Objective

To develop a logistic regression model in R that predicts the presence of heart disease based on clinical features. The goal is to build an interpretable and reliable classifier using a real-world medical dataset for early risk detection and healthcare decision support.

Dataset

- **Samples:** 303 patients
- **Target variable:** target (1 = disease, 0 = no disease)
- **Features include:** Age, Sex, Chest Pain Type (cp), Resting BP, Cholesterol, Fasting Blood Sugar, Max Heart Rate, etc.

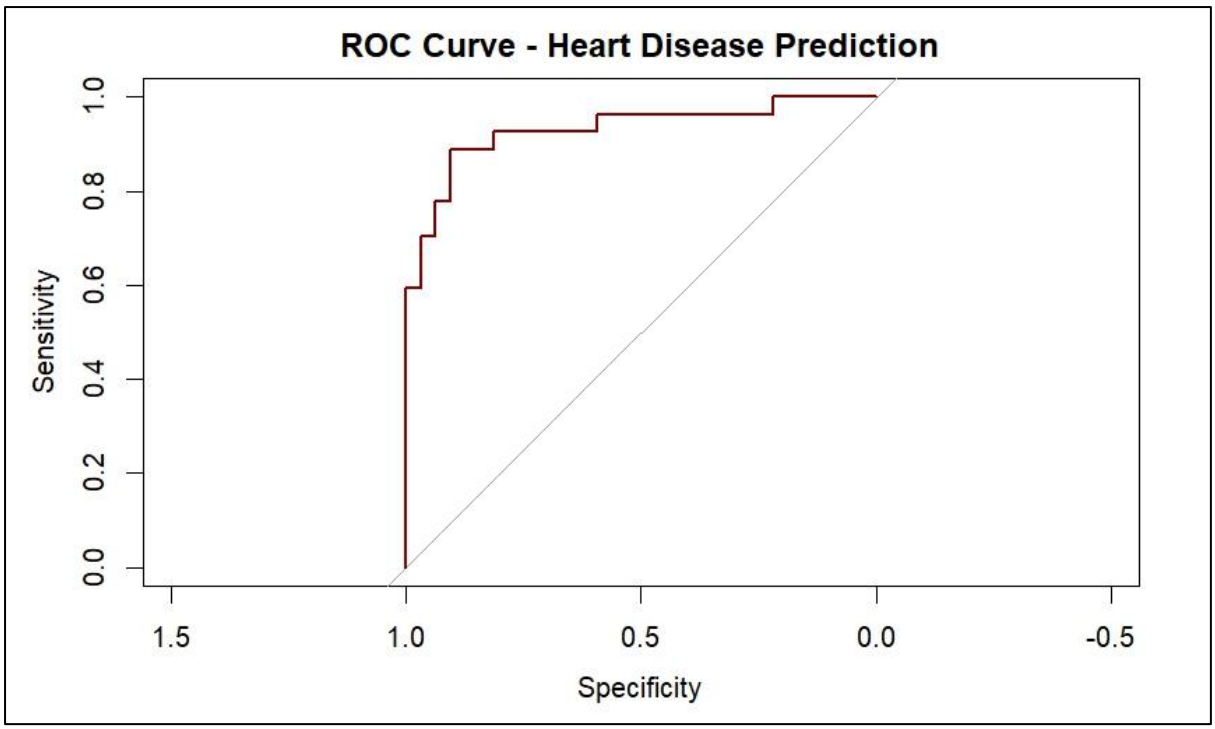
Methods

- Pre-processed categorical variables as factors (sex, cp, thal, etc.)
- Split dataset: 80% training, 20% testing
- Fitted logistic regression model using glm() in R
- Evaluated using confusion matrix and ROC/AUC

Results

Metric	Value
Accuracy	89.8%
Sensitivity	90.6%
Specificity	88.9%
AUC (ROC Curve)	0.93
Kappa Score	0.795

The model performed with high accuracy and discriminative power, showing strong potential for clinical screening applications.



Conclusion

This project demonstrates the use of logistic regression for binary medical classification using real patient data. The model achieved strong predictive metrics and was further enhanced by interpretability. The workflow can be extended to more advanced ML models or deployed as a web app for screening support.