DETECTION AND PREDICTION OF ALZHEIMER'S DISEASE

[1]Manushri Venkatesh, [2]Abhishek MM, [3]Varshini Rao, [4]Meraj Ahmed,
[1,2,3,4] Student,

[1]Department of Computer Science,
[1] JSSSTU,Karnataka,India

Abstract *:* The topic of this paper is Alzheimer's disease detection and prediction of categories of individuals vulnerable to the disease.There are many methods used to detect Alzheimer's,this study considers machine learning and deep learning techniques.We have used SVM model and CNN model for alzheimer's detection.We have also compared the results obtained and concluded the better method.We have done exploratory data analysis on the ADNI data to get insights of categories of individuals who are more vulnerable to Alzheimer's.

1.INTRODUCTION

The topic of this paper is Alzheimer's disease detection and prediction of categories of individuals vulnerable to the disease.There are many methods used to detect Alzheimer's,this study considers machine learning and deep learning techniques.

2.BACKGROUND

Alzheimer's disease affects around 44 million individuals globally. Alzheimer's disease is a long-term illness. It's a neurologic condition that causes the brain to shrink over time. Alzheimer's disease is the most frequent cause of dementia, which diminishes a person's capacity to live independently and results in a steady deterioration in thinking abilities.

3.DATASETS

3.1. ADNI DATASET

Data consists of 12741 entries and 1907 features. Features include cognitive scores, PET ,MRI ,CSF features, genetic features & Demographic data.Each column represents a feature is a biomarker from the subject at that particular visit, and each row represents data for one specific visit of a subject. The spreadsheet's first columns contain unique identifiers: The RID (roster ID) is a unique identifier for each subject, and the VISCODE (visit code) represents the date and time of the visit.

- The main measures to be predicted: DX
- Cognitive tests: CDRSB, ADAS11, MMSE, RAVLT_immediate
- MRI measures: Hippocampus, WholeBrain, Entorhinal, MidTemp
- PET measures: FDG, AV45
- CSF measures
- Risk factors: APOE4, AGE

Out of 1907 features in our study we use the following features-

Column	Meaning
RID	Participant Roster ID
Hippocampus	UCSF Hippocampus-Hippocampus is a complex brain structure embedded deep into temporal lobe.
Wholebrain	UCSF WholeBrain-Brain volume
Fusiform	UCSF Fusiform-The fusiform gyrus is a large gyrus that spans across the basal surface of the temporal and occipital lobes of the cerebral hemispheres
MidTemp	UCSF Med Temp-Middle temporal gyrus is a gyrus in the brain on the Temporal lobe
Ventricles	UCSF Ventricles-The ventricles of the brain are a communicating network of cavities filled with cerebrospinal fluid (CSF) and located within the brain parenchyma
Entrohinal	UCSF Entorhinal-The entorhinal cortex (EC) is a locality of the brain's allocortex, situated within the medial lobe, whose functions embrace being a widespread network hub for memory, navigation, and therefore the perception of your time.
DX	Dx status(AD/NL)
DXCHANGE	1=Stable:NL to NL 2=Stable:MCI to MCI 3=Stable:AD to AD, 4=Conv:NL to MCI, 5=Conv:MCI to AD, 6=Conv:NL to AD, 7=Rev:MCI to NL, 8=Rev:AD to MCI, 9=Rev:AD to NL, -1=Not available
VISCODE	Visit code

Table 1.0-MRI measures

Column	Meaning
Age	Age of the participant in years
APOE4	Number of alleles of the gene in participant
MMSE	Mini Mental State Examination score
PTEDUCAT	Number of years of formal education

Tabe 1.1-Risk factors

3.2. Kaggle MRI Image Dataset

The data consists of segmented MRI images. It contains 4 classes of image in testing and training sets ..

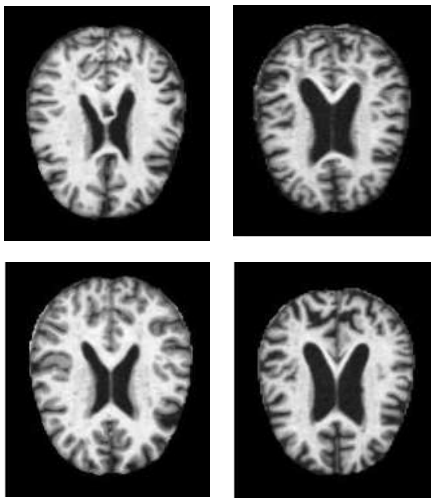


Fig 1-Segmented MRI images

- Mild Demented
- Moderate Demented
- Non Demented
- Very Mild Demented

4. METHODOLOGY

4.1.Detection

4.1.0.SVM Model

Initially,we have used a machine learning model.We have adoped the SVM(Support Vector Machine) model for the Alzheimer's detection.When compared to other classifiers like logistic regression and decision trees, SVM has a relatively high accuracy. The classifier uses the hyperplane with the most margin to separate data points. As a result, an SVM Classifier is sometimes referred to as a discriminative classifier.

Steps-

- Read Data Using Pandas
- Filter required columns –Table 1.0
- The dataset contains data of the same patients visiting multiple times.
  Only the data from the first visit is considered in our analysis.
- Cleaning of NaN values.
  Selecting features required for training and the target value.
- Test-Train Split A 90/10 ratio(90 % train) is used for the test-train split.
- Fit model & Predict A support vector classifier (SVC) model is used for the purpose.

```
clf = SVC(C=2, kernel="linear").fit(X_train,y_train)
```

Fig 2-Code snippet of fit model

4.1.1CNN Model

A detailed model using CNN is built in order to get greater accuracy compared to the SVM model.A Convolutional Neural Network (CNN) is a Deep Learning algorithm that can take an image as input, assign priority (weights and biases) to distinct features of the image, and distinguish one from the other. A ConvNet's architecture is similar to the connectivity network of Neurons in the Human Brain. A number of similar fields can be stacked on top of each other to span the full visual field.

The following steps are performed on the MRI image dataset from Kaggle

- Import datasets and necessary libraries- The main module used to built CNN module is Keras.Other libraries include imblearn,pandas,seaborn,matplotlib,sklearn
- Perform Image Augmentation-Set parameters such ass zoom range,brightness range and horizontal flip.

- Oversampling of imbalanced data-In a classification challenge, if one or more of the classes for which we wish to forecast have an unusually low number of samples, you may have an unbalanced classes problem in your data.

The dataset details are as follows:

VeryMildDemented-2240 mri images

NonDemented-3200 mri images

ModeratelyDemented-64 mri images

MildDemented-896 mri images

SMOTE is one of the most widely utilised oversampling approaches for resolving the problem of imbalance.It evens o ut the distribution of classes by replicating minority class samples at random.

- Build model

Model main functions-

1.Convolution Block Function

Fig 3-Convolution block function components

2.Dense Block function

Fig 4-Dense block function components.

3.Custom Callback function

Stops training model when the val_acc reaches threshold.

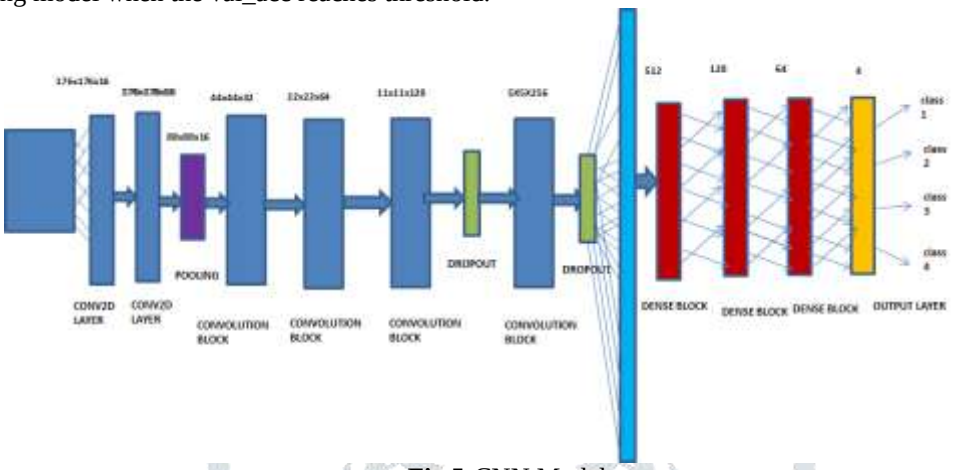


Fig 5:CNN Model

• Evaluation

The performance of the model evaluated based on the accuracy obtained and the confusion matrix method.

4.2.Risk Prediction

In order to find the categories of people who are more vulnerable to Alzheimer's,we used the ADNI dataset.We have used exploratory data analysis in order to analyse the ADNI and identify the catgories more prone to Alzheimer's from the graph plot.We have used kdeplots for plotting the data.From the graph plotted we can conclude the categories the individuals vulnerable to Alzheimer's disease.

5.Results

The SVM Model results in a accuracy of 84.05%.
The CNN Model resulted in a validation accuracy of 95%.

	precision	recall	f1-score	support
Dementia	0.83	0.65	0.73	23
NL	0.84	0.93	0.89	46
accuracy			0.84	69
macro avg	0.84	0.79	0.81	69
weighted avg	0.84	0.84	0.83	69

Fig 6-Classification report of SVM Model

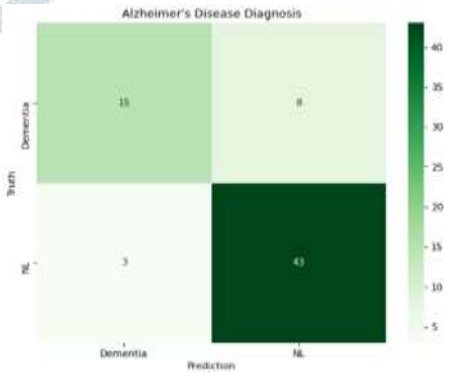


Fig 7-Confusion matrix of SVM Model

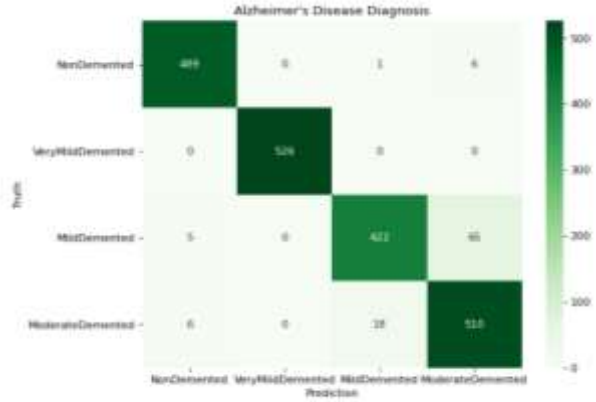
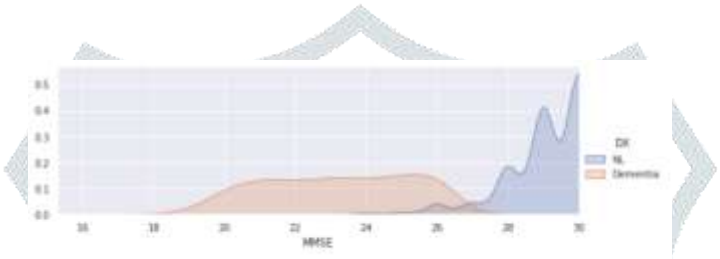


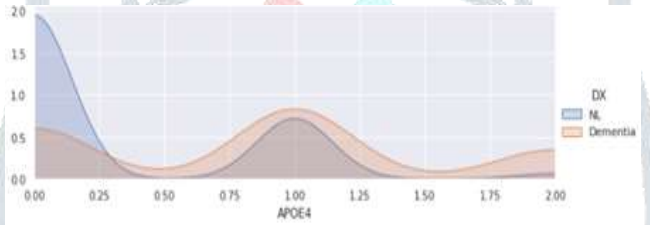
Fig 8-Confusion matrix of CNN model

From the results of the study we conclude that CNN model could be a better option to detect alzheimer's.

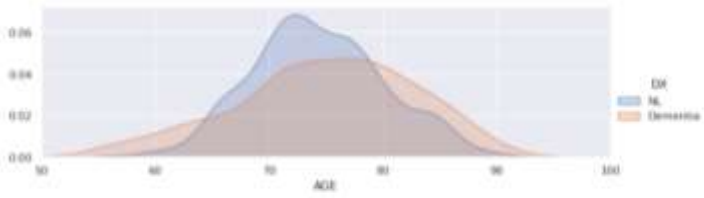
RISK PREDICION CONCLUSIONS



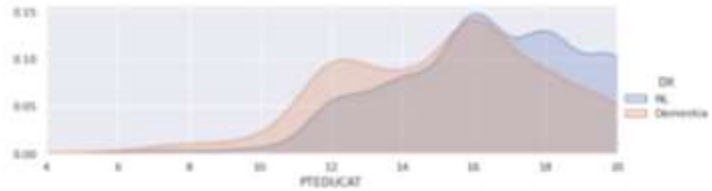
Graph 1-Kdeplot of MMSEscore in individuals **vs** density with a DX value



Graph 2-Kdeplot of APOE4 gene in individuals vs density with a DX value



Graph 3-Kdeplot of age of individuals vs density with a DX value



Graph 4-Kdeplot of education of individuals vs density with a DX value

From the above plotted graphs we could conclude-
Graph 1-More people in the mmse score range of 18-26 have dementia .More People with high MMSE score that is more than 27 are normal.Low MMSE Score is linked to alzheimer's
Graph 2-More people with APOE4=1&2 are demented and APOE4=0 are non demented.
Graph 3-More people in the age groups of >=80 are demented than normal.
Graph 4-More people with EDUC<=15 are demented more than normal.Hence low level of eduation may be linked to dementia.

6.Conclusion

We conclude that CNN model could be a good method in order to detect the individuals with alzheimer's .The model detects the disease with a good accuracy and hence could be quite reliable to be used in real time.From this study we also conclude some of the most crucial factors related to the Alzheimer's risk.This could help the individuals to be self aware of their condition and take care of not having additional complications.We have also observed the following are linked to Alzheimer's disease-

- Low MMSE score
- Old age
- APOE 4 gene
- Education level

7.Reference

[1] https://braininformatics.springeropen.com/articles/10.1186/s40708-020-00112-2
[2] https://www.frontiersin.org/articles/10.3389/fnagi.2019.00220/full
[3] https://tadpole.grand-challenge.org/Data/
[4]https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447
[5] https://www.physio-pedia.com/Mini-Mental_State_Examination
[6]https://pubmed.ncbi.nlm.nih.gov/18579155
[7] http://adni.loni.usc.edu/
[8] https://www.kaggle.com/tourist55/alzheimers-dataset-4-class-of-images