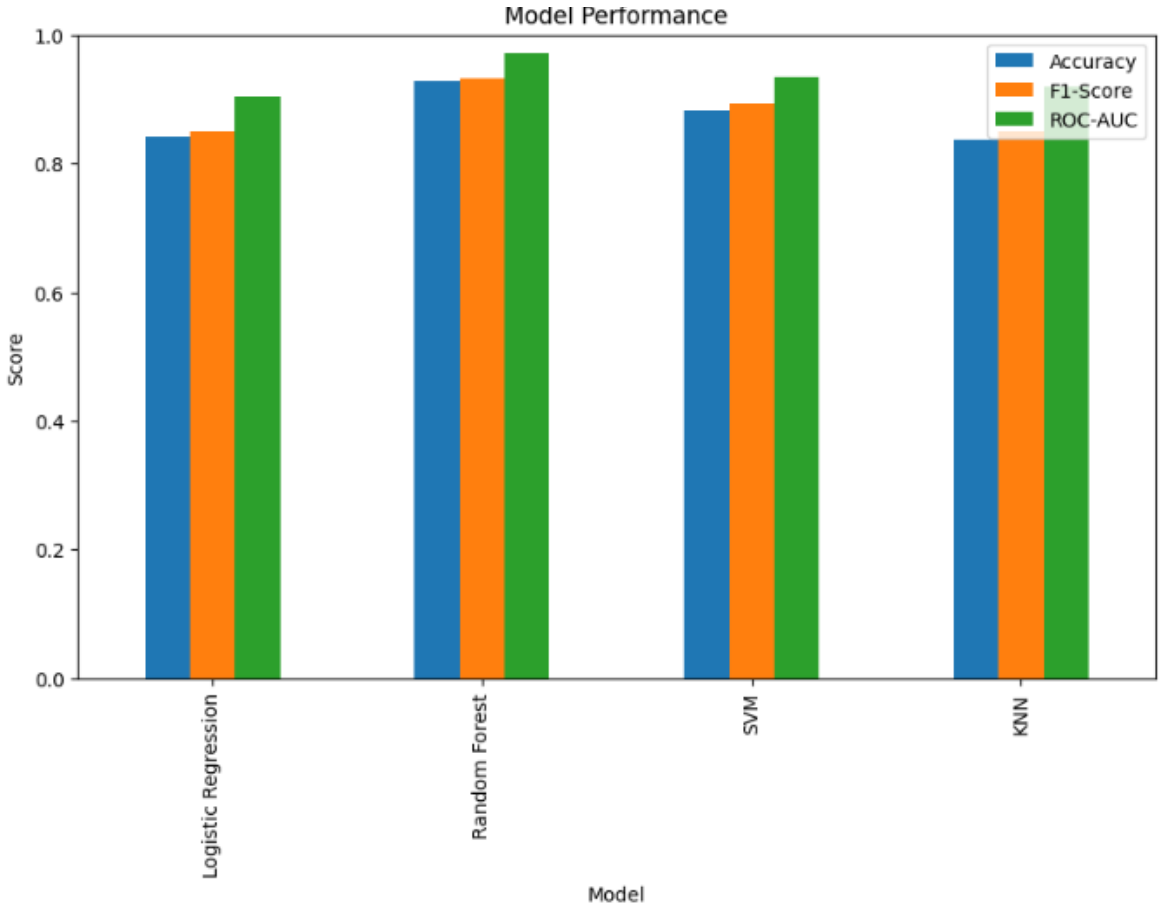
PHASE 4 – MODEL TRAINING REPORT

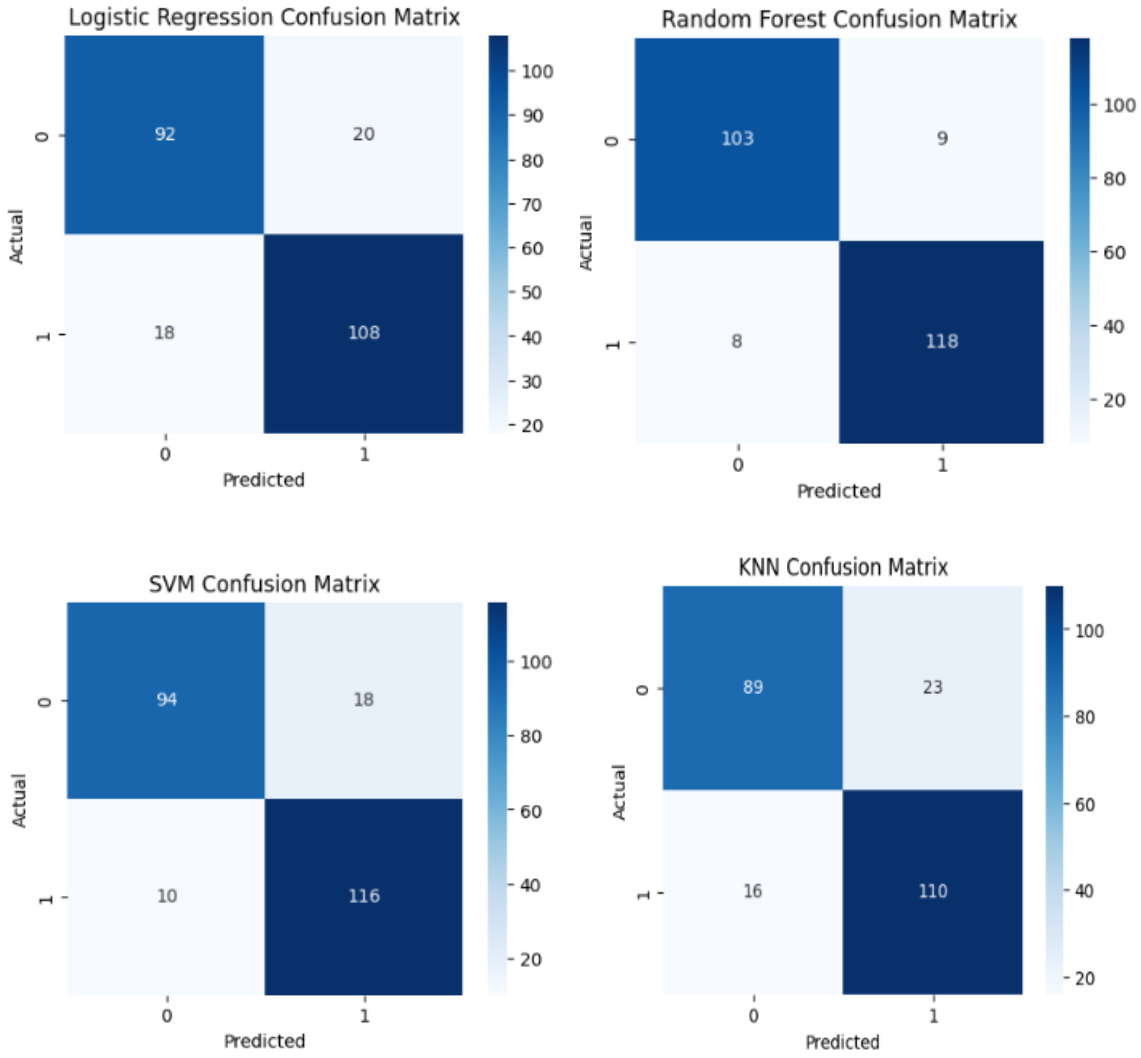
In this phase, I trained and evaluated four supervised learning models (Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN)) to predict heart disease using the cleaned dataset. The goal was to compare their performance and identify the best model

	Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
0	Logistic Regression	0.840336	0.843750	0.857143	0.850394	0.904124
1	Random Forest	0.928571	0.929134	0.936508	0.932806	0.971159
2	SVM	0.882353	0.865672	0.920635	0.892308	0.935197
3	KNN	0.836134	0.827068	0.873016	0.849421	0.918722

Model Comparison Plot



Confusion Metrics



Logistics Regression worked as a simple baseline. Random Forest gave the best overall performance and provided feature importance. SVM also performs well but slows to train. KNN was easy to implement but less accurate, I would choose ***Random Forest*** as the final model for this project.