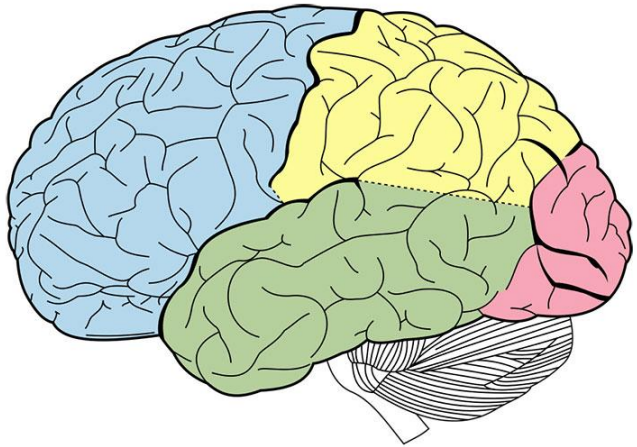
Alzheimer's Disease Classification Using MRIs and Gene Expression Data



Presented By: Abdelrahman HABIB

Introduction

- 60% to 70% of elderly adults with progressive cognitive impairment have Alzheimer's disease (AD) [1].

- There are 3 (macro)-stages for AD:
  - CTL (Controls): no deficit.
  - MCI (Mild Cognitive Impairment): few defects.
  - AD (Alzheimer's Disease): dementia.

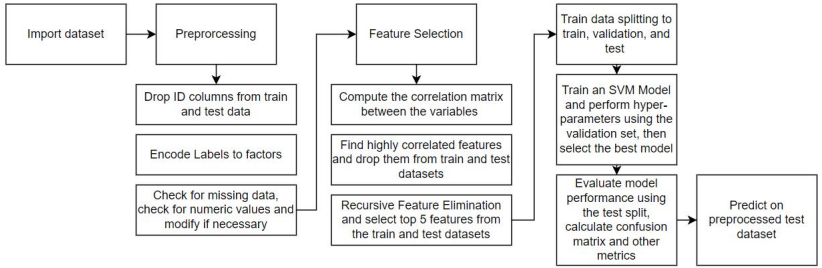
- The objective of this challenge is to build three binary classification solutions to three different datasets, to classify between different stages on data obtained from MRI and Gene Expression data.

[1] Cummings JL, Cole G. Alzheimer Disease. *JAMA*. 2002;287(18):2335–2338. doi:10.1001/jama.287.18.2335

Challenge 1: AD vs CTL

- Dataset has been processed and prepared splitted into train, validation and testing (validation and testing to evaluate on training data).
- Top 5 feature has been selected with both correlation and Recursive Feature Elimination.
- SVM model has been used, hyper-tuned and validated using 10-fold cross-validation. The validation and prediction was using the best model.

Approach



Selected Features `LDHB, DNAJC7, NACA, Caudate_R, Thalamus_R`

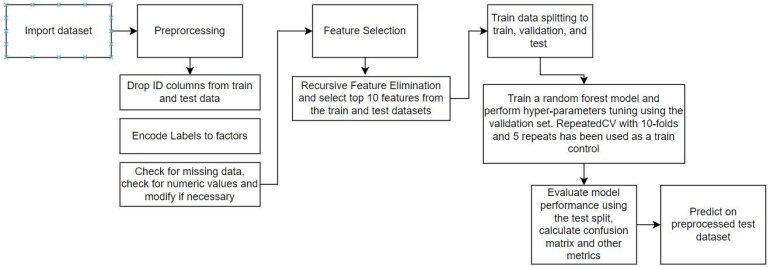
Performance on the training dataset

	Accuracy	Sensitivity	Prec	F1	AUC	MCC	BA
AD vs CTL	0.9167	1	1	0.90909	0.9167	0.8451543	0.9166667

Challenge 2: AD vs MCI

- Various feature selection approaches has been explored for this dataset. Best result has been obtained using Recursive Feature Elimination to select top 10 features.
- Other feature selection approaches like Boruta algorithm has been explored, however, it obtained very low evaluation results.
- Selected features on the train dataset has been trained using a random forest model with repeated Cross-Validation of 10-folds and 5 repeats, with tuning hyper-parameters.

Approach



Selected Features `Right.Middle.Temporal.Gyrus, Left.Angular.Gyrus, Left.Middle.Temporal.Gyrus, Right.Angular.Gyrus, Right.Inferior.Temporal.Gyrus, Right.Parahippocampal.Gyrus, Left.Hippocampus, Left.Supramarginal.Gyrus, Right.Hippocampus, PIP5K2A`

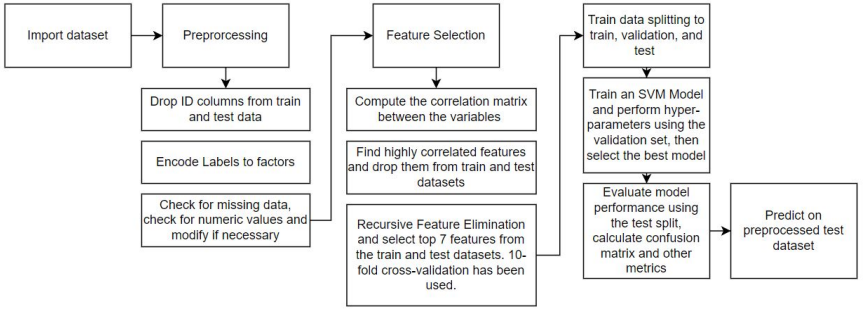
Performance on the training dataset

	Accuracy	Sensitivity	Prec	F1	AUC	MCC	BA
AD vs MCI	0.76	0.6667	0.8	0.727272727	0.756410256	0.52297636	0.756410256

Challenge 3: MCI vs CTL

- Various feature selection approaches has been explored for this dataset. Best result has been obtained using both correlation features elimination and Recursive Feature Elimination to select top 10 features.
- Other feature selection approaches like Lasso regularization has been explored, and couldn't obtain the best features based on the evaluation results.
- Both GBM and SVM models has been explored. Selected features on the train dataset has been trained using an SVM model with repeated cross-validation of 10-folds and 3 repeats, with tuning hyper-parameters.

Approach



Selected Features `SHFM1, LOC728499, S100P, Hippocampus_L, SELPLG, ITPRIPL2, IKZF1`

Performance on the training dataset

	Accuracy	Sensitivity	Prec	F1	AUC	MCC	BA
MCI vs CTL	0.8333	0.8824	0.8823529	0.8823529	0.7983	0.5966387	0.7983193

Conclusion

- Among the three challenges, best results were obtained in the ADvsCTL dataset using SVM model and both correlation and RFE feature selection approaches.
- SVM model has shown promising results on other datasets using different feature selection approaches.
- Other models as random forest and GBM has been explored on other datasets, with obtaining close results to the expected evaluation metrics.