Challenge-2

Operations: It is a Binary classification challenge for three training dataset. To obtain the classification 4 approach was used:

1. Logistic Regression
2. Linear Discriminant Analysis(LDA)
3. Support Vector Machine(SVM)
4. Random Forrest Classifier

Data Sets:

Training Dataset: ADCNtrain.csv, ADMCItrain.csv, MCICNtrain.csv

Testing Dataset: ADCNtest.csv, ADMCItest.csv, MCICNtest.csv

Data Analysis and Feature Selection:

These training sets contain a lot of features. To obtain the most important features Boruta algorithm was used for 2000 iterations. After that, total 45 very important features were found for ADCN challenge, 8 for ADMC challenge, and 11 for the MCICN challenge.
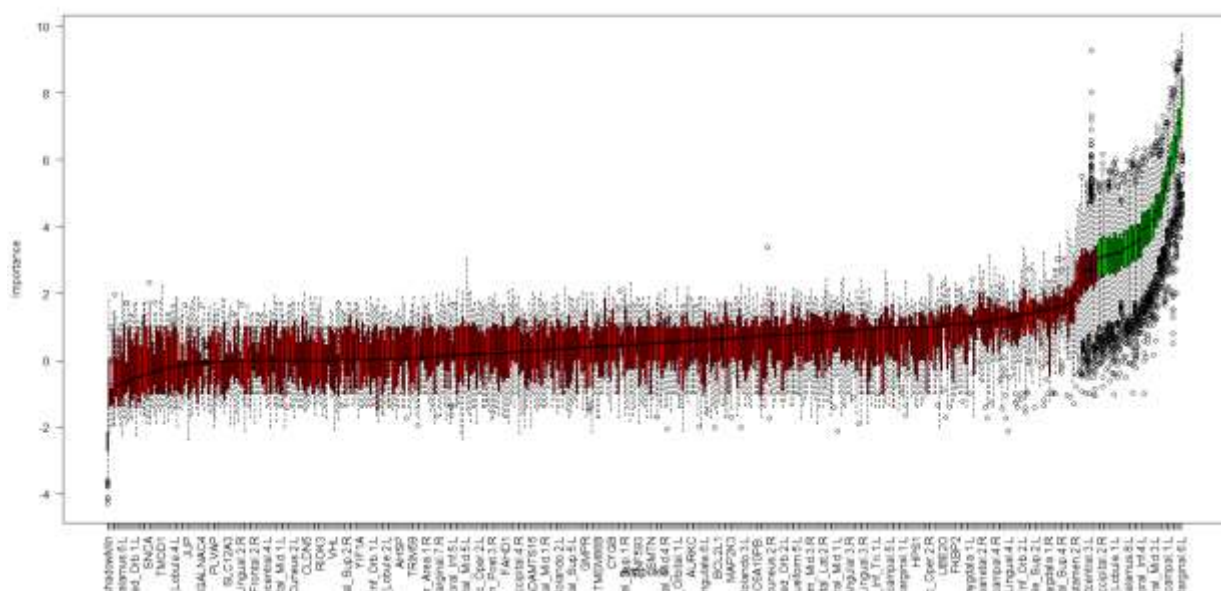

Fig-1: Feature Selection for ADCN (Green = Important Features, Red = Unimportant Features)
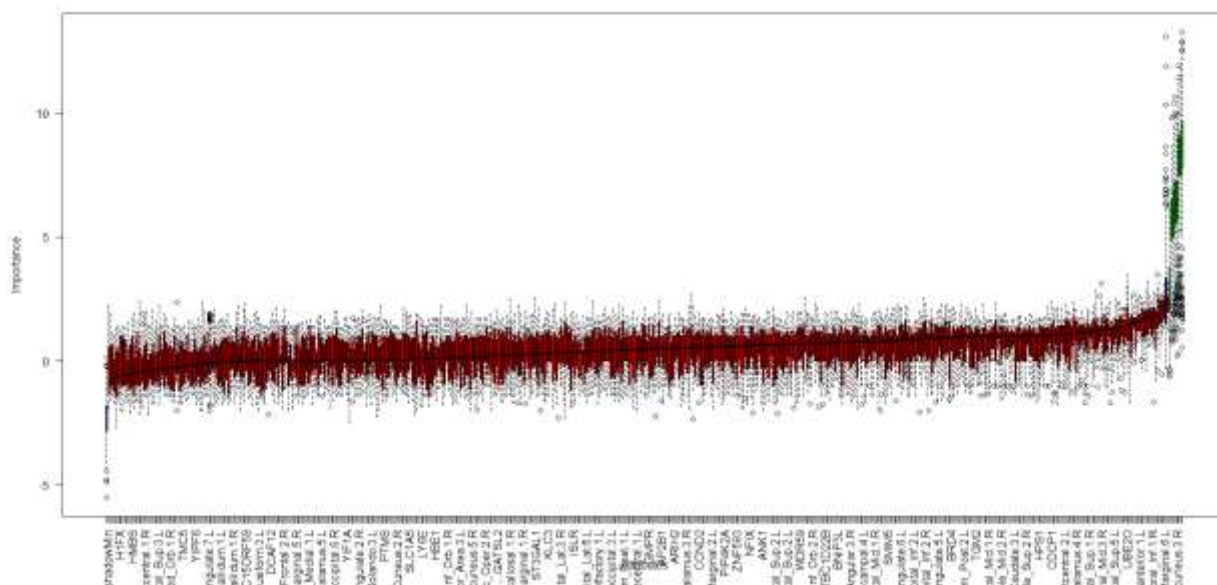
Fig-2: Feature Selection for ADMCI (Green = Important Features, Red = Unimportant Features)
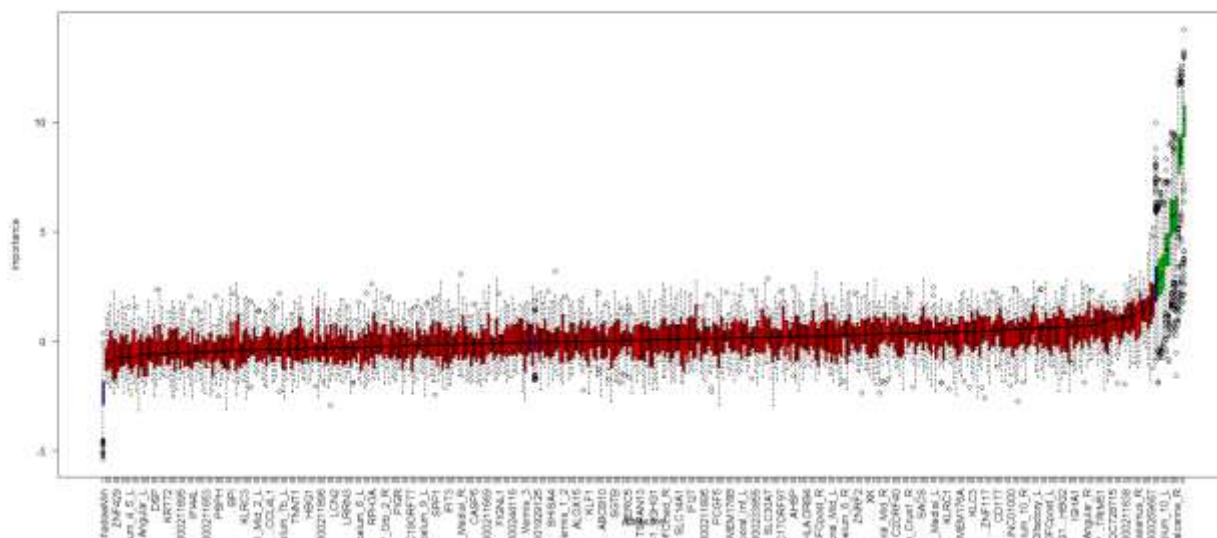

Fig-3: Feature Selection for MCICN (Green = Important Features, Red = Unimportant Features)

Approach: The main training datasets were splited into two parts. The training data and the validation data with the the 80% and 20% data from the main training data available. Then, after training the models, the model was used to predict the target variable in the test dataset.

Results: After training each model, a confusion matrix was generated. From that confusion matrix some performance parameters were calculated.

Logistic Regression	Acc	Sens	Spec	Prec	F1	AUC	MCC
AD vs CN	0.925	0.6	0.9714	0.75	0.66667		
AD vs MCI	0.9342	0	1	NA	NA		
MCI vs CN	0.6757	0.45	0.8028	0.5625	0.5		

Table-1: Performance Table for Logistic Regression.

Linear Discriminant Analysis	Acc	Sens	Spec	Prec	F1	AUC	MCC
AD vs CN	0.1	0.4	0.05714	0.05714	0.099995		
AD vs MCI	0.0658	1	0	0.06579	0.123458		
MCI vs CN	0.2973	0.55	0.1549	0.2683	0.36066		

Table-2: Performance Table for Linear Discriminant Analysis.

Support Vector Machine	Acc	Sens	Spec	Prec	F1	AUC	MCC
AD vs CN	0.875	0.6	0.9143	0.5	0.545454	0.757	
AD vs MCI	0.9342	0	1	NA	NA	0.5	
MCI vs CN	0.7027	0.4250	0.8592	0.6396	0.5106	0.5	

Table-3: Performance Table for Linear Support Vector Machine.

Random Forest	Acc	Sens	Spec	Prec	F1	AUC	MCC
AD vs CN	0.9	0.4	0.9714	0.6667	0.5	0.5	
AD vs MCI	0.9342	0	1	NA	NA	0.5	
MCI vs CN	0.7117	0.4750	0.8451	0.6333	0.5428	0.5	

Table-4: Performance Table for Linear Random Forest.