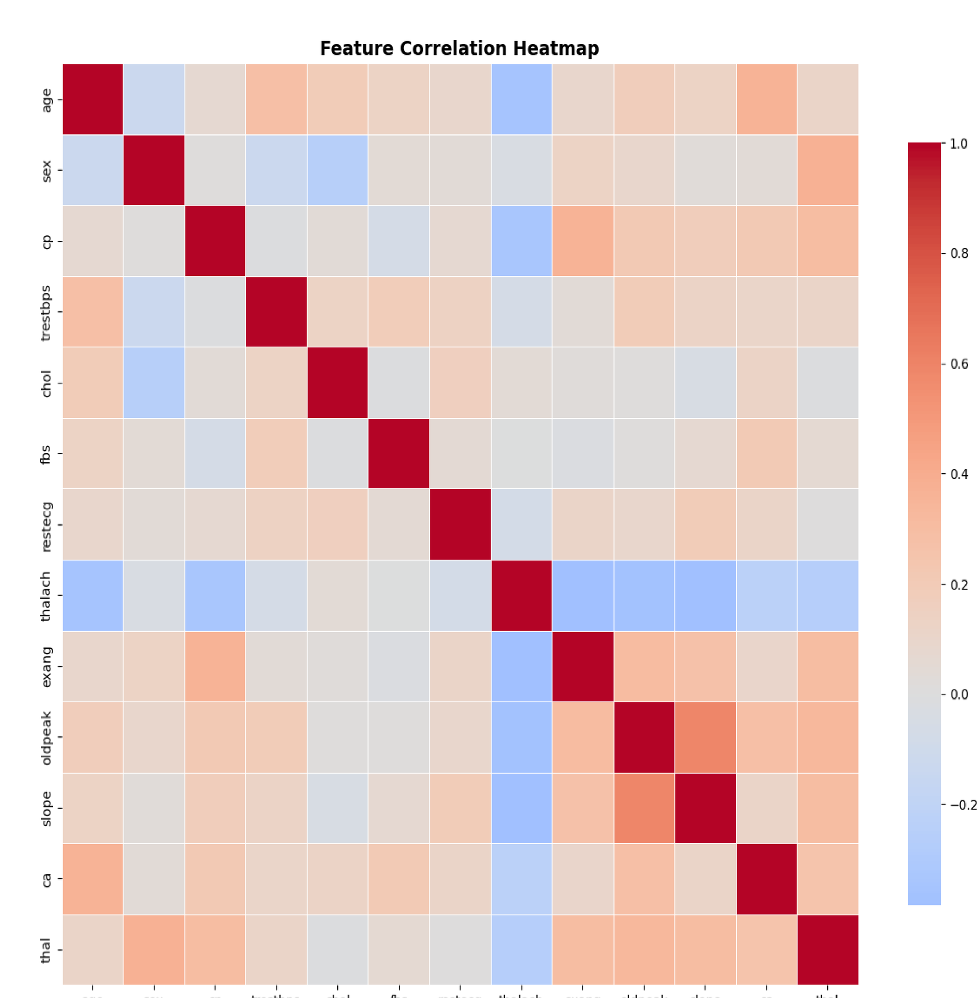
In this project, we compared logistic regression with a custom fully connected neural network to examine how model complexity affects performance, interpretability, and generalization. By training both models under consistent preprocessing and tuning procedures, we aimed to understand not only which method performs better, but also how dataset size, feature structure, and learning dynamics influence their behavior in a clinical prediction setting.

Methodology

We used the UCI Cleveland Heart Disease dataset, consisting of 303 samples and 13 clinical features such as chest pain type, cholesterol level, blood pressure, maximum heart rate, and ECG results. After removing rows with missing critical values, 297 samples remained. The target variable was binarized into "disease" vs. "no disease." We standardized all features using a training-fit scaler and appled an 80/20 stratified split.

We reviewed feature distributions and correlations to understand clinical patterns and detect outliers.





Models

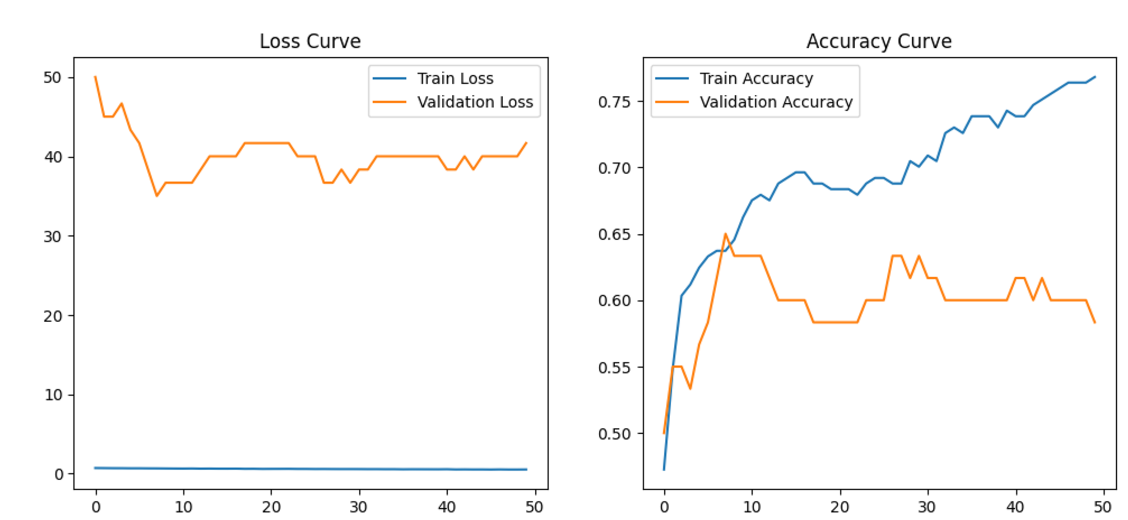
We implemented two models:

Logistic Regression

Chosen for its interpretability, stability on small datasets, and strong baseline performance. Hyperparameter tuning included different penalty types and regularization strengths.

Neural Network (PyTorch)

A fully connected network with two hidden layers (32 – 16 neurons), LeakyReLU activation, and a sigmoid output layer. We tuned learning rate, batch size, epochs, and weight decay.



These curves illustrate that while training accuracy improved steadily, the validation performance remained unstable.

Conclusion

This study demonstrates that classical machine learning models can outperform deep learning methods on small tabular medical datasets. Logistic regression achieved higher accuracy, better calibration, and more stable learning behavior compared to the neural network. The deep learning model struggled due to limited sample size for this dataset and hight variance. We might include in future directions expanding the dataset, using cross-validation, and exploring ensemble approaches, or applying explainable AI techniques to improve transparency for clinical use.

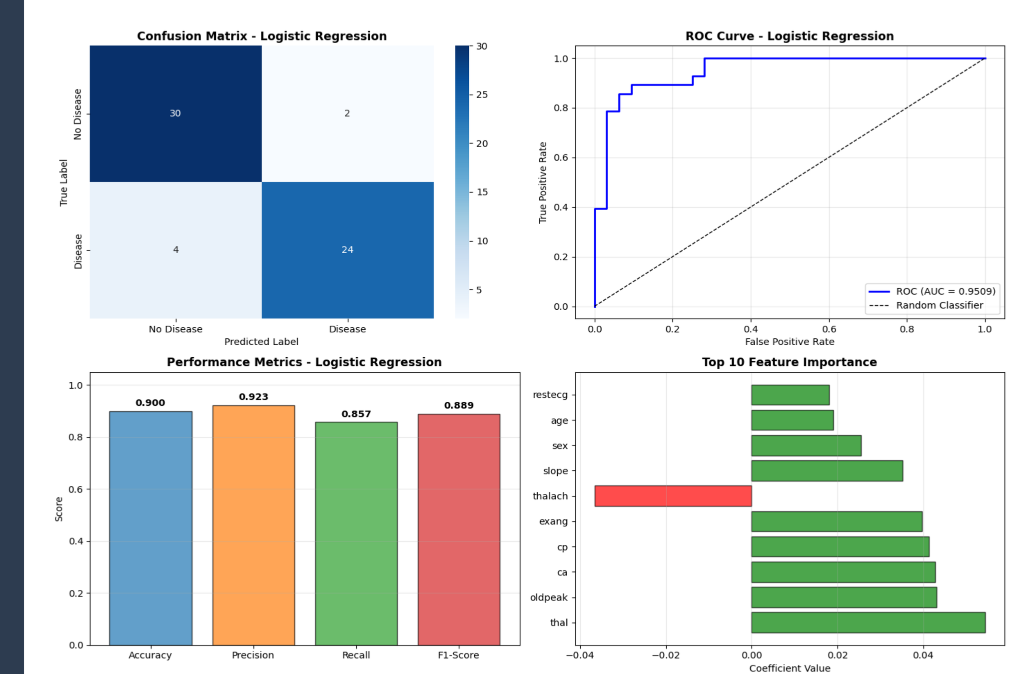
Result

Logistic regression outperformed the neural network:

Logistic Regression Performance:

Accuracy: 0.90, ROC-AUC: 0.95, F1-score: 0.89

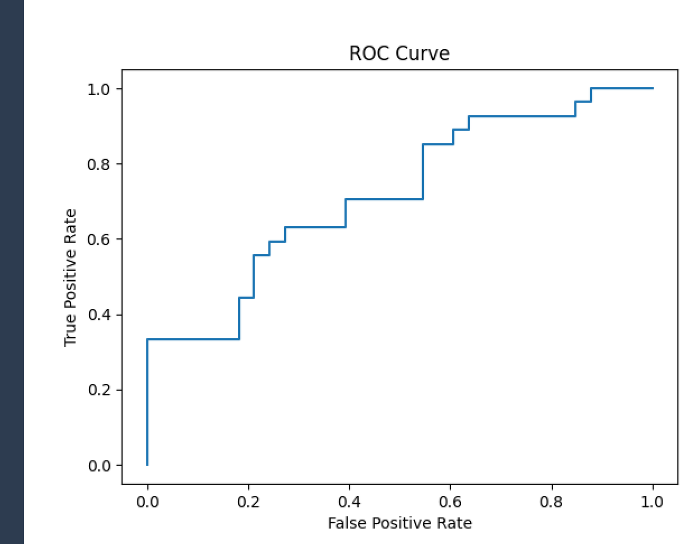
The confusion matrix showed only six total misclassifications, demonstrating strong generalization.



Neural Network Performance:

Accuracy: 0.68, ROC-AUC: 0.71, F1-score: 0.61

The network produced more false negatives, making It less reliable in a health setting.



Confusion Matrix:

[[26 7]

[12 5]]

Acknowledgements

We thank the University of Minnesota and our course instructor and teaching assistants for providing the dataset, computational resources, and all the guidance. We also acknowledge our own members for their contributions to data processing, modeling, and evaluation.