Detection of dementia using machine learning techniques

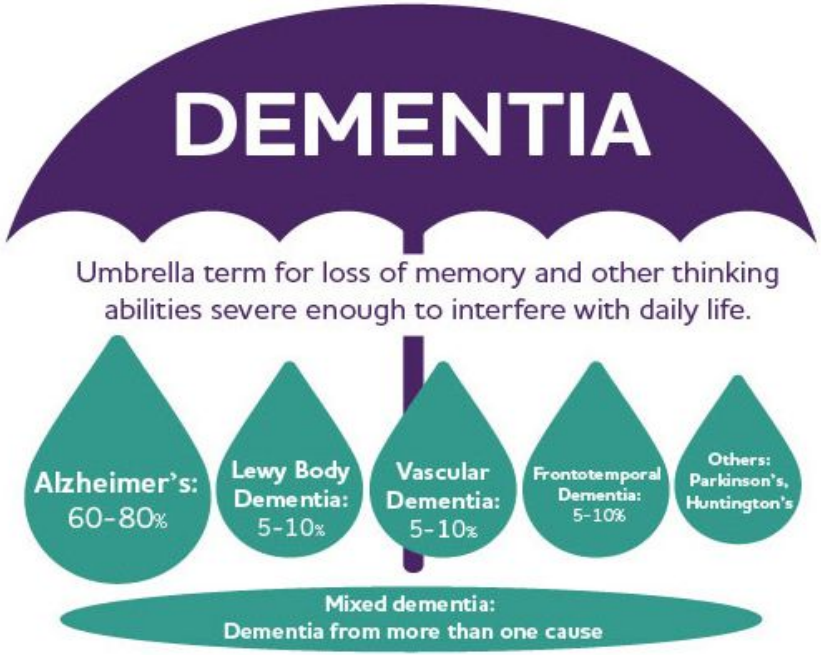
BY: NIDHI ZARE 2017A7PS0139G
JASLEEN 2017A7PS0077G

INTRODUCTION

What is Dementia?

A general term for loss of memory, language, problem-solving and other thinking abilities that are severe enough to interfere with daily life.

Alzheimer's is the most common cause of dementia.

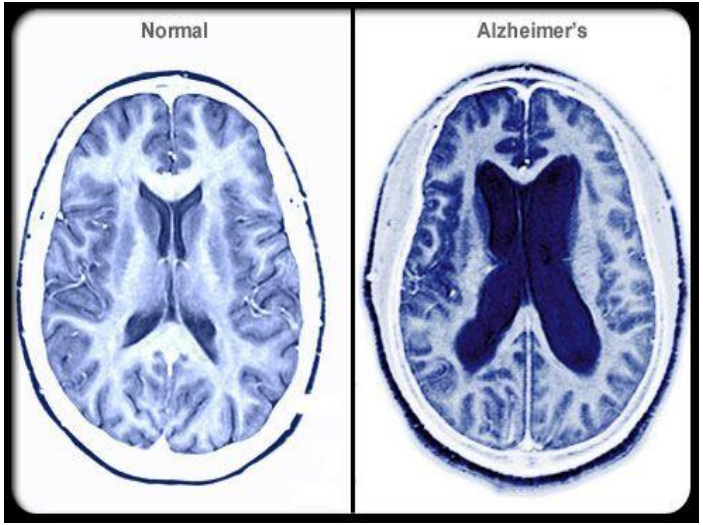


Detection of dementia

- Patients suffering from Dementia have **shrinkage** of brain tissues and change in their **speech patterns**.
- Thus dementia can be detected in two major ways:

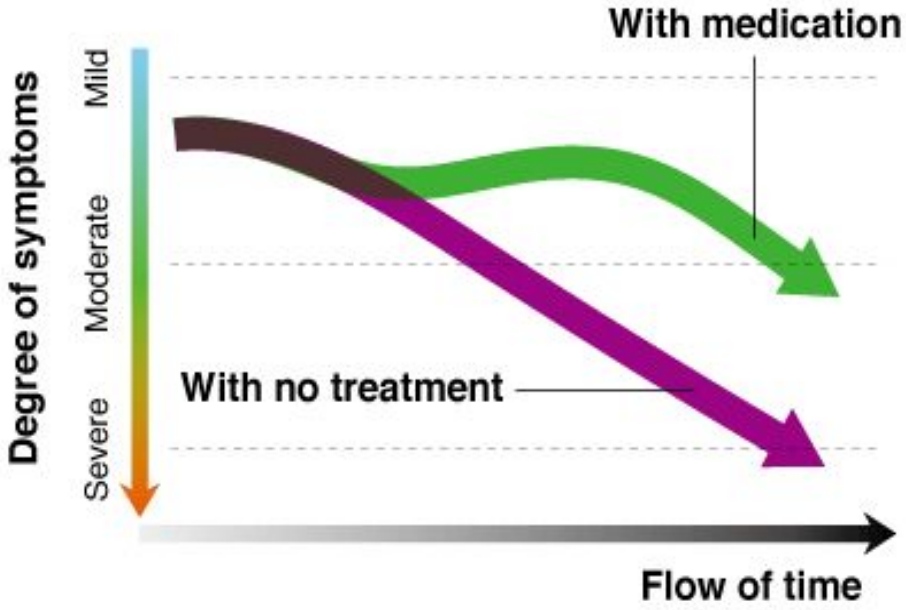
Speech patterns

MRI images



Detection of dementia

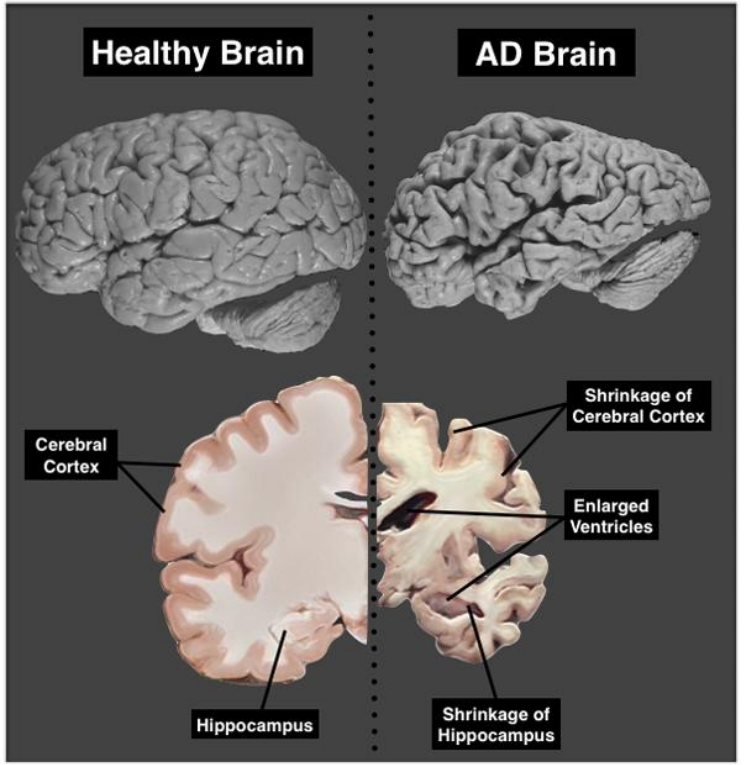
- The current treatment cannot stop the disease but rather slow down the damage caused.
- Detecting at early stage can reduce damage to large extent.



USING MRI IMAGES

MRI images of Alzheimer Disease(AD) patients show both **local and generalized shrinkage of brain tissues**. There are various other factors like age,education and socioeconomic status that determines the course of the disease.

We use Machine Learning techniques to detect dementia even at **mild damage** so that immediate treatment can be given.



DATASET

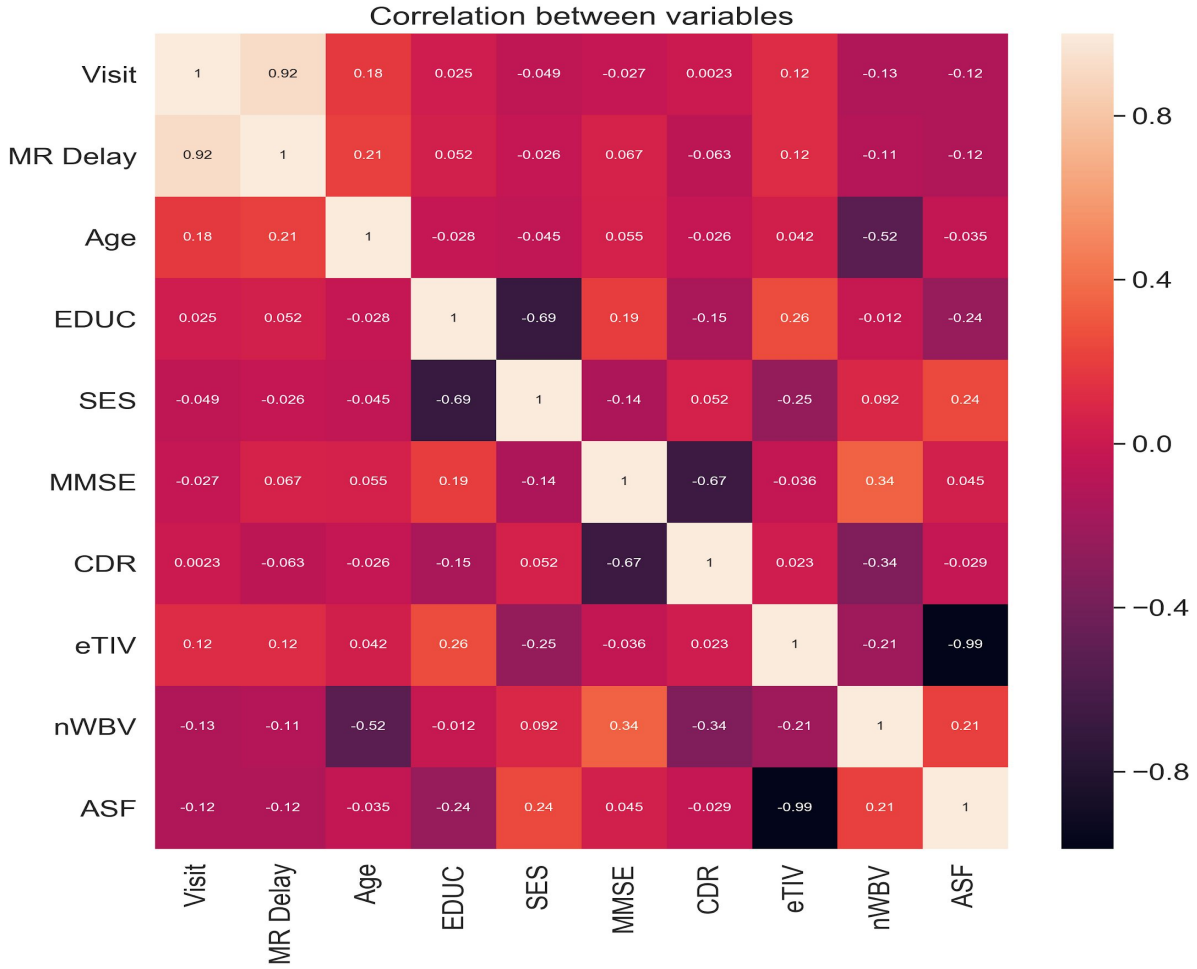
1. We used MRI related dataset from Open Access Series of Imaging Studies **(OASIS)**
2. It is longitudinal MRI data consisting of 150 people aged between 60 to 96.
3. Everyone is right handed.
4. Some patients were under **NonDemented** category at first visit but later on were found **Demented**. These fall under category **Converted**.

COLUMN DESCRIPTION

Column Name	Description
EDUC	Years of education
SES	Socioeconomic Status
MMSE	Mini Mental State Examination
CDR	Clinical Dementia Rating
eTIV	Estimated Total Intracranial Volume
nWBV	Normalized Whole Brain Volume
ASF	Atlas Scaling Factor

SAMPLE DATA

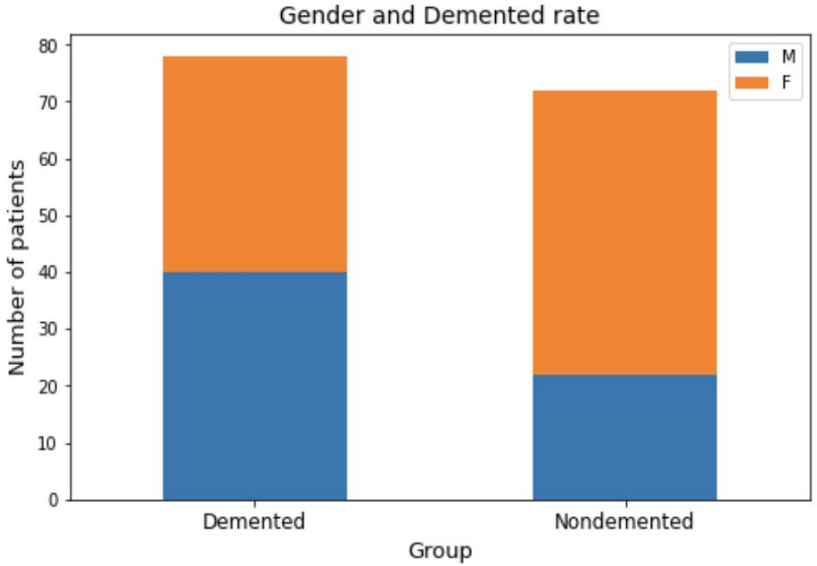
	Subject ID	MRI ID	Group	Visit	MR Delay	M/F	Hand	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
0	OAS2_0001	OAS2_0001_MR1	Nondemented	1	0	M	R	87	14	2.0	27.0	0.0	1987	0.696	0.883
1	OAS2_0001	OAS2_0001_MR2	Nondemented	2	457	M	R	88	14	2.0	30.0	0.0	2004	0.681	0.876
2	OAS2_0002	OAS2_0002_MR1	Demented	1	0	M	R	75	12	NaN	23.0	0.5	1678	0.736	1.046
3	OAS2_0002	OAS2_0002_MR2	Demented	2	560	M	R	76	12	NaN	28.0	0.5	1738	0.713	1.010
4	OAS2_0002	OAS2_0002_MR3	Demented	3	1895	M	R	80	12	NaN	22.0	0.5	1698	0.701	1.034


Correlation between variables
Visit 1 0.92 0.18 0.025 -0.049 -0.027 0.0023 0.12 -0.13 -0.12
MR Delay 0.92 1 0.21 0.052 -0.026 0.067 -0.063 0.12 -0.11 -0.12
Age 0.18 0.21 1 -0.028 -0.045 0.055 -0.026 0.042 -0.52 -0.035
EDUC 0.025 0.052 -0.028 1 -0.69 0.19 -0.15 0.26 -0.012 -0.24
SES -0.049 -0.026 -0.045 -0.69 1 -0.14 0.052 -0.25 0.092 0.24
MMSE -0.027 0.067 0.055 0.19 -0.14 1 -0.67 -0.036 0.34 0.045
CDR 0.0023 -0.063 -0.026 -0.15 0.052 -0.67 1 0.023 -0.34 -0.029
eTIV 0.12 0.12 0.042 0.26 -0.25 -0.036 0.023 1 -0.21 -0.99
nWBV -0.13 -0.11 -0.52 -0.012 0.092 0.34 -0.34 -0.21 1 0.21
ASF -0.12 -0.12 -0.035 -0.24 0.24 0.045 -0.029 -0.99 0.21 1
Visit
MR Delay
Age
EDUC
SES
MMSE
CDR
eTIV
nWBV
ASF
0.8
0.4
0.0
−0.4
−0.8

Dementia vs Gender

	M	F
Demented	40	38
Nondemented	22	50

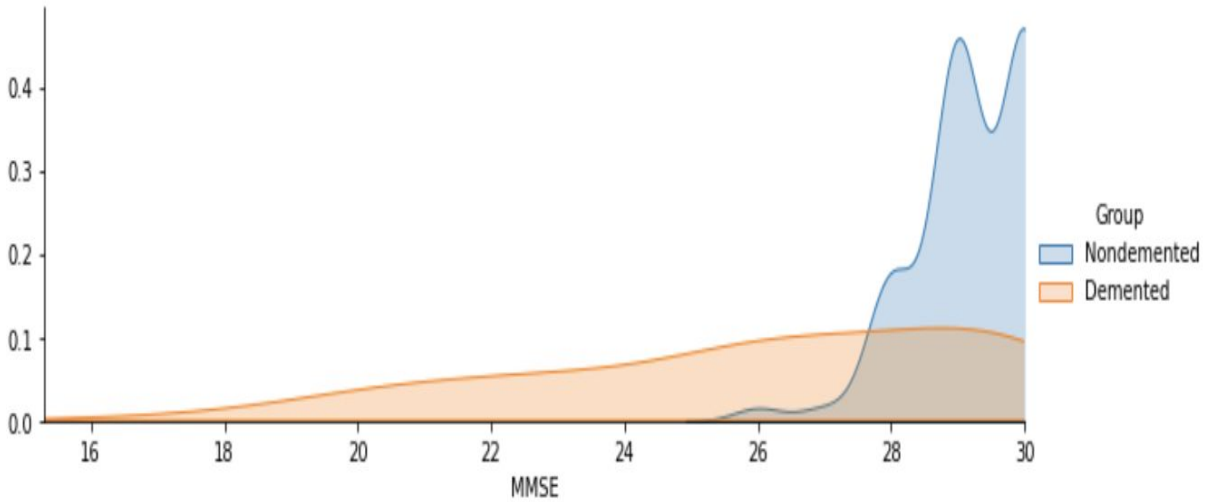
More number of males have dementia than females.



Dementia vs MMSE

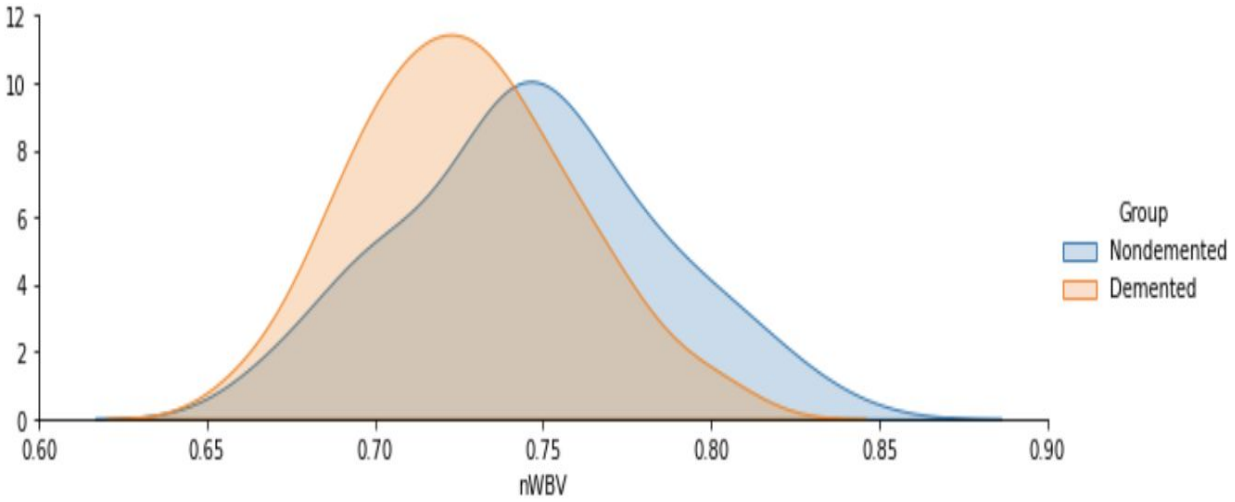
- Plot between MMSE (Mini Mental State Examination) and Dementia

- Nondemented has higher test result ranging from **25 to 30.**



Dementia vs nWBV

- Plot between nWBV (Normalized Whole Brain Volume) and Dementia
- Nondemented group has **higher brain volume** than Demented group. This is true because the disease causes the shrinking brain tissue.



Dementia vs Age

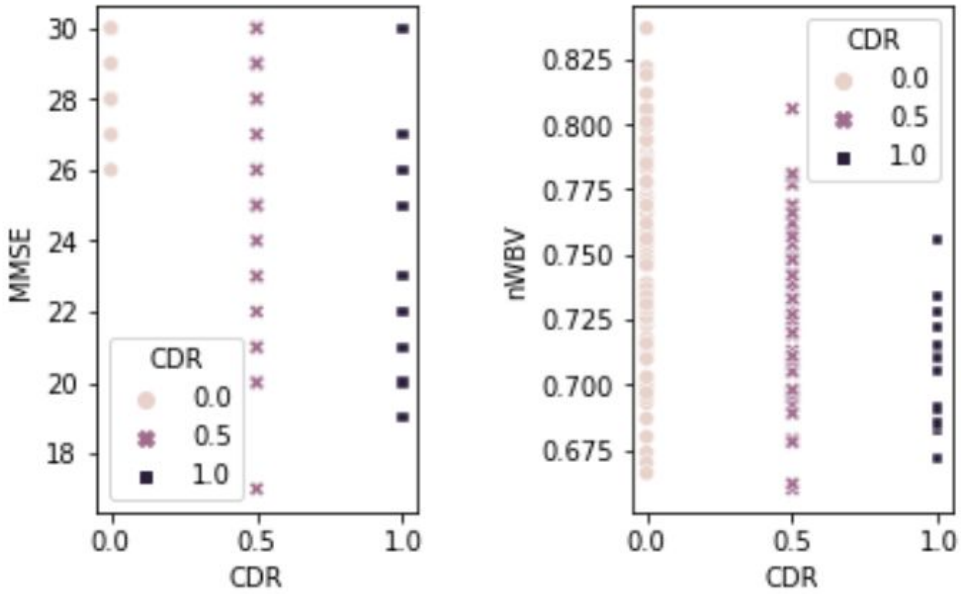
- There is a higher concentration of **70-80 years** old in the Demented patient group than the Nondemented group.
- We guess patients who suffered from the disease have lower survival rate and hence there are very few over 90 years old.

CDR - Clinical Dementia Rating

Relation between variables and **CDR (Clinical Dementia Rating)**

- While the Dementia/Non-Dementia classification had relation with age and gender, this direct connection is not found with CDR.
- No obvious connection between Education Level/Socio Economic Status and CDR as well.
- While the MMS examination results of objects not diagnosed with Dementia concentrate near 27-30 point rate, MMSE results of objects diagnosed with Dementia seems to be more spreaded. The objects had the highest MMSE score but still have Clinical Dementia Rating of 0.5 or 1. No obvious connection between Estimated total intracranial volume and Dementia Diagnosis.
- Normalized whole-brain volume seems to be more spreaded for objects with CDR = 0 and narrows as CDR grows up. No obvious connection between Atlas scaling factor and Dementia Diagnosis.

CDR vs MMSE and NWBV



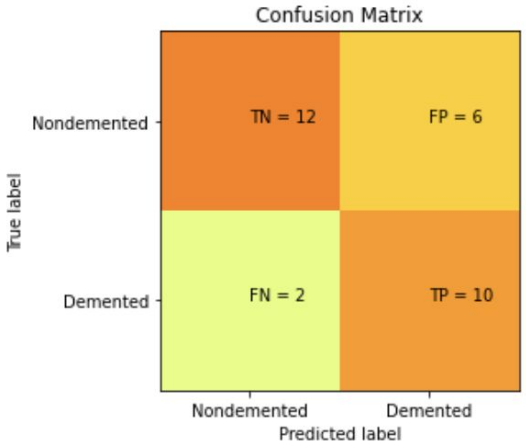
MACHINE LEARNING MODELS

Random Forest

- It is an ensemble learning method that constructs multiple decision trees and outputs the class that is the mode of the classes or **mean/average prediction** of the individual trees.
- Classifying Dementia vs Non-Dementia prediction using random forest classifier.
- **Accuracy = 73%**

Classification Report:

	precision	recall	f1-score	support
0	0.86	0.67	0.75	18
1	0.62	0.83	0.71	12
accuracy			0.73	30
macro avg	0.74	0.75	0.73	30
weighted avg	0.76	0.73	0.74	30

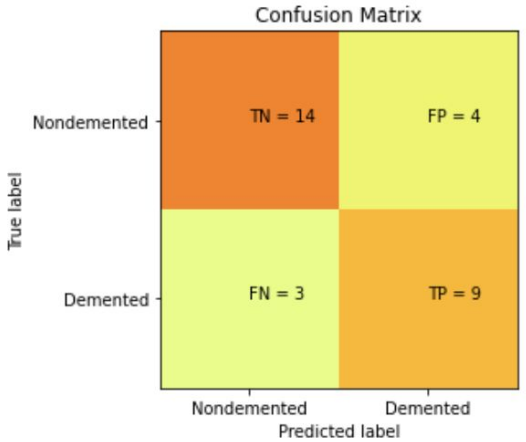


Support Vector Machine

- Support Vector Machine or SVM is a **supervised** machine learning model which can solve linear as well as non-linear problems.
- Classifying Dementia vs Non-Dementia prediction using support vector machine(SVM)
- **Accuracy = 76%**
- Nondemented = 0,Demented = 1

Classification Report:

	precision	recall	f1-score	support
0	0.82	0.78	0.80	18
1	0.69	0.75	0.72	12
accuracy			0.77	30
macro avg	0.76	0.76	0.76	30
weighted avg	0.77	0.77	0.77	30

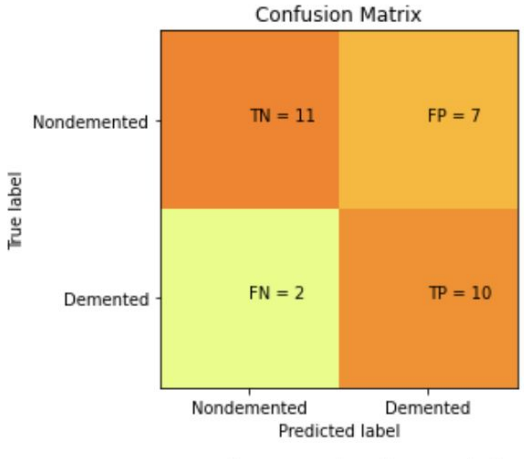


XGBOOST

- XGBOOST is a decision tree based ensemble machine learning model.
- It uses **gradient boosting** framework.
- Classifying Dementia vs Non-Dementia prediction using XGB Classifier.
- **Accuracy = 70%**
- Nondemented = 0,Demented = 1

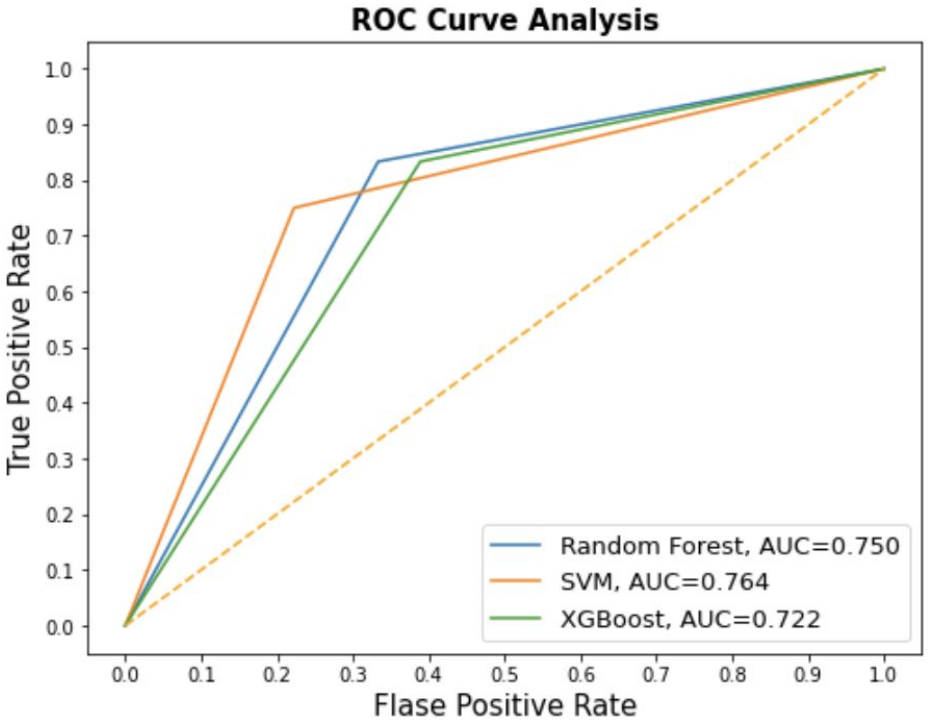
Classification Report:

	precision	recall	f1-score	support
0	0.85	0.61	0.71	18
1	0.59	0.83	0.69	12
accuracy			0.70	30
macro avg	0.72	0.72	0.70	30
weighted avg	0.74	0.70	0.70	30



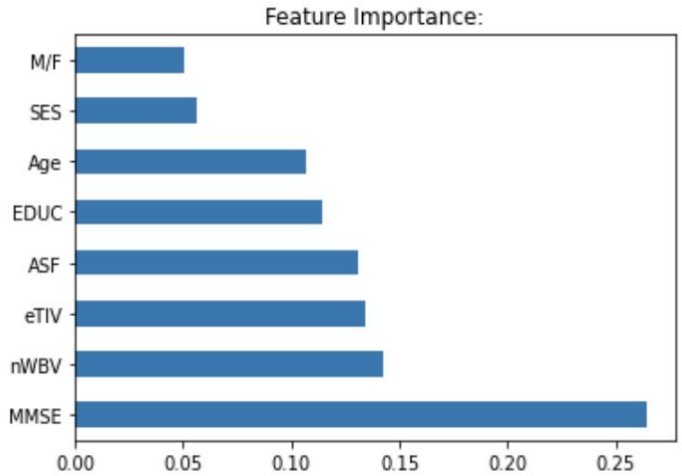
RESULTS

- An **ROC curve** (receiver operating characteristic curve) is a **graph** showing the performance of a classification model at all classification thresholds. This **curve** plots 2 parameters: True Positive Rate & False Positive Rate.
- Area under ROC curve (AUC) is between **0.5 to 1**, 1 being best classifier.
- We plot AUC for all three models to see difference between them.



RESULTS

- Following graph shows relative feature importance in classification of Demented vs Non-Demented.



- We observe that MMSE is the most important feature whereas gender and socio economic status are least important features.

Classifying CDR

- CDR is **Clinical Dementia Rating**
- Instead of binary classification between Dementia and Non-Dementia we can classify into CDR classes which will give the severity of dementia.
- Classification model classifies CDR into 0,0.5 and 1 where:

0: Non Dementia 0.5: Mild Dementia 1: Severe Dementia

MACHINE LEARNING MODELS

- We use the classification based models: **random forest**, **support vector machine** (SVM) and **XGBOOST** to classify CDR on scale 0,0.5 and 1.

- Here, 0 stands for 0 CDR, 1 stands for 0.5 CDR and 2 stands for 1 CDR.

Random Forest

Accuracy = 73%

Confusion Matrix:

```
[[17  5  0]
 [ 0  5  2]
 [ 0  1  0]]
```

Classification Report:

	precision	recall	f1-score	support
0	1.00	0.77	0.87	22
1	0.45	0.71	0.56	7
2	0.00	0.00	0.00	1
accuracy			0.73	30
macro avg	0.48	0.50	0.48	30
weighted avg	0.84	0.73	0.77	30

Support Vector Machine

Accuracy = 80%

Confusion Matrix:

```
[[17  5  0]
 [ 0  7  0]
 [ 0  1  0]]
```

Classification Report:

	precision	recall	f1-score	support
0	1.00	0.77	0.87	22
1	0.54	1.00	0.70	7
2	0.00	0.00	0.00	1
accuracy			0.80	30
macro avg	0.51	0.59	0.52	30
weighted avg	0.86	0.80	0.80	30

XGBOOST

Accuracy = 76%

Confusion Matrix:

```
[[17  5  0]
 [ 0  6  1]
 [ 0  1  0]]
```

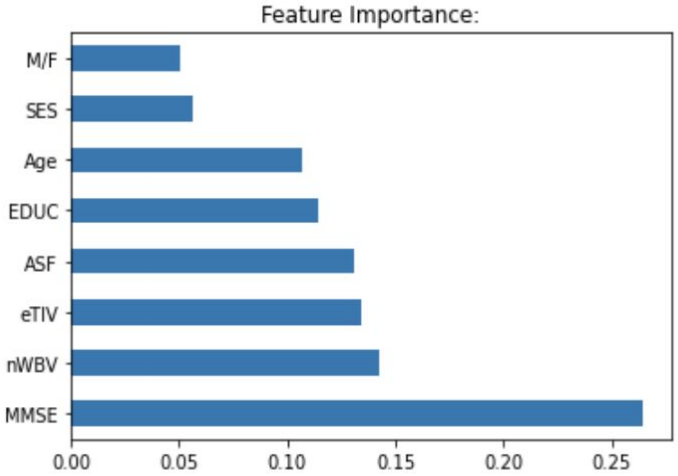
Classification Report:

	precision	recall	f1-score	support
0	1.00	0.77	0.87	22
1	0.50	0.86	0.63	7
2	0.00	0.00	0.00	1
accuracy			0.77	30
macro avg	0.50	0.54	0.50	30
weighted avg	0.85	0.77	0.79	30

RESULTS

- Following graph shows relative feature importance in classification of CDR.

- Similar to Dementia vs Non Dementia classification, we observe that **MMSE** is the most important feature whereas **gender and socio economic status** are least important features.



Future Direction

- Due to non-availability of open data source, we used only MRI images to detect Dementia. The current work can be extended by inculcating speech patterns. A combined model can be created which uses both **MRI images and speech patterns** to detect Dementia and it's severity which can give more accurate results.

- Deep learning techniques couldn't be incorporated due to fewer data points. In case of availability of more data deep learning can be used.


THANK YOU