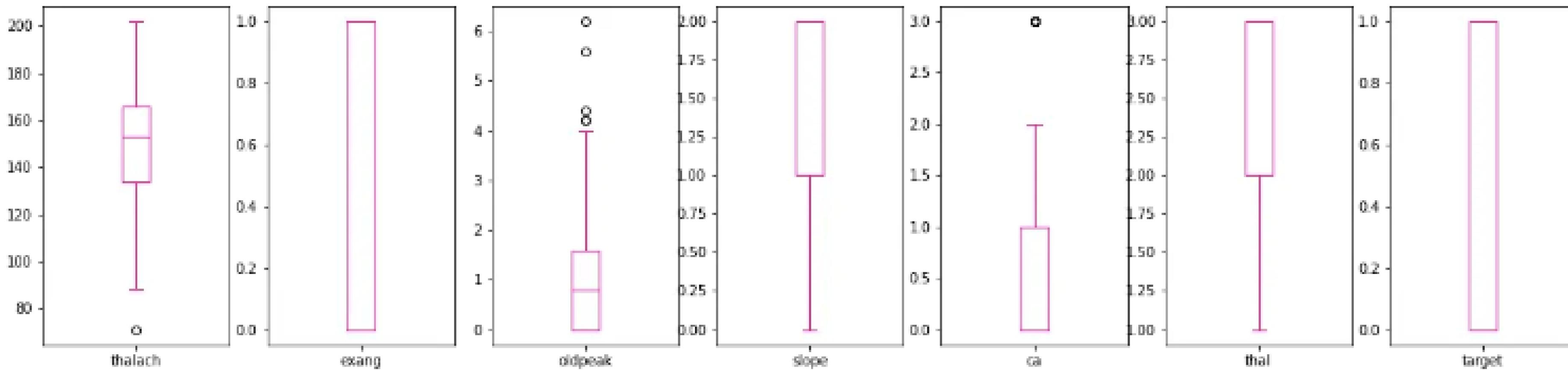
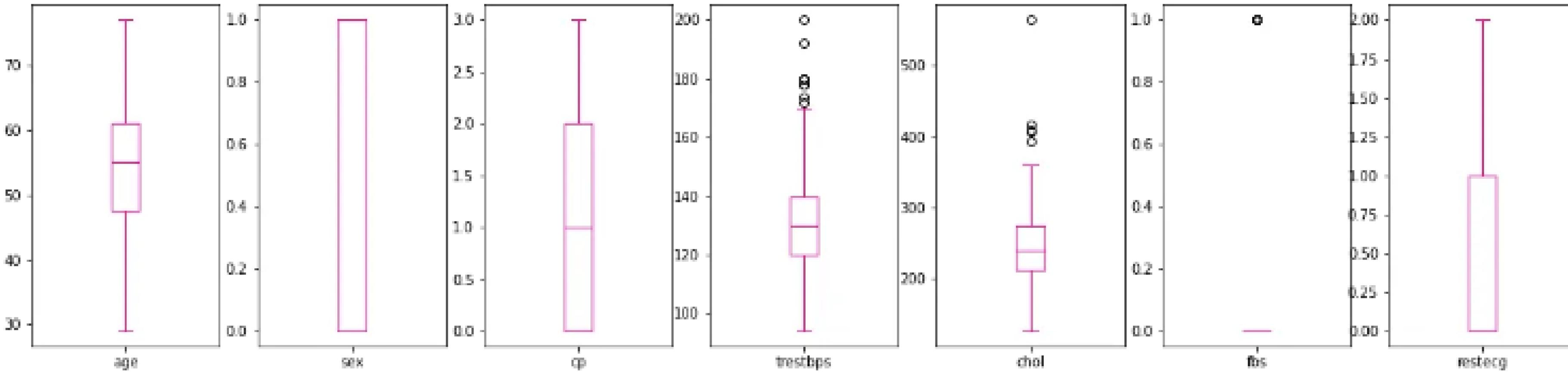
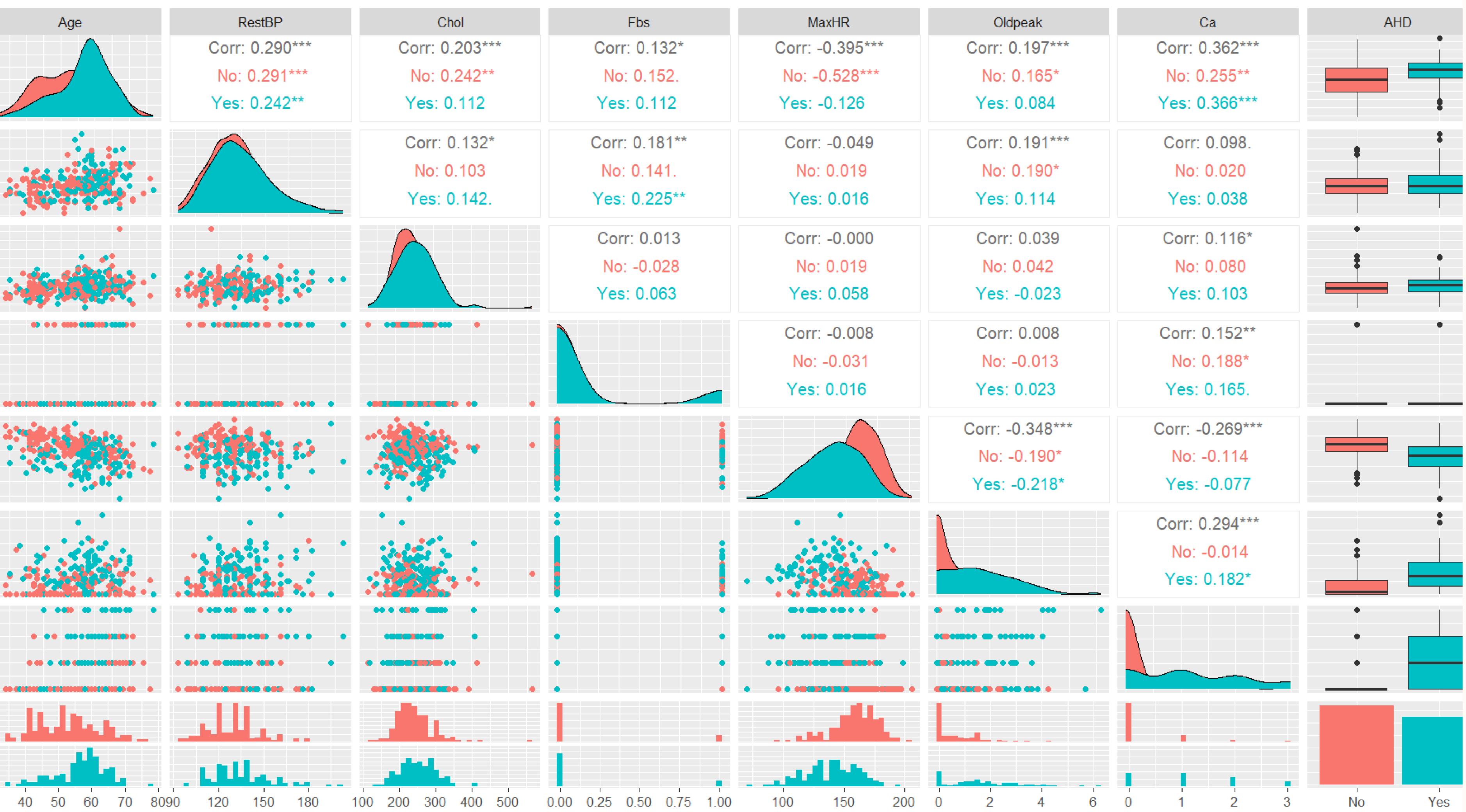


# Heart Disease Prediction

The Heart disease dataset includes 14 variables, with 13 predictors telling about patient's information and symptoms such as Age , cholesterol (Chol), ... and a variable to classify whether or not that patient has heart disease problem (AHD).

X	Age	Sex	ChestPain	RestBP
Min. : 1	Min. :29.00	Min. :0.0000	asymptomatic:142	Min. : 94.0
1st Qu.: 75	1st Qu.:48.00	1st Qu.:0.0000	nonanginal : 83	1st Qu.:120.0
Median :149	Median :56.00	Median :1.0000	nontypical : 49	Median :130.0
Mean :149	Mean :54.54	Mean :0.6768	typical : 23	Mean :131.7
3rd Qu.:223	3rd Qu.:61.00	3rd Qu.:1.0000		3rd Qu.:140.0
Max. :297	Max. :77.00	Max. :1.0000		Max. :200.0
Chol	Fbs	RestECG	MaxHR	ExAng
Min. :126.0	Min. :0.0000	Min. :0.0000	Min. : 71.0	Min. :0.0000
1st Qu.:211.0	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:133.0	1st Qu.:0.0000
Median :243.0	Median :0.0000	Median :1.0000	Median :153.0	Median :0.0000
Mean :247.4	Mean :0.1448	Mean :0.9966	Mean :149.6	Mean :0.3266
3rd Qu.:276.0	3rd Qu.:0.0000	3rd Qu.:2.0000	3rd Qu.:166.0	3rd Qu.:1.0000
Max. :564.0	Max. :1.0000	Max. :2.0000	Max. :202.0	Max. :1.0000
Oldpeak	Slope	Ca	Thal	AHD
Min. :0.000	Min. :1.000	Min. :0.0000	fixed : 18	No :160
1st Qu.:0.000	1st Qu.:1.000	1st Qu.:0.0000	normal :164	Yes:137
Median :0.800	Median :2.000	Median :0.0000	reversible:115	
Mean :1.056	Mean :1.603	Mean :0.6768		
3rd Qu.:1.600	3rd Qu.:2.000	3rd Qu.:1.0000		
Max. :6.200	Max. :3.000	Max. :3.0000		





# LDA Analysis

F1 score=0.8438

Precision=0.871

## Coefficients of linear discriminants:

	LD1
Age	0.005617004
Sex1	0.669600864
ChestPainnonanginal	-1.054413113
ChestPainnontypical	-0.837888840
ChestPaintypical	-0.846239877
RestBP	0.008418292
Chol	0.001400804
Fbs1	-0.198798400
RestECG	0.152508886
MaxHR	-0.004937565
ExAng1	0.479822521
Oldpeak	0.101705621
Slope	0.237843745
Ca	0.455593341
Thalnormal	-0.277224823
Thalreversible	0.615768805

## Confusion Matrix and Statistics

		Reference	
		Prediction	No Yes
No	45	6	
Yes	4	27	

Accuracy : 0.878  
95% CI : (0.7871, 0.9399)  
No Information Rate : 0.5976  
P-Value [Acc > NIR] : 2.358e-08

Kappa : 0.7439  
McNemar's Test P-Value : 0.7518

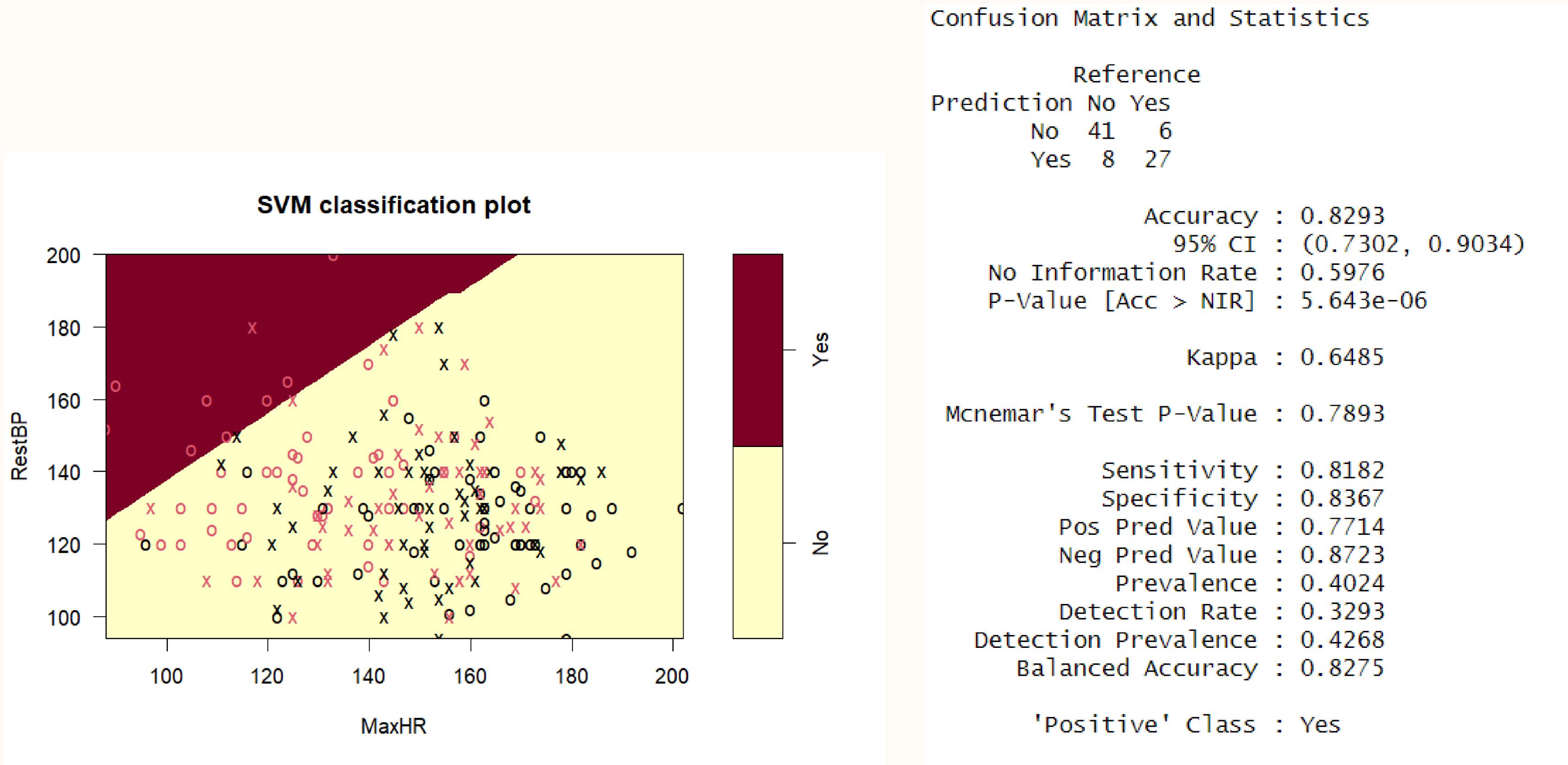
Sensitivity : 0.8182  
Specificity : 0.9184  
Pos Pred Value : 0.8710  
Neg Pred Value : 0.8824  
Prevalence : 0.4024  
Detection Rate : 0.3293  
Detection Prevalence : 0.3780  
Balanced Accuracy : 0.8683

'Positive' Class : Yes

# SVM linear kernel (cost=0.1)

F1 score=0.7941

Precision=0.7714



# Random Forest

F1 score=0.75

Precision=0.7742

## Confusion Matrix and Statistics

		Reference	
		Prediction	No Yes
Prediction	No	42	9
	Yes	7	24

Accuracy : 0.8049  
95% CI : (0.7026, 0.8842)

No Information Rate : 0.5976  
P-Value [Acc > NIR] : 5.228e-05

Kappa : 0.5903

McNemar's Test P-Value : 0.8026

Sensitivity : 0.7273  
Specificity : 0.8571

Pos Pred Value : 0.7742

Neg Pred Value : 0.8235

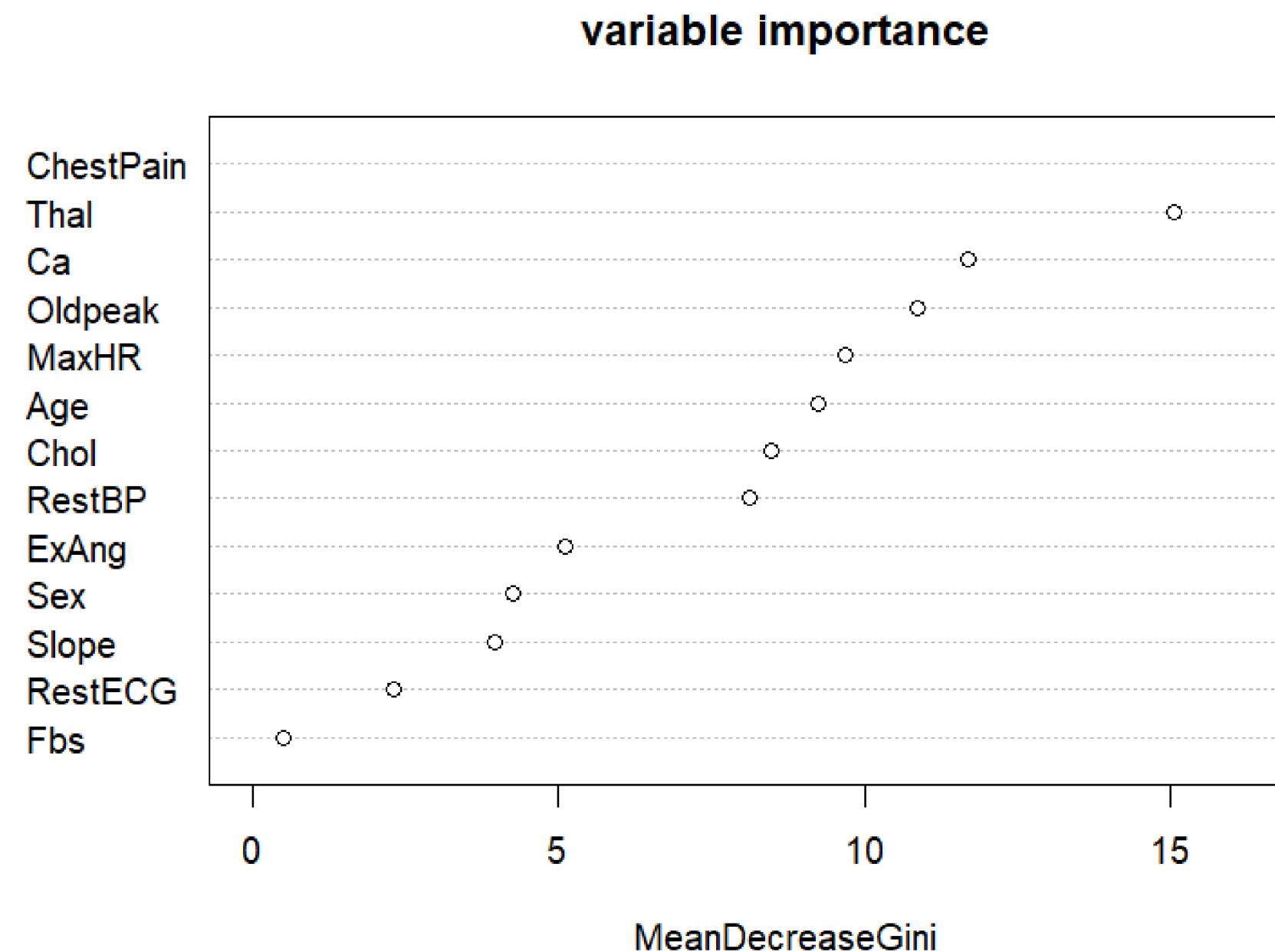
Prevalence : 0.4024

Detection Rate : 0.2927

Detection Prevalence : 0.3780

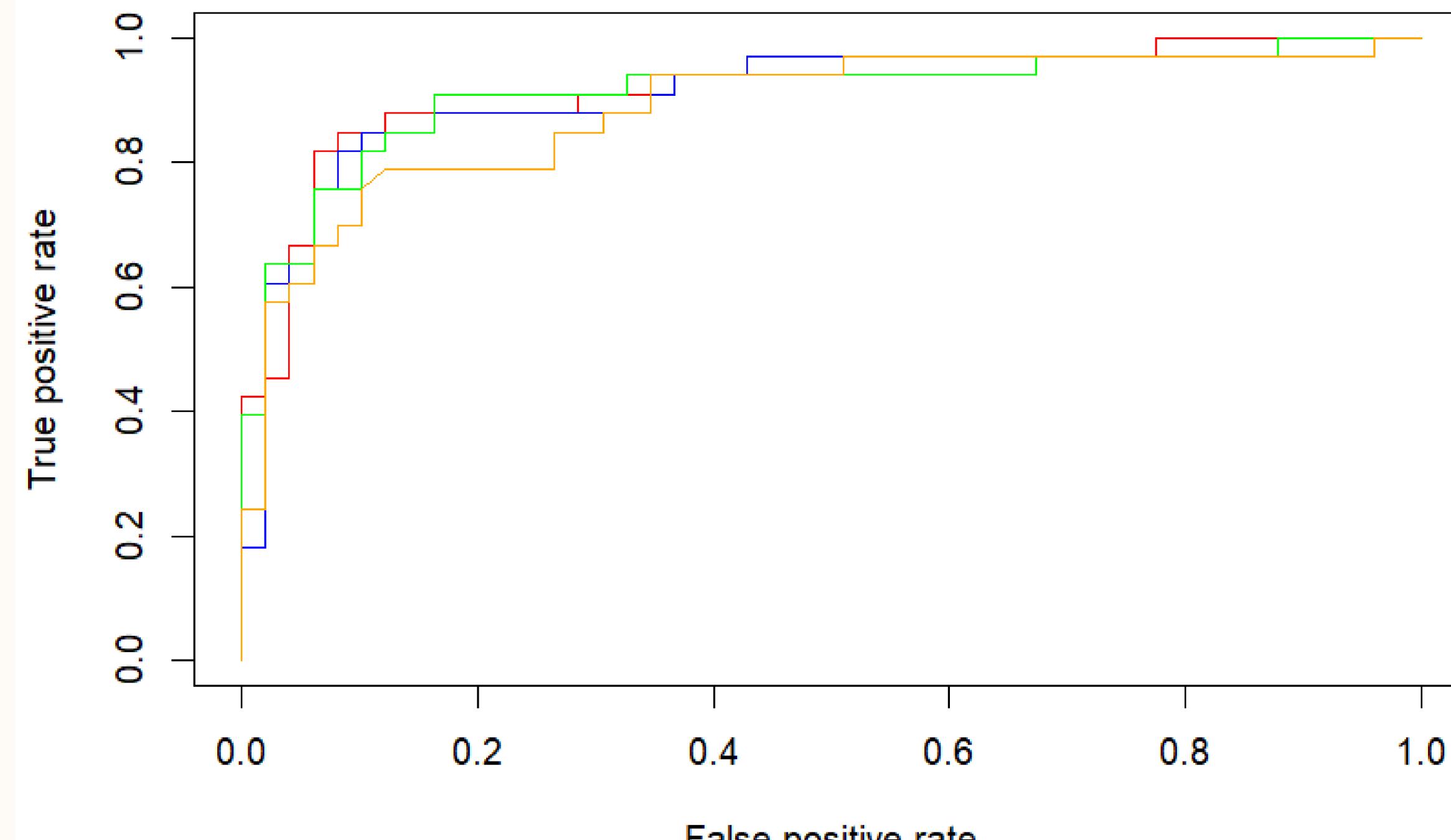
Balanced Accuracy : 0.7922

'Positive' Class : Yes



# ROC curve

Testing Data



SVM linear kernel shows the best performance according to ROC curve



# Conclusion

LDA analysis yields the best result on the Heart Disease dataset.

According to the model coefficients, Chestpain is the most important variable for disease diagnostic.

**THANK YOU FOR LISTENING**