

ALZHEIMER'S DISEASE PREDICTION



A Project Report in partial fulfillment of the degree

Bachelor of Technology

in

Computer Science & Engineering /

Electronics & Communication Engineering

By

19K41A0539

G. Anjali

19K41A0540

G. Sai Srujan

19K41A0450

N. Bhavith

Under the Guidance of

Naresh Kumar

Assistant Professor School of CS & AI SRU

Submitted to

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

S.R.ENGINEERING COLLEGE (A), ANANTHASAGAR, WARANGAL

(Affiliated to JNTUH, Accredited by NBA) Dec-2021.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that the Project Report entitled “ALZHEIMER’S DISEASE PREDICTION” is a record of bonafide work carried out by the student(s) Anjali, Sai srujan, Bhavith bearing Roll No(s) 19K41A0539, 19K41A0540, 19K41A0450 during the academic year 2020-21 in partial fulfillment of the award of the degree of **Bachelor of Technology in Computer Science & Engineering / Electronics & Communication Engineering** by the S.R. ENGINEERING COLLEGE, Ananthasagar, Warangal.

Supervisor

Head of the Department

External Examiner

ABSTRACT

Alzheimer's disease is one in all the foremost devastating brain disorders of elderly humans. It's an undertreated and under-recognized disease that's becoming a significant public pathological state. Prediction of this disease may be a critical challenge within the area of clinical data analysis. Machine learning (ML) is effective in assisting in making decisions and predictions from the massive quantity of knowledge produced by the healthcare industry. We've got also seen ML techniques getting used in recent developments in several areas of the net of Things (IoT). Various studies give only a glimpse into predicting Alzheimer's disease with ML techniques. In this report, we propose a unique method that aims at finding significant features by applying machine learning techniques leading to improving the accuracy within the prediction of this disease. The prediction model is a combination of different features and a number of other known classification techniques. We produced an enhanced performance level with an accuracy level of 94.36% through the prediction model for Alzheimer's disease with the Random Forest and 93.66% with the SVM classifier.

Table of Contents

S.NO	Content	Page No
1	Introduction	5
2	Literature Review	7
3	Design	7
4	Dataset	8
5	Data Pre-processing	10
6	Methodology	15
7	Results	18
8	Conclusion	19
9	References	19

1. INTRODUCTION

Alzheimer's infection is the most widely recognized sort of dementia. It is a dynamic illness starting with gentle cognitive decline and conceivably prompting loss of the capacity to carry on a discussion and react to the climate. This infection includes portions of the cerebrum that control thought, memory, and language. It can genuinely influence an individual's capacity to do every day exercises.

In 2020, as numerous as 5.8 million Americans were living with Alzheimer's disease. More youthful individuals might get Alzheimer's infection however, it is less normal. The number of individuals living with the infection pairs like clockwork past age 65. This number is projected to almost significantly increase to 14 million individuals by 2060. Side effects of the infection would first be able to show up after age 60, and the danger increments with age.

Researchers don't yet completely get what causes Alzheimer's sickness. There likely is certainly not a solitary reason but instead, a few factors that can suddenly influence every individual, and age is the most popular danger factor for Alzheimer's sickness.

Family ancestry analysts accept that hereditary qualities might assume a part in fostering Alzheimer's infection. In any case, qualities don't rise to fate. A sound way of life might assist with lessening your danger of fostering Alzheimer's infection. Two huge, long-haul studies demonstrate that sufficient actual work, a nutritious eating routine, restricted liquor utilization, and not smoking might help individuals.

Notwithstanding memory issues, somebody with indications of Alzheimer's illness might encounter at least one of the accompanying:

- The cognitive decline that disturbs day-to-day existence, like losing all sense of direction in a natural spot or rehashing questions.
- Inconvenience taking care of cash and taking care of bills.
- The cognitive decline that disturbs day-to-day existence, like losing all sense of direction in a very natural spot or rehashing questions.
- Inconvenience taking care of money and taking care of bills.
- Trouble wrapping up natural jobs reception, working or at relaxation.
- Diminished or misguided thinking.
- Losing things and being ineffective to recollect steps to trace down them.
- Changes in temperament, character, or conduct.

Regardless of whether you or somebody you recognize contains a few or perhaps the greater a part of these signs, it does not imply it's Alzheimer's infection. Getting checked by your medical services supplier can help decide whether the indications you're encountering are identified with Alzheimer's infection or more treatable conditions sort of a nutrient lack or a secondary effect from drug.

Early and exact analysis likewise gives freedom to you and your family to consider monetary preparation, foster development orders, tackle clinical preliminaries, and expect care needs. Clinical administration can work on personal satisfaction for people living with Alzheimer's illness and for his or her parental figures. There is straight away no known solution for Alzheimer's infection.

Albeit an excellent many folks eagerly help to their friends and family and companions, really specializing in a private with Alzheimer's infection reception will be a hard assignment and will become overpowering now and again, each day brings new difficulties because the parental figure adapts to changing degrees of capacity and new samples of conduct, because the illness deteriorates, individuals living with Alzheimer's sickness frequently need more concentrated consideration.

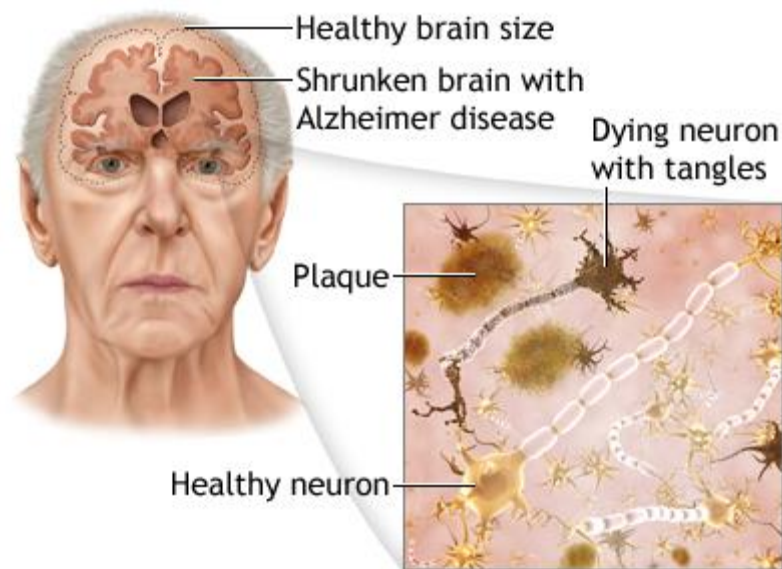


Figure 1: Neuron of an Alzheimer's patient

2. LITERATURE REVIEW

Machine learning is helpful for a variety of situations. The prediction of dependent variable values from independent variables is one of the uses of this methodology. The field of healthcare is a data mining field since it contains extensive data resources that are challenging to manually handle. Even in wealthy countries, Alzheimer's disease was identified as one of the major problems faced by elderly humans. The fact that the hazards are either not discovered or only later is one of the reasons for suffering from this disease. Machine learning methods can nonetheless be helpful in solving this challenge and in anticipating danger early. Some of the techniques used for such prediction problems are Decision tree, Random Forest, and SVM algorithms. The existing research has used ensemble methods to improve the accuracy in the prediction of Alzheimer's disease but they are not much accurate than our proposed system. Our model showed the highest accuracy rate with Decision tree [92.95%], Random Forest [94.36%], SVM [93.66%].

3. DESIGN

3.1 REQUIREMENT SPECIFICATION(S/W & H/W)

Hardware Requirements

- | | |
|--------------------|---|
| ✓ System | : Pentium 4, Intel Core i3, i5, i7 and 2GHz Minimum |
| ✓ RAM | : 4GB or above |
| ✓ Hard Disk | : 10GB or above |
| ✓ Input | : Keyboard and Mouse |
| ✓ Output | : Monitor or PC |

Software Requirements

- | | |
|---------------------------|----------------------------------|
| ✓ OS | : Windows 8 or Higher Versions |
| ✓ Platform | : Jupiter Notebook, Google Colab |
| ✓ Program Language | : Python |

3.2 FLOW CHART

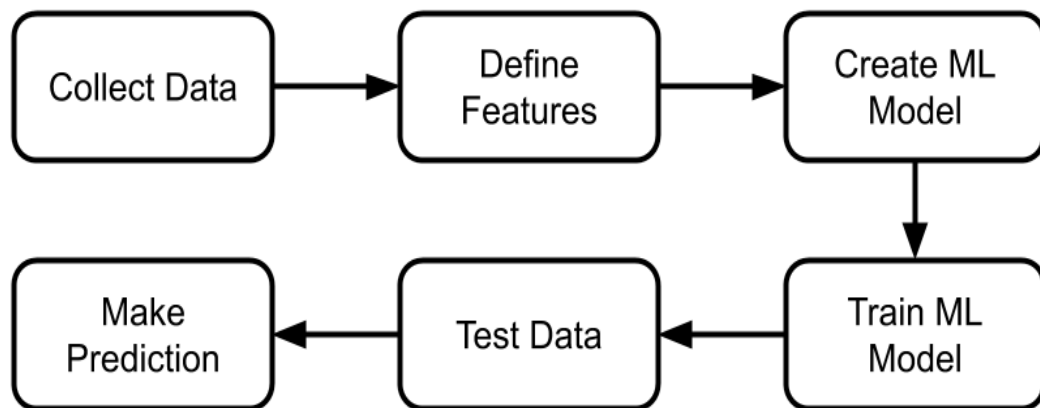


Figure 2: Flow chart

This is an analysis of Alzheimer's disease using different machine learning models. It predicts whether a person will suffer from Alzheimer's or not. For analysis of such models, we use different machine learning tools like pandas, numpy, matplotlib, sci-kit-learning, etc.

4. DATA SET

The team has found MRI related data that was generated by the Open Access Series of Imaging Studies (OASIS) project that is available both, on their website and on kaggle that can be utilized to train various machine learning models to identify patients with mild to moderate Alzheimer's.

The dataset consists of a total of 15 different columns out of which 10 attributes are used for our machine learning model to predict whether a person will suffer from Alzheimer's or not.

- Group : Demented / Non-demented / Converted
- M/F : M-Male, F-Female
- Hand : R-Right, L-Left
- Age : Age in years
 - Its range is between 60 to 98
- EDUC : Education
 - Its range is between 6 to 23
- SES : Socio Economic Status

- Socio-economic status is the social standing or class of an individual or group. It is often measured as a combination of education, income, and occupation.
- Its range is between 1 to 5
- MMSE : Mini mental state examination
 - A mini-mental state examination is a set of 30 questions that doctors and other health care professionals commonly used to check for cognitive impairment (problems with thinking, communication, understanding, and memory).
 - Its range is between 4 to 30
- CDR : Clinical Dementia Rating
 - Clinical dementia rating is a widely used clinical tool for grading the relative severity of dementia.
 - Its range is between 0 to 2
- eTIV : Estimated total intracranial volume
 - It refers to the estimated volume of the cranial cavity as outlined by the supratentorial dura matter or cerebral contour when dura is clearly detectable.
 - Its range is between 1106 to 2004
- nWBV : Normalize whole brain volume
 - Investigation of whole-brain anomalies considers the collective volume of the entire brain in structural imaging, without considering regionally specific differences in the volume of any individual structures.
 - Its range is between 0.644 to 0.837
- ASF : Atlas scaling factor
 - Automated atlas transformation generated the atlas scaling factor is defined as the volume scaling factor required to match each individual to the atlas target. Because atlas normalization equates to head size, the ASF should be proportional to total intracranial volume (TIV).
 - Its range is between 0.875 to 1.587

5. DATA PRE-PROCESSING

The dataset consists of a total of 374 rows and 15 columns, of which 8 columns consist of numerical data and the remaining are categorical data.

Group	Categorical [Demented, Nondemented, Converted]
M/F	Categorical [M,F]
Hand	Categorical [R]
Age	Numerical [60 to 98, mean = 77.03]
EDUC	Numerical [6 to 23, mean = 14.7]
SES	Numerical [1 to 5, mean = 2.46]
MMSE	Numerical [4 to 30, mean = 27.4]
CDR	Numerical [0 to 2, mean = 0.27]
eTIV	Numerical [1106 to 2004, mean = 1489.99]
nWBV	Numerical [0.644 to 0.837, mean = 0.72]
ASF	Numerical [0.875 to 1.587, mean = 1.19]

Table 1: Dataset range and datatype

Dropping unwanted data:

We are dropping unwanted columns which are not used for the prediction of Alzheimer's disease those are Subject ID, MRI ID, Visit, MR Delay.

Missing values treatment:

The column SES consists of 19 missing values and MMSE consists of 2 missing values. For its treatment, the correspondent row of the missing value is removed.

Converting categorical data to numerical data:

The columns M/F, Hand, and Group consist of categorical data.

In the column M/F, M - 1 and F - 0

In the column Hand, R - 1 and L - 0

In the column Group, Demented - 1, Nondemented - 0, Converted - 1

Evaluating the data:

The input / feature columns are M/F, Hand, Age, EDUC, SES, MMSE, CDR, eTIV, nWBV, ASF, and the target / output column is Group.

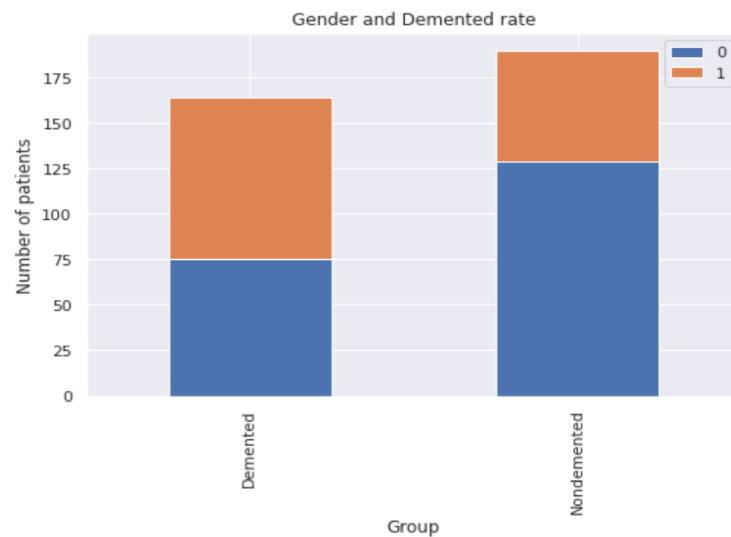


Figure 3: M/F and Demented rate

The figure 3 indicate M-1, F-0 and the men are more likely with dementia than women.

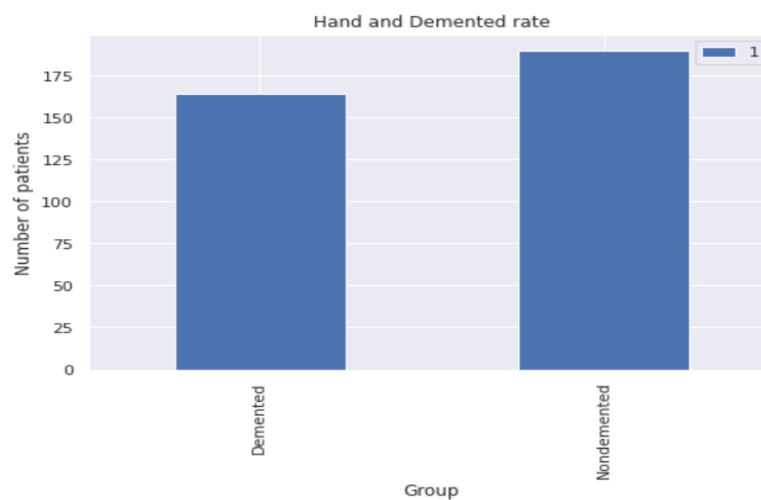


Figure 4: Hand and Demented rate

In the figure 4 R – 1

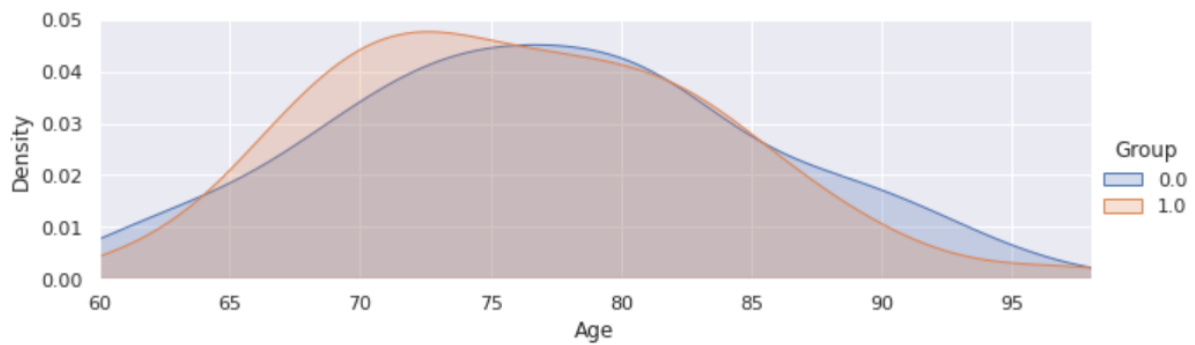


Figure 5: Age vs Group

The figure 5 indicate, there is a higher concentration of 70-80 years old in the Demented patient group than those in the Nondemented patients. We guess patients who suffered from that kind of disease have a lower survival rate so that there are a few of 90 years old.

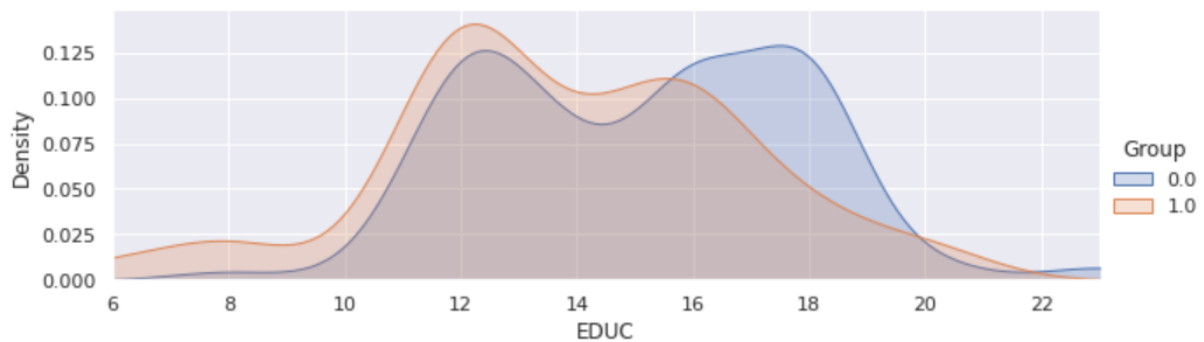


Figure 6: EDUC vs Group

The figure 6 indicate that patients with less education are more likely to be demented.

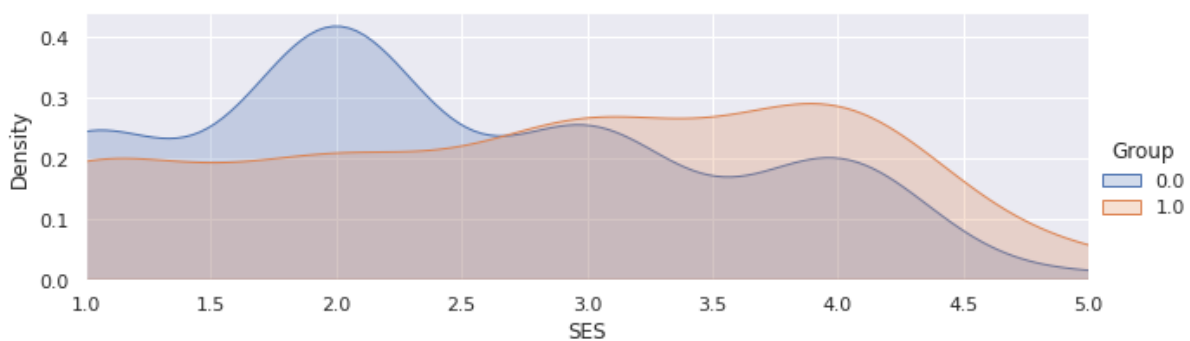


Figure 7: SES vs Group

The figure 7 indicate SES values 1 and 2 indicate the patient is economically good in position, the SES values 3 and 4 indicate economically average while SES of 5 indicates economically poor. From the above graph, a patient with SES 3 to 4 is more likely to suffer from dementia.

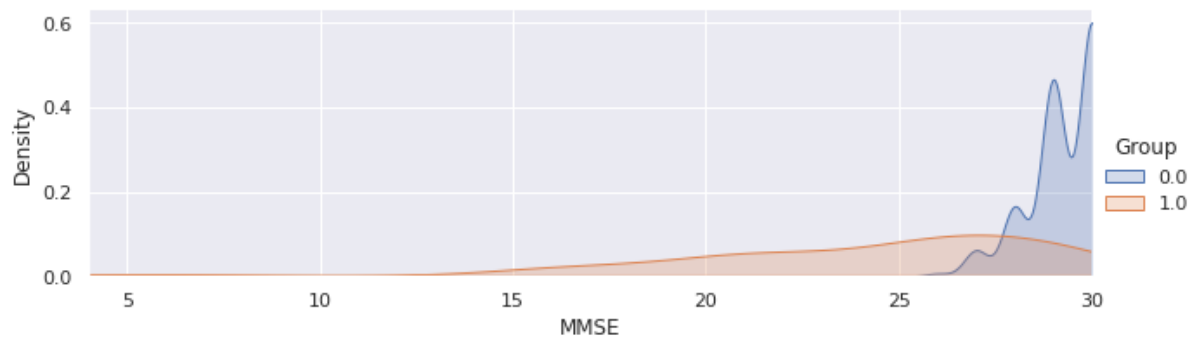


Figure 8: MMSE vs Group

The figure 8 indicate that the Nondemented group got much higher MMSE scores than the Demented group.

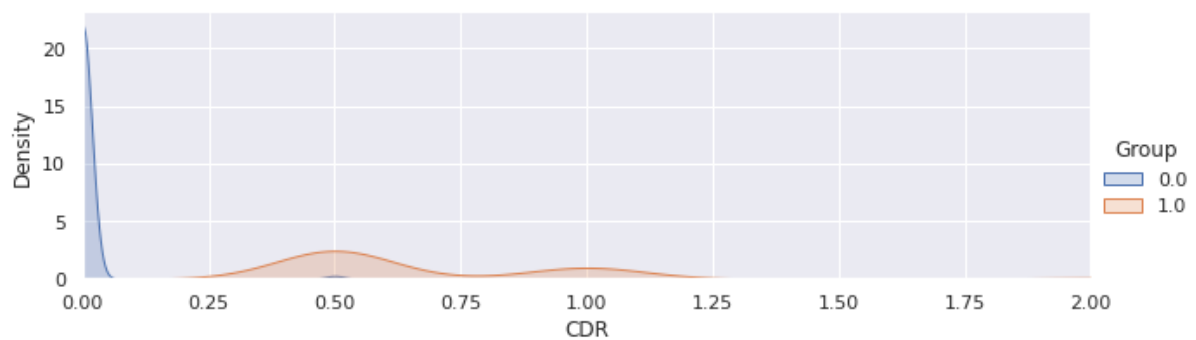


Figure 9: CDR vs Group

The figure 9 indicate that the patients with CDR values above 0.5 are more likely to suffer from dementia.

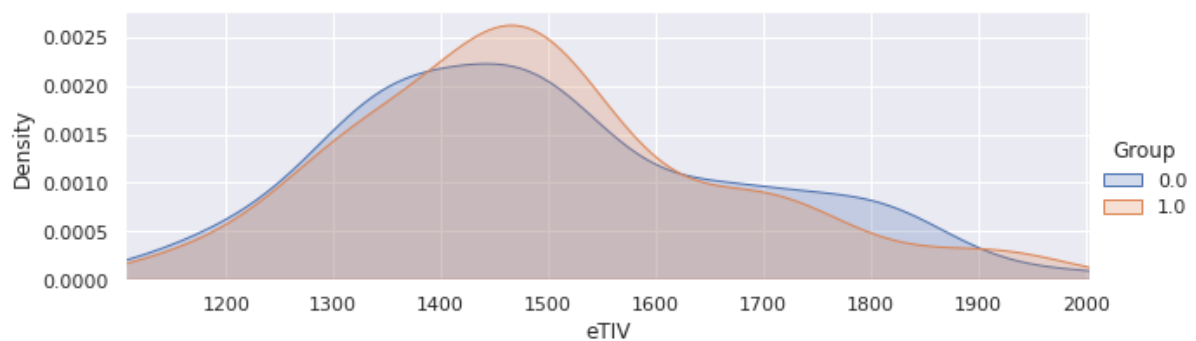


Figure 10: eTIV vs Group

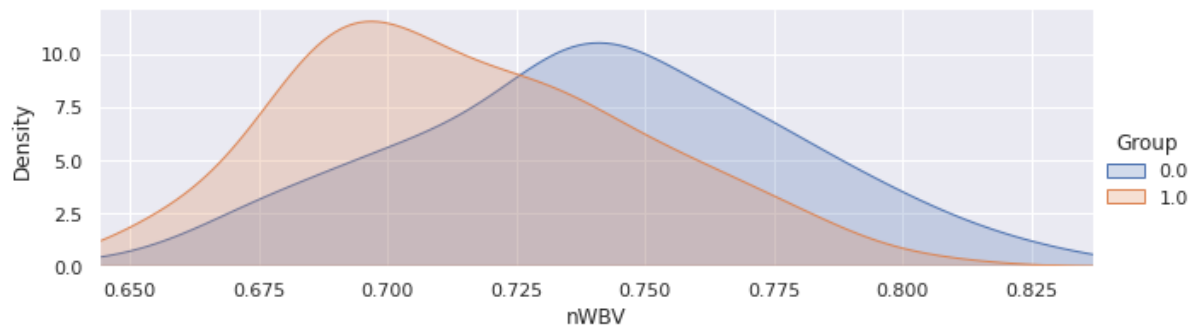


Figure 11: nWBV vs Group

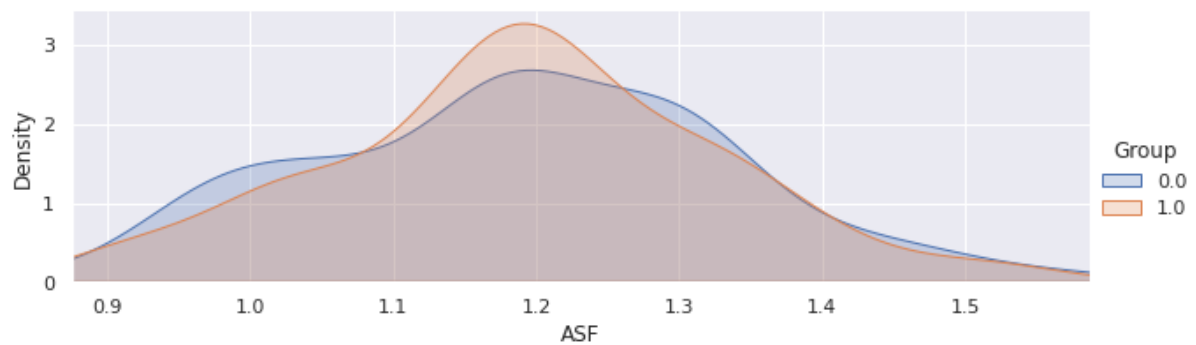


Figure 12: ASF vs Group

The figures 10, 11, 12 indicate that the Nondemented group has a higher brain volume ratio than the Demented group. This is assumed to be because the diseases affect the brain to be shrinking its tissue.

Normalization of data:

By applying min-max- normalization the original data is going to be transformed in the range from 0 to 1.

Split dataset:

Firstly split the dataset into features and target variable, then by using the `train_test_split` method, split the data into a training set and test set.

6. METHODOLOGY

This section talks about the algorithms used for the project. We used three different algorithms, they are Decision tree, Random forest, SVM.

Decision Tree:

A Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

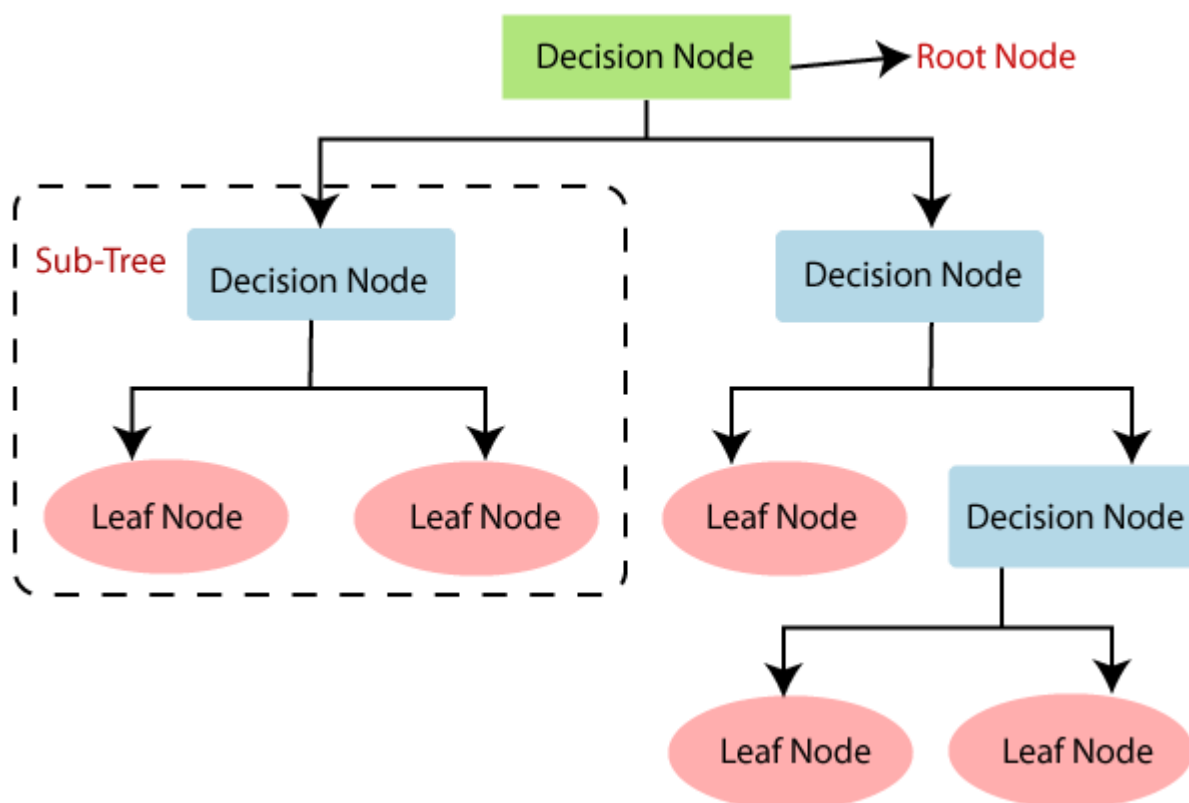


Figure 13: Decision Tree flowchart

Step-1: Begin the tree with the root node, says S, which contains the complete dataset.

Step-2: Find the best attribute in the dataset using the **Attribute Selection Measure (ASM)**.

Step-3: Divide the S into subsets that contains possible values for the best attributes.

Step-4: Generate the decision tree node, which contain the best attribute.

Step-5: Recursively make new decision trees using the subsets of the dataset created in step - 3. Continue this process until a stage is reached where you cannot further classify the nodes and call the final node as a leaf node.

Random forest:

Random forest is used for both regression and classification-based applications. This algorithm is flexible and easy to use. Most of the time this algorithm gives accurate results even without hyper tuning the parameters. It builds many decision trees which on merging forms as a forest. While building the decision trees, adds more randomness to the model. This algorithm searches for the best feature in the random subset of features, which results in the formation of a better model.

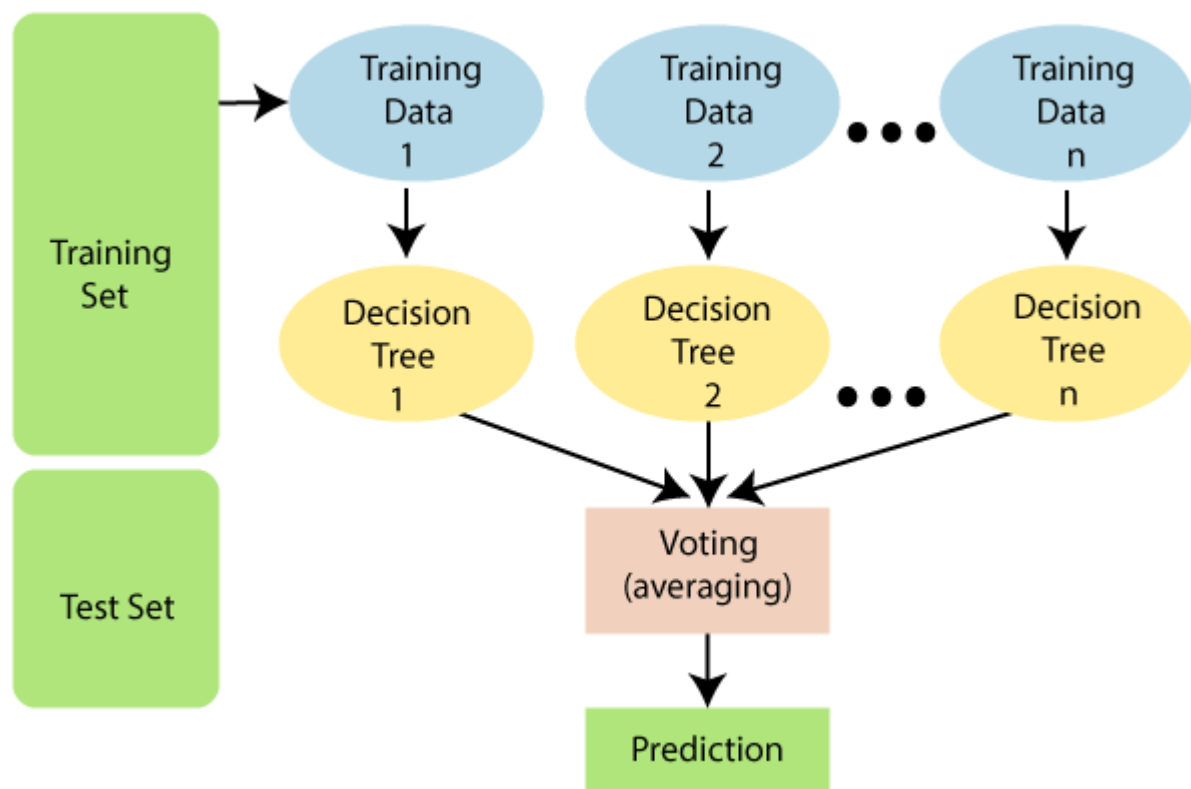


Figure 14: Random Forest flowchart

Random Forest works in two-phase first is to create the random forest by combining the N decision tree, and the second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

Support Vector Machine (SVM):

The objective of the support vector machine algorithm is to find a hyperplane in N -dimensional space (N - the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e. the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

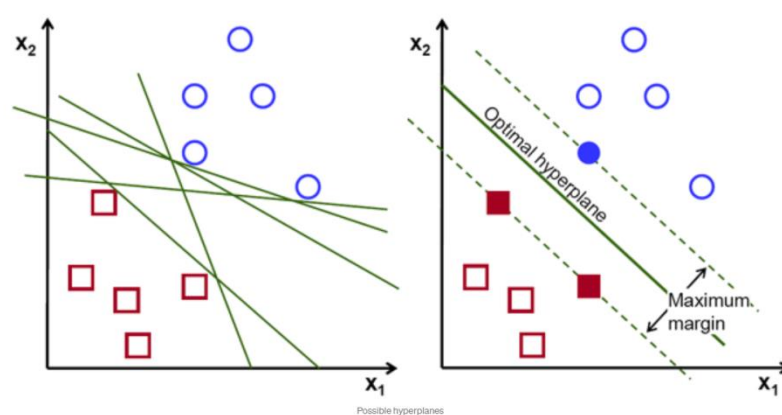


Figure 15: SVM

7. RESULTS

	Training Accuracy	Testing Accuracy
Decision tree classifier	97.6%	92.9%
Random Forest classifier	98.1%	94.3%
SVM classifier	94.8%	93.6%

Table 2: Accuracy rates

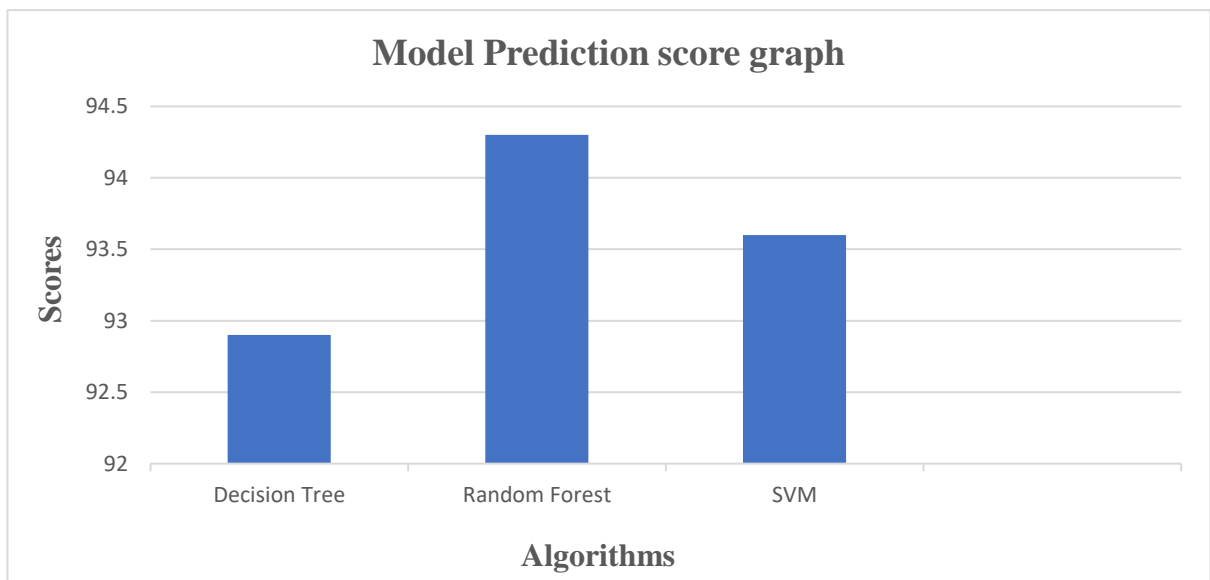


Figure 16: Accuracy of Machine learning models

8. CONCLUSION

This study is based on the comparison of different machine learning algorithms. In the proposed method, the missing values were cleansed by rows deletion in the column of SES and MMSE respectively. Random forest features selection is a reliable technique for early prediction of Alzheimer's. Random Forest algorithm is the best classifier outperformed among all proposed algorithms, with the accuracy of 94.3%.

9. REFERENCES

1. M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri, 'The global prevalence of dementia: A systematic review and metaanalysis', *Alzheimer's and Dementia*. 2013.
2. C. Dillon, C. M. Serrano, D. Castro, P. P. Leguizamón, S. L. Heisecke, and F. E. Taragano, 'Behavioral symptoms related to cognitive impairment.', *Neuropsychiatr. Dis. Treat.*, vol. 9, pp. 1443–55, 2013.
3. A. Khan and M. Usman, 'Early Diagnosis of Alzheimer's Disease using Machine Learning Techniques', 2015 7th Int. Jt. Conf. Knowl. Discov. Knowl. Eng. Knowl. Manag. (IC3K), 2015.
4. J. Islam and Y. Zhang, 'Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks, *Brain informatics*, vol. 5, no. 2, p. 2, May 2018.
5. D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, 'Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults', *J. Cogn. Neurosci*, 2007.
6. S. Teipel et al., 'Multimodal imaging in Alzheimer's disease: Validity and usefulness for early detection', *The Lancet Neurology*. 2015.
7. Alzheimer's Association (2019). Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 15(3):321–387.
8. Kundaram, S. S. and Pathak, K. C. (2021). Deep learning-based alzheimer disease detection. In Springer Science and Business Media LLC, pages 587–597. Springer.
9. Patterson, C. (2018). The state of the art of dementia research: New frontiers. World Alzheimer Report.