International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Alzheimer Disease Classification using Machine Learning

Gousiya Begum, A. Manisha

Abstract: ***Alzheimer's disease is the most popular and persuading dementia that affects our memory power, reasoning and deportment. Symptoms rise up slowly and worsen with time, becoming an obstacle in doing our routine tasks. Alzheimer is not conventional wedge of aging. The substantial and known risk factor is up surging age. The prevalence of AD is depicted to be around 5% after an age of 65 years and took a leap of 30% for people of 85 years old in developed countries [1]. In this project we proposed a detection and classification technique using Random Forest(RF) and Support Vector Machine(SVM) algorithms on the oasis longitudinal data set and compare their respective accuracies to come to a conclusion that which algorithm best suits for this detection and classification.***
Paper Setup must be in A4 size with Margin: Top 0.7",

Keywords: Alzheimer, Random Forest, SVM.

I. INTRODUCTION

Alzheimer's disease is now predominant everywhere especially in US and other western countries. From 2000 to 2013, deaths due to cardiac disease, prostate cancer and stroke declined by 14%, 11% and 23%, respectively, whereas deaths due to AD inflated by 71% [2]. It is a specific kind of dementia where in the person's mental ability is deteriorated and he loses his thinking ability even affecting his ability to perform daily routine activities. Alzheimer's disease detection is not yet explicit till the patient arrives at the moderate phase of AD. For precise medicinal evaluation of AD, corporeal and neurological examinations, Folstein's test and patient's comprehensive record is needed [1]. Early diagnosis is critical to prevent its progression**.**

In the below figure, (a) is MRI of a non demented person, (b) is the MRI of person suffering from very mild dementia, (c) is the MRI of a person suffering from mild dementia and (d) is the MRI of the person suffering from moderate dementia. Investigations reveal that Alzheimer cannot be detected until the person reaches moderate stage. Hence there can be no prevention but only cure once it is detected**.**

Machine learning has now gained significant importance in the fields of classification, prediction, analysis tasks.

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
Gousiya Begum*, CSE Department, Mahatma Gandhi Institute of Technology, Hyderabad, India, Email: gousiyabegum@gmail.com
A Sai Manisha, CSE Department, Mahatma Gandhi Institute of Technology, Hyderabad, India, Email: manishaannam22@gmail.com

We here use the SVM and Random forest algorithm for the classification of OASIS dataset [5] prepared by Dr. Randy Buckner. CNN does well at visual images recognition where as SVM is generally used in classification problems. Further, CNN increases model complexity by adding more layers to it. We will use the longitudinal collection of MRI data. We classify the subjects into demented and non demented at the end after applying the algorithms based on MMSE , ETIV, ATLAS and even find which algorithm best predicts the results by comparing their accuracy levels.

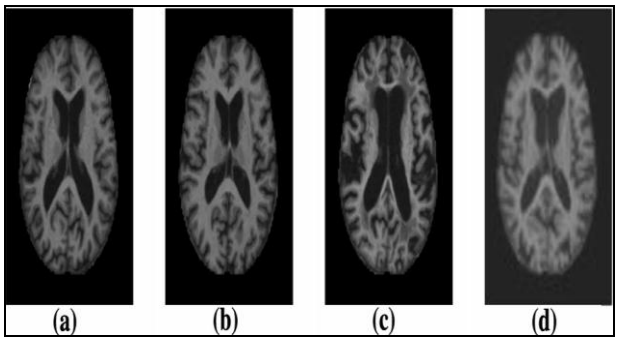

FIGURE 1: STAGES OF ALZHEIMER

II. LITERATURE SURVEY

MRI analysis is ritualized for detection of Alzheimer's disease. There are two core blocks that make up an automatic AD scrutiny system [1]. They are extracting of features from the MRI data and classifying model or a classification algorithm based on those features. There are 3 major categories of feature extraction methods. They are discussed below. Voxel intensity based extraction approaches consider voxel intensities as the classification feature. It measures the differences in local concentrations of brain tissue. The connotation of the results in voxel based representation [7] is simple and instinctive. This approach abide over fitting problem as there are finite in the range of tens to hundreds subjects but they are with many high dimensional features [6]. Voxel-based approaches ignore regional information.

Region of interest (ROI)-based extraction approach uses the functionally or structurally preset regions of brain and bring out features from every region [4]. But in this approach, the extracted features cannot amount to tenuous changes analogous to abnormalities in the brain. The changes in brain due to neurological chaos are not constricted to a peculiar region and are usually extended to numerous regions of the brain. Considering the fact that the affected area can be a part of a single region or extend to numerous ROIs, voxel intensity based or ROI extraction approaches cannot coherently apprehend the disease allied symptoms.

Further, Region of interest extraction approaches requires domain information.

Patch wise extraction approaches [3] split the entire image of brain into compact patches and this patches can be used to extract feature vector. Patch extraction do not demand Regional information identification, so the use of domain knowledge is reduced.

Data is retrieved from whole brain and this approach better apprehends the disease-related pathologies and traits that leads in better diagnosis. In Comparison to Voxel intensity based approach, Patch wise extraction can bring out even small changes in brain with sufficiently reduced dimensionality. The major challenge in this approach is to select informative patches out of all the patches.

III. SYSTEM DESIGN

The following is the brief implementation details of the proposed system

A. UNDERSTANDING THE DATA

Oasis dataset is a longitudinal collection of 150 subjects ranging from age of 60 years to 96 years. The subjects include men and women and everyone is right handed. Out of 150 subjects, 72 of them were found out to be nondemented and 64 subjects were reported to be demented. They include 51 subjects who were in mild phase or moderate phase of Alzheimer's disease and 14 subjects who were reported as nondemented during their initial visit and later were reported as demented. These 14 subjects hence come under the converted category.

The following are the parameters to be considered for the Alzheimer detection and classification in our proposed model based on analysis.

Mini–Mental State Examination (MMSE)

It is a questionnaire to measure cognitive impairment. It also acts as a potent means to log a persons' response to the treatment. It can also be referred as Folstein test and plays a vital role in the analysis of Dementia.

Cognitive Impairment

A person is said to be suffering from cognitive impairment if he/she faces issues in recollecting things, assimilate new things, focussing or decision making.

Clinical Dementia Rating (CDR)

The CDR is a 5-point numeric scale which can differentiate and measure level or severity of dementia. An overall CDR score may be calculated using an algorithm.

Estimated Total Intracranial Volume (eTIV)

The only notable foretell of TIV in case of Alzheimer disease is sex. Men were characterized with approximately ~12% larger eTIV than women. We measured TIV using a semiautomated segmentation technique on T1-weighted and T2-weighted MRI in 55 controls, 10 AD patients.

TABLE 1: CDR SCORE AND THEIR INTERPRETATION

PHASE OF AD	CDR RATING
Severe	3
Moderate	2
Mild	1
Very mild	0.5
Normal	0

Atlas Scaling Factor (ASF)

Atlas scaling factor commonly known as atlas normalization takes total intracranial volume i.e., TIV as a reference value and must be proportional to TIV. It can sometimes also act as a proxy for the manually determined TIV value.

EDUC refers to the figure of years of education

SES refers to the socio economic status of the subject that is the living environment and conditions.

B. DATA PREPROCESSING

Data preprocessing is used to mutate the raw data to the form convenient for our analysis. In general, data may be incomplete, not containing some values, lacking certain attributes required for analysis, or containing summarized information only. So its important that we preprocess data for accurate results.

STEP 1: IMPORT THE DATASET

We will be analysing oasis longitudinal data set for classification and dementia detection which includes longitudinal MRI data of 150 subjects ranging from 60years to 96years of age who are right handed whose Mini mental state examination, eTIV, CDR, ASF(atlas scaling factor), SES(socio economic status), years of education(EDUC) has been considered. According to the rules we took 2/3rd of data as training set and rest data i.e., $1/3^{rd}$ as test data for prediction on response values. But in the proposed model we can also take entire data as training data so as to find and analyze unnecessary attributes values. But in the proposed model we can also take entire data as training data so as to find and analyse unnecessary attributes which are used in training and when used for prediction can result in inaccurate results.

STEP 2: CHECK OUT THE MISSING VALUES

Initially we check for null values and missing values in the data set and count total rows in column which contain null values using pandas. In case of undetected or unhandled missing values we may arrive at erroneous deductions from the data. We use pandas duplicated() function to return the duplicate rows using Boolean values that is true for duplicate rows and others as false.

STEP 3: FILL OUT THE MISSING VALUES

In this project we fill the missing values of categorical attributes with column or attribute median. For example the attribute socioeconomic status (SES) has null values so fill them with column median.

We can fill the missing values of discrete or continuous attributes with column or attribute mean. For instance the attribute mini mental state examination (MMSE) has null values so fill them with column mean

STEP 4: FINDING OUTLIERS IN DATA

Outliers are the attribute values that deviate from the other observations in the dataset.

For this purpose we take help of inter quartile ranges. Interquartile ranges can be found using seaborn library. We can find the outliers in the dataset by first multiplying the interquartile range by 1.5 and then adding the result to Q3 and subtract it from Q1 where Q1, Q2, Q3, Q4 are the quartiles. In this project we are detecting outliers in socioeconomic status(SES) , mini mental state exam(MMSE), educational status(EDUC), atlas scaling factor(ASF), estimated total intracranial volume (ETIV), normalized whole brain volume(NWBV), MR delay using inter quartile plots created using seaborn and matplotlib.

C. CLASSIFICATION AND DETERMINATION OF ACCURACY

First step is to load the dataset and this can be done using pandas effectively. Next in order to determine accuracy of a model use train dataset by using feature matrix (X) which are simply the variables of our data and predict the response values(Y) for the same dataset using that model and hence, find the accuracy of the model. In this project we used 2 machine learning classifiers ie., SVM and random forest and used oasis dataset to train the models and acquire classification report and find accuracies and compare them.

IV. RESULTS

CORRELATION

It is the mutual association or relationship between the attributes.

The below figure 2 is a table showing correlation coefficients between sets of variables which helps to find the pairs having highest correlation.

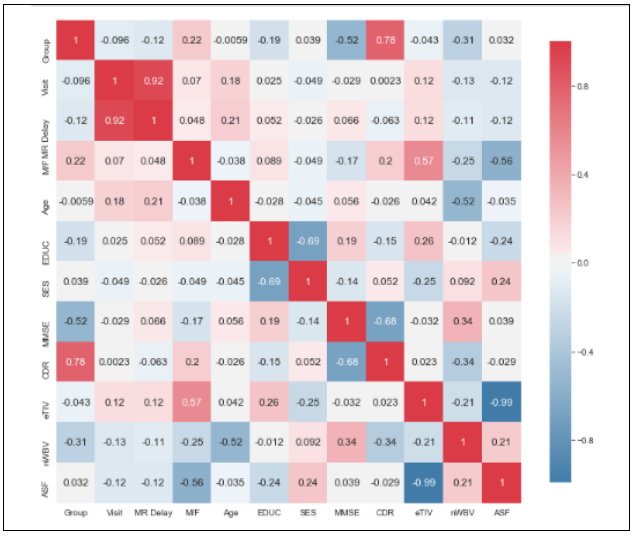

FIGURE 2 : CORRELATION MATRIX

BOX PLOT FOR ATTRIBUTES

The below figure 3 is a boxplot for finding out the outliers present in the numerical attributes graphically. It is interpreted using the two parameters that are the ranges and the inter quartile ranges.

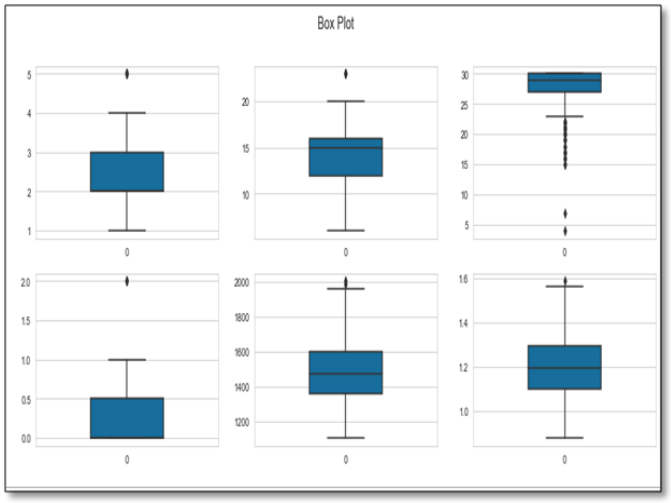

FIGURE 3 : BOX PLOT FOR ATTRIBUTES

CONFUSION MATRIX OF RANDOM FOREST

In the below figure 4, confusion matrix reveals that

TN values ie., Nondemented classified as nondemented are 45

TP values ie., demented classified as demented are 49

FP values ie., Nondemented classified as demented are 7

FN values ie., demented classified as nondemented are 11

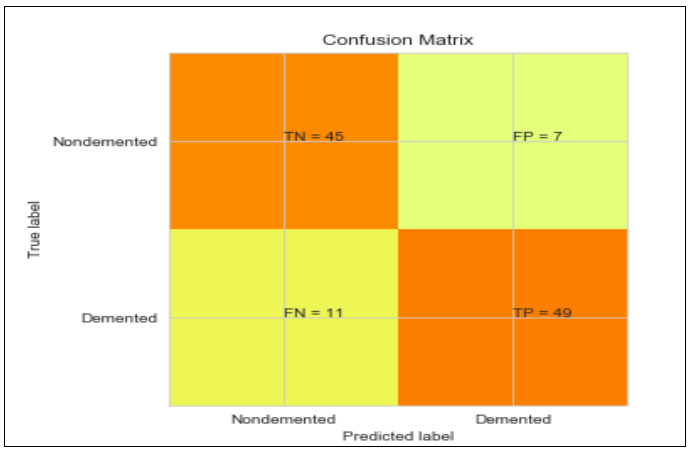

FIGURE 4 : CONFUSION MATRIX OF RF

CONFUSION MATRIX OF SUPPORT VECTOR MACHINE

In the below figure 5, confusion matrix reveals

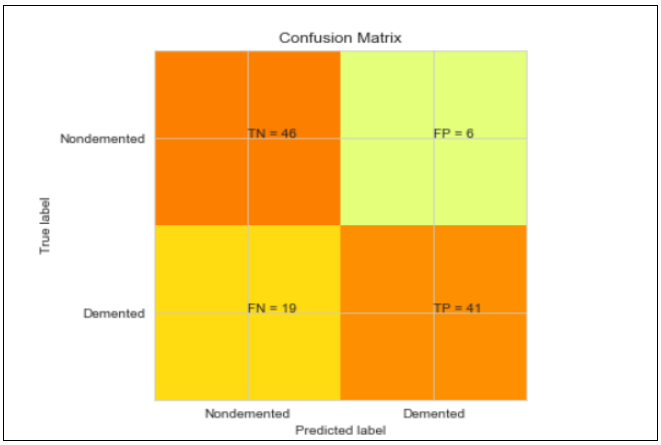

FIGURE 5 : CONFUSION MATRIX OF SVM

TN values ie., Nondemented classified as nondemented are 46

TP values ie., demented classified as demented are 41

FP values ie., Nondemented classified as demented are 6

FN values ie., demented classified as nondemented are 19

CLASSIFICATION REPORT OF RANDOM FOREST

Classification Report:

	precision	recall	f1-score	support
0	0.80	0.87	0.83	52
1	0.88	0.82	0.84	60
micro avg	0.84	0.84	0.84	112
macro avg	0.84	0.84	0.84	112
weighted avg	0.84	0.84	0.84	112

FIGURE 6 : ACCURACY OF RANDOM FOREST CLASSIFIER

Therefore, the accuracy of random forest classifier is found to be **0.83928**

CLASSIFICATION REPORT OF SVM CLASSIFIER

Classification Report:

	precision	recall	f1-score	support
0	0.71	0.88	0.79	52
1	0.87	0.68	0.77	60
micro avg	0.78	0.78	0.78	112
macro avg	0.79	0.78	0.78	112
weighted avg	0.80	0.78	0.78	112

FIGURE 7: ACCURACY OF SUPPORT VECTOR MACHINE

Therefore , the accuracy of random forest classifier is found to be **0.77672.**

V. CONCLUSION

We observe that out of all the factors taken into account for the Alzheimer's classification, MMSE (Mini Mental State Examination) plays a vital role and subjects who are demented have low MMSE scores compared to the subjects who are not suffering from dementia. Dementia is mostly found in 70-80 years old subjects than in young and middle aged subjects. Once the old subjects got attacked with it, chances of their survival is very low. Hence if Alzheimer if detected at the earliest possible can be diagnosed using various treatments being researched. We also notice that the accuracy of Random forest classifier is 0.839 whereas the accuracy of Support Vector Machine(SVM) is 0.776. Hence we conclude that Random forest classifier best suits for predicting the response values and for Alzheimer's classification. We can further extend our investigations to detect Alzheimer at the earliest stages possible for proper diagnosis.

REFERENCES

1. Jyoti Islam and Yanqing Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks", *Brain Informatics 5:2, Springer*, pp.1-4 (2018).
2. Noel O'Kelly, "Use of Machine Learning Technology in the Diagnosis of Alzheimer's Disease ", *doras.dcu.ie*, Vol 2, pp.3-7 (2016).
3. Suk HI, Lee SW, Shen D, Initiative ADN et al, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis" , NeuroImage, pp.569–582 (2014).
4. Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehéricy S, Habert MO, Chupin M, Benali H, Colliot O, " Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database", *Neuroimage* 56(2), pp.766–781(2011).
5. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL," Open access series of imaging studies (OASIS):cross-sectional MRI data in young, middle aged, nondemented, and demented older Adults", *J Cogn Neurosci* 19(9), pp.1498–1507 (2007).
6. Friedman J, Hastie T, Tibshirani R, "The elements of statistical learning", *Springer*, vol.1 (2001).
7. Davatzikos C, Genc A, Xu D, Resnick SM ," Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy", NeuroImage, pp.1361–1369 (2001).
8. Satyabrat, Rishi agarwaal, Deepak et al, www.github.com/topics/alzheimer.

AUTHORS PROFILE

Gousiya Begum, is pursuing her PhD(CSE) from JNTUA, Anantapuramu, She has received M.Tech Degree in Computer Science and Engineering from JNTU, Hyderabad. She is currently working as Assistant Professor in Computer Science and Engineering Department in Mahatma Gandhi Institute of Technology, Hyderabad. She is a Life Member of **ISTE**. She has published 7 papers in International Journals and presented 1 paper in International Conference. Her areas of research are Big Data, Data Mining, Natural Language Processing, Network Security, Software Engineering, and Cloud Computing.

A. Sai Manisha, is a student of CSE dept of MGIt. She is studying final year B.Tech. She is very talented,hardworking. Her areas of interest are data mining, Machine learning.

.