IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115153

Detecting Alzheimer using Shallow Learning and Deep Learning Techniques

Sakshi Singh[1], Komal Gaikwad[2], Asma Nehal[3], Sukanya Pawal[4], Poonam Gupta[5]

[1-4]UG Student, Department of Computer Engineering, G H Raisoni College of Engineering and Management, Pune,Maharashtra, India

[5]Assistant Professor, Department of Computer Engineering, G H Raisoni College of Engineering and Management Pune,Maharashtra, India

Abstract: Alzheimer's is considered to be a syndrome where the cognitive functionalities of the human brain decline beyond the expected consequences of ageing. As stated by the World Health Organization (WHO) currently there are nearly about 55 million people who are affected from Dementia. It has been found that not much have been discovered about the consequences of Dementia in India but there are currently nearly 4 million people who are affected with different forms of Dementia in India whereas nearly 44 million people are living with this condition worldwide. In 2019 it became the 9th leading cause of deaths amongst all diseases worldwide leaving behind strokes. Almost 60-70% cases of Dementia are caused by Alzheimer's alone which makes it one of the leading causes of Dementia. With no cure currently available Alzheimer's surely needs to be addressed.

Along with the clinical aspects that are related with detection of Alzheimer's, the psychological and socio-economic impacts are also studied that the disease might have on the affected people with the help of variousshallow learning methods in Machine Learning. The dataset used consists of 372 patient's records. For the deep learning implementation, the dataset comprises of 5121 files belonging to 4 classes. Both the datasets are accessible through Open Access Series of ImagingStudies (OASIS) and also available on Kaggle.

From a combination of linear, distance and tree-based algorithms we have used 12 such algorithms with four clinical features i.e., Mini Mental State Examination (MMSE), Normalized Whole Brain Volume (nWRV), Atlas Scaling Factor (ASF) and Estimated Total Intracranial Volume (eTIV). The analysis have been performed to find certain patterns and correlations inthe dataset on the data fields such as age, gender, education social status etc., furthermore an analysis has been performed using CNN, ResNet-50 transfer learningtechnique to figure out at what stage is the progression of alzheimer is, the stages are namely 'NonDemented', 'VeryMildDemented', 'ModerateDemented', and'MildDemented,

Keywords: Dementia, Alzheimer's Disease, MMSE, nWRV,eTIV, ASF, Machine Learning, Feature Selection, Deep Learning CNN, ResNet50.

I. BACKGROUND

Before working on the project "Alzheimer's Detection" we should first understand what Alzheimer is and, why have we chosen this project. Alzheimer's is actually a type of dementia that has an effect on memory, thinking and behavior. Symptoms in time grow severe enough to hinder the daily tasks. Alzheimer's is known to be the most common cause of dementia, which is a general term for memory loss and Alzheimer's disease reports nearly 60-80% of dementia cases. This disease is a progressive disease, where the symptoms of dementia worsen over time. Alzheimer's disease consistently progresses slowly in three general stages named as early stage, middle stage and last stage.

Nearly 55 million people around the globe are sufferingfrom this tragic disease and more people are getting affectedday by day. Alzheimer's have no cure and no treatments canguarantee to stop this disease from developing. So, we have decided to work on Alzheimer's Detection to contribute to science and medicines. Our main target is to develop utilization strategies that would positively impact early diagnosis to lead to better outcomes and lower costs for patients, caregivers and the healthcare systems. We are implementing our project with the help of ensemble learningmethods combined with other deep learning technologies. Therefore, we are working to develop a research-based model for detecting Alzheimer's in the early stage of development using machine learning techniques.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115153

II. LITERATURE REVIEW

There has been sufficient research on detection of Alzheimer's disease with respect to both physical /clinical parameters and MRI based diagnosis. Amongst the researches that we considered we found the following conclusions:

[1] suggested a Machine learning algorithm along with psychological parameters and clinical information like MMSE, CDR, EDUC, etc. they have used SVM and Decision tree algorithm to detect Alzheimer disease anddistinguish between cognitive impairment.

[2] MRI images and process them to get the numerical data, on which they performed different machine learning algorithms such as SVM, Gradient Boosting, Neural network, KNN, and random forest to detect Alzheimer disease. In their research paper, they found that Neural Network and Random Forest have much better performance.

[3] This paper used the oasis dataset and they use feature extraction using the EM algorithm, feature selection using a best-first algorithm, classification is done using the Gaussian process algorithm, and for the missing data, they have used mean and mode method.

[4] In this research paper they have used both hard voting classifier and soft voting classifier and they observe that when soft voting classifier used on bio-markers such as MMSE, CDR, normalized whole brain volume produces the best accuracy among all other algorithm used by them.

III. DATASET

[A] Clinical Dataset: We have used "oasis_longitudinal.csv"dataset which is an Open-source dataset available on variousplatforms. This dataset consists of 15 features which reportseducation status, social economic status, and various clinical records of the patients that we will briefly discuss here. After removing the NaN values from a total of 372 patients we were left with 354 which further classifies into 204 Female and 150 Male patients with their age ranging between 60 to 98 years

Feature	Explanation	Range
Group	Demented or Nondement	Demented, Nondemented
Visit	Number of visits to the clinic	[0,5]
M/F	Male or Female	Male, Female
Hand	Dominant Hand	Right, Left
Age	Age of the patient	[60….,98]
EDUC	Years of Education	[6……,23]
SES	Socio Economic Status	[1,5]
MMSE	Mini Mental State Examination	[4…...,30]
CDR	Clinical Dementia Rating	[0,2]
eTIV	Estimated Total Intracranial Volume	[1106…,2004]
nWBV	Normalized Whole Brain Volume	[0.644..,0.837]
ASF	Atlas Scaling Factor	[0.876..,1.587]

Fig 1 Description of data fields

[B] MRI Dataset: We have used "oasis_longitudinal.csv" dataset which is an open source dataset available on various platforms. This dataset consists of test and train directories with a total of 6400 images.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115153

IV. UNDERSTANDING THE DATA

Based on the various data visualization techniques we havegot the following correlations amongst the columns from our clinical dataset keeping Demented Group as target variable.:For all the below represented graphs:

Non-Demented Group
Demented Group

[A] SES

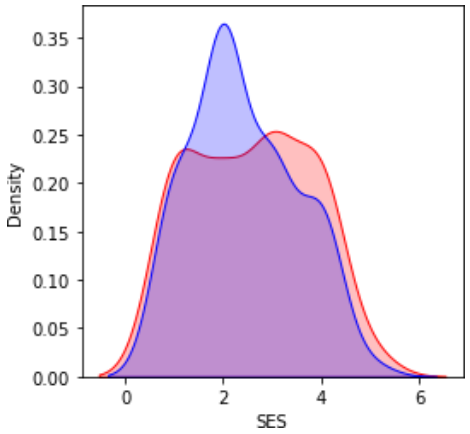

Fig. 2 Socio Economic Status

From (Fig. 1) it can be clearly observed that lower levels of SES are related with higher risk of having Alzheimer's Disease.

SES is the socioeconomic status of an individual or a group. It is measured as a combination of education, income as well as total family income. Socioeconomic status is also referred to as a combination of a person's work experience and how that person is access to all the resources as compared to others. Socioeconomic status is a key featurefor detecting Alzheimer's disease Because some studies show that lower SES will be associated with a high risk of Alzheimer's. Lowering the SES means that a low level of education that will impact one's professional life will increase the risk of Alzheimer's disease and dementia.

[B] MMSE

Mini-Mental State Examination is used to check the short- term as well as long-term memory, concentration, language, and communication skills. It is mainly used for older adults to measure their cognitive impairment. It is also known as the Folstein test and it is 30 points questionnaires. The duration of the test is between 5 to 10 minutes which includes registration, languages, recall, attention and simple calculation, ability to follow simple steps. A person who scores 25 and above will be classified as normal with no impairment, If the score is between 21 to 24 then that personwill have mild impairment. A score between 10 to 20 meansmoderate and less than 10 will have a severe impairment.

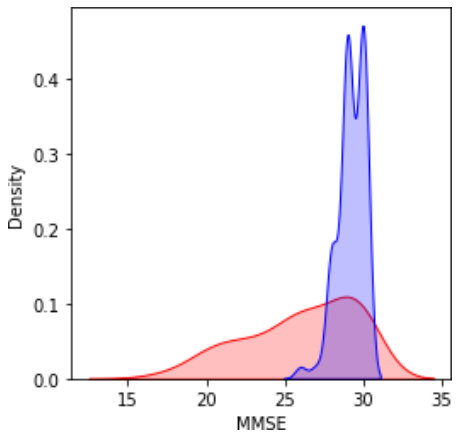

Fig. 3 Mini-Mental State Examination

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115153

Low Scores of MMSE are associated with higher risks of having Alzheimer's Disease.

[C] EDUC

It refers to the years of education that one has received throughout their life. In the dataset that we have its value ranges from 6 years to 23 years from the below visualizationwe can conclude that low values of education results tohigher risks of having Alzheimer's. various studies also confirm the same interpretation as of ours, but no solid reason as of why it happens is discovered yet.

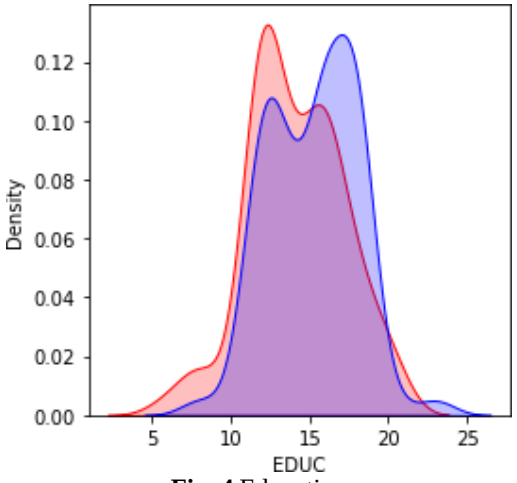

Fig. 4 Education

[D] Age

The greatest risk factor associated with AD is increasing ageand the majority of people with AD worldwide are beyond the age of 65. From (Fig. 4) we can state that the concentration of people affected with AD is higher between 70-80 years of age.

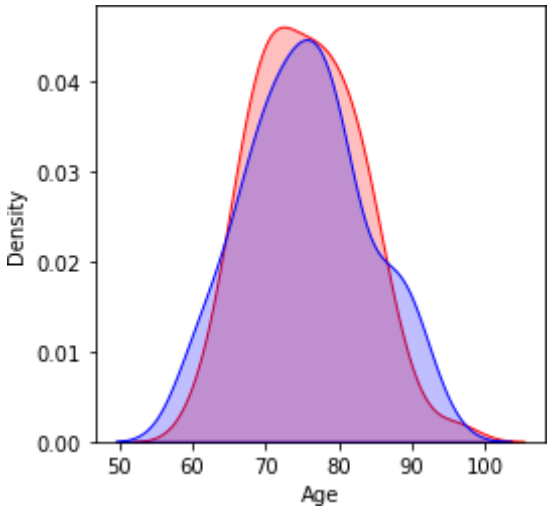

Fig. 5 Age

[E] eTIV, nWRV and ASF

Estimated Total intracranial volume (eTIV) (Fig. 6), is the estimated measurement of Total Intracranial Volume (TIV) of the brain which is commonly used in studies related to volume inside the cranium (skull). The ICV (intracranial volume) is often preferred over the brain volume as it is a good measure of premorbid brain size (It determines the volume of the brain which is affected by Alzheimer's disease)

Normalized Whole Brain Volume (nWRV) (Fig. 7), considers the collective volume of the entire brain which is often measured and normalized with TIVs (Total IntracranialVolume). It has been found out that it significantly reduces inpeople with AD.

Atlas Scaling Factor (ASF) (Fig. 8), is the volume scaling factor that equates the head sizes of different individuals as allows the comparison of the eTIV It should be proportional to the Estimated Total intracranial volume (eTIV).

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115153

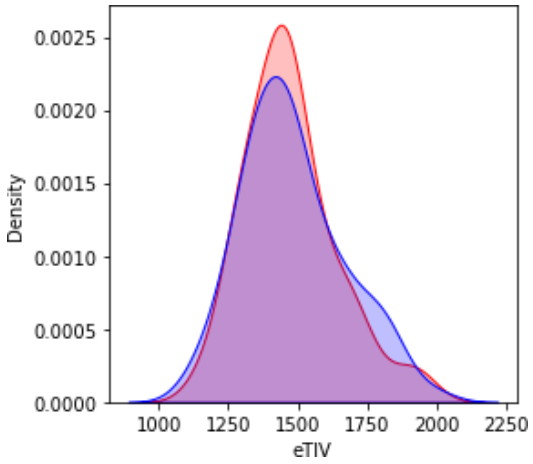

Fig. 6 Estimated Total intracranial volume

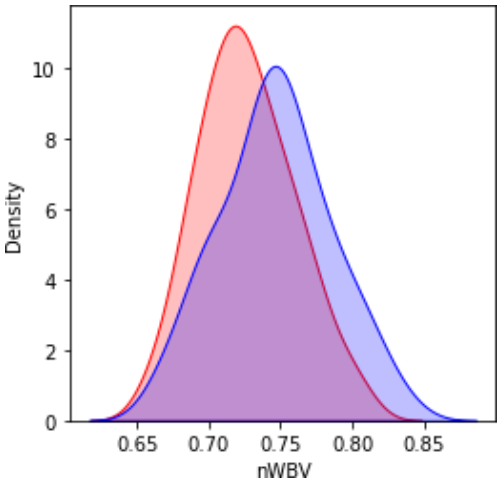

Fig. 7 Normalised Whole Brain Volume

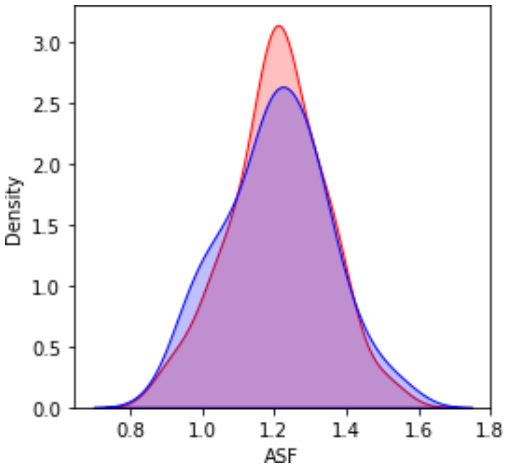

Fig. 8 Atlas Scaling Factor

The above all graphs indicates that those with Dementia havelower ratios of Brain Volumes than the Non Demented group of people. This observation is also backed with scientific researches and is associated with the shrinking of the brain tissues or dead brain cells in the people affected with AD.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115153

V. FEATURE RANKING

For selecting the appropriate features, we have used theMann–Whitney's test and correlation matrix using Seaborn library which performs the common univariate Pearson's test. We have also implemented the Shapiro-Wilk Test in order to detect if a feature follows a normal distribution or not.

The Mann–Whitney U test stands for Wilcoxon rank-sum test, here it is applied to each feature in relation to the Group field i.e., target, it detects whether we can reject the null hypothesis that the distribution of each feature for the groupsof samples defined by target variable (Group-Demented,Nondemented) are the same. A low p-value of this test (closeto 0) indicates that the analyzed feature strongly relates to having AD, while a high p-value (close to 1) indicates the opposite.

The Shapiro-Wilks test is used for checking the normality and is one of the three general normality tests that are designed to detect all departures from normality. It rejects the hypothesis of normality for a p-value less than or equal to 0.05. Here we can see that some features in our data do not follow a normal distribution whereas some features such as Age, ASF and nWBV follow a normal distribution.

After performing all the above-stated methods we found out that Gender, Age, nWRV, MMSE, EDUC, were some of the strongly related features where SES, CDR, EDUC, MMSE, eTIV, Gender does not follow a normal distribution. Since a lot of research has already been carried out to study the effects of features such as Age, Gender, and Education on AD we Decided to go with the clinical set of features from our dataset i.e., ASF, nWBV, MMSE and eTIV for implementing the Machine learning algorithms.

VI. METHODS

A. Logistic regression:

It is a machine-learning algorithm to find the relationship between an independent (x) and dependent variable(y) where the probability of a particular class is predicted using ywhich is plotted between 0 and 1 on scale and x which is plotted accordingly. It can be represented using the sigmoid function which is y=1/1x where e is Euler's constant which gives an S-shaped curve on a graph. Data points are divided into two classes, where one class is true, yes, etc. and the other is vice versa.

B. Random forest:

It is a machine learning technique that is used forclassification. It consists of many decision trees. It is used forpredicting the values and making decisions. The first bootstrapping dataset is created randomly using the original dataset which may or may not include all the records and records might be repeated. From this dataset, decision trees are made, and the final decision or prediction will be made by voting amongst all trees. Nodes of decision trees are also selected randomly as the name suggests.

C. Decision Tree:

It is a machine learning algorithm that is used in regression as well as classification. In this, a decision tree is used to make a decision. The test is performed on a root node/attribute/feature. A leaf node is a classified value. Basically, a decision tree is a tree-shaped classifier

D. Boosting:

It is a machine learning technique in which numerous base models are considered and they are trained using different sets of datasets. If a tuple/record is misclassified by a model then it is prioritized and used for the training of another model tuple and a final/strong model is created using these models. After training of various weak models one final/strong model is created. Test tuple is provided to the models and their decision is passed to a final model which uses voting for making decisions.

E. Gaussian Naïve Bayes:

The Gaussian Processes Classifier is a classification machinelearning algorithm. Gaussian Processes are a generalization of the Gaussian probability distribution and can be used as the basis for sophisticated non-parametric machine learning algorithms for classification and regression.

F. Adaboost:

In the case of AdaBoost, higher points are assigned to the data points which are miss-classified or incorrectly predicted by the previous model. This means each successive model will get a weighted input. It works on correcting the errors of previous models by building a new model on the error.

G. XGBoost

XGBoost is an ensemble Machine Learning algorithm that isdecision-tree-based and uses a gradient boosting framework. It handles the missing values and regularization for over- fitting. It Can be used to solve regression, classification.

H. ResNet50

Residual Neural Network or popularly known as ResNet-50 model is an Artificial Neural Network of a kind that stacks residual blocks on top of each other to form a 50 layers' deepconvolutional neural network. It is considered as a backbone for many computer vision tasks and is widely used for classification of images i.e., in our case classification of MRIbrain

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115153

scans into 4 different classes with input image of 224- by-224 in size.

VII. RESULT

A. Confusion Matrix

It presents a table layout of outcomes of predictions and results of a classification problem and helps us to visualize its outcomes with various parameters such as accuracy, MCC (Matthews correlation coefficient), F1 score,precision, recall etc. A confusion matrix consists of rows that are predicted values and columns are the actual values.

	+ve	**-ve**
+ve	True Positives(TP)	False Positives(FP)
-ve	False Negatives(FN)	True Negatives(TN)

B. Accuracy

It tells us the number of correct predictions made by an algorithm it is all the true values divided by the sum of total values i.e.,

TP/(TP+TN+FP+FN)

C. Precision

It is the ability of the model to classify correct values accurately. It is total positives (TP) divided by sum of all positives i.e., (TP+TN).

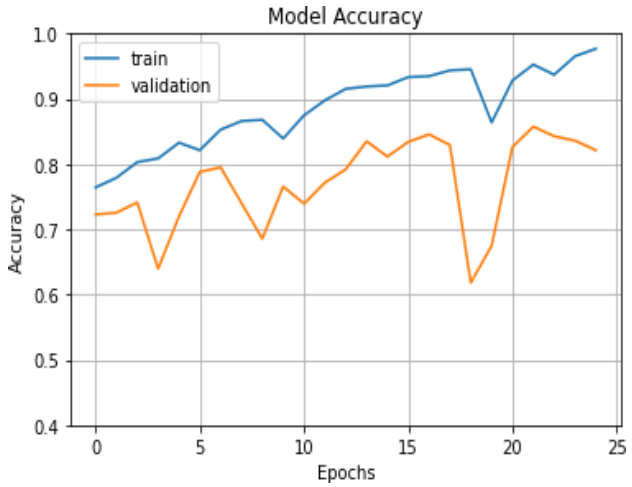

Fig. 9 Accuracy for ResNet50

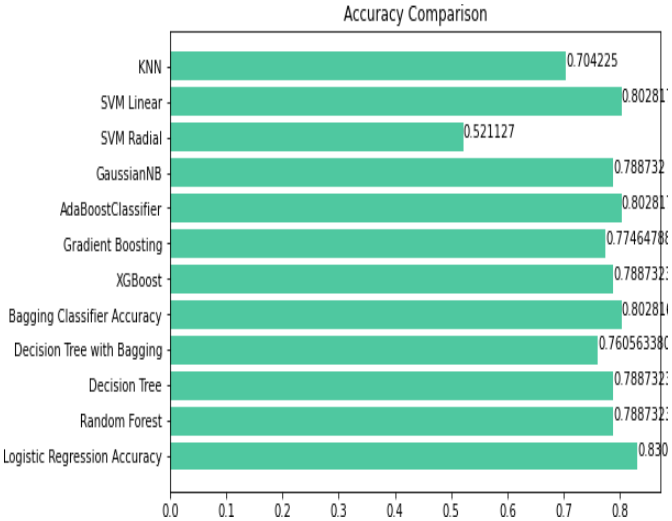

Fig. 10 Accuracy comparison for clinical dataset.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115153

SrNo	Algorithms	F1 Score	MCC	Precision	Recall	Accuracy
1.	Gradient BoostingClassifier	0.75	0.54	0.80	0.70	0.77
2.	XG Boost	0.76	0.58	0.82	0.70	0.78
3.	Bagging Classifier	0.78	0.57	0.77	0.79	0.78
4.	RFC	0.75	0.52	0.74	0.76	0.76
5.	Ada Boost Classifier	0.77	0.61	0.85	0.70	0.80
6.	Naive Bayes	0.72	0.61	0.95	0.58	0.78
7.	SVM Radial	0.00	0.00	0.00	0.00	0.52
8.	SVM Linear	0.75	0.63	0.95	0.61	0.80
9.	KNN	0.68	0.40	0.69	0.67	0.70
10.	Logistic Regression	0.79	0.68	0.95	0.67	0.83
11.	Decision Tree Bagging Classifier	0.72	0.52	0.81	0.64	0.76
12.	Decision Tree Classifier	0.71	0.63	1.0	0.55	0.78

Table 2 Prediction Results on clinical features.

VIII. CONCLUSION

In our work, we found that based on the clinical features related to brain volume and MMSE Logistic regression gives the best accuracy closely followed by Linear SVM and Adaboost classifier. The best precision is reported by Decision tree classifier and Bagging classifier performs the best on Both recall and F1 score. The accuracy for both clinical dataset and MRI dataset stands at 83%. Moreover, our approach clearly states that machine learning can be used very effectively for binary classification of electronic health records of patients with Alzheimer 's disease or Dementia with significant scope of improvement in the accuracies of both deep learning and shallow learning approaches.

As a limitations of the study, we have to use the available small size dataset (372 patients) for clinical dataset, a larger dataset from different geographical locations would have permitted us to research on a larger domains and obtain more reliable results and conclusions.

The entire model can be summed up in 2 parts where firstly by using clinical values one can get the chances of being affected by Alzheimer's and secondly by using the MRIscans the stage of the disease can be determined.
Regarding the future developments, we plan to apply our machine learning approach to alternative MRI datasets (2D and 3D) to figure out early detection of Alzheimer's.
The model can be used for prescribing medicines and other medical practices which are helpful in the case of Alzheimer's completely by using artificial intelligence in thenear future.

IX. REFERENCES

[1] J.Neelaveni and M.S.Geetha Devasana (2020). Alzheimer's Disease prediction using machine learning algorithm. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). 23 April 2020, Coimbatore, India, doi: 10.1109/ICACCS48705.2020.9074248

[2] Priyanka Lodha, Ajay Talele, & Kishori Degaonkar. Diagnosis of Alzheimer's Disease using Machine Learning. 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 16-18

DOI: 10.17148/IJARCCE.2022.115153

Aug. 2018, Pune, India, doi:10.1109/ICCUBEA.2018.8697386

[3] Sivakani. R and Gufran Ahmad Ansari, Machine Learning Framework for Implementing Alzheimer's Disease, 2020 International Conference on Communication and Signal Processing (ICCSP), 28-30 July 2020 Chennai, India, doi: 10.1109/ICCSP48568.2020.9182220.

[4] Aakash Shah, Dhruvi Lalakiya, Shekha Desai, Shreya & Vibha Patel, Early Detection of Alzheimer's Disease Using Various Machine Learning Techniques: A Comparative Study, 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI), June 1 2021, pp.522-526, doi: 10.1109/ICOEI48184.2020.9142975.

[5] A Chakraborty, NM de Wit, WM van der Flier and HE de Vries, "The blood brain barrier in Alzheimer's disease", Vasc Pharmacol., vol. 89, pp. 12-18, 2016.

[6] D.A. Nation, M.D. Sweeney, A. Montagne, A.P. Sagare, L.M. D'Orazio, M. Pachicano et al., "Blood-brain barrier breakdown is an early biomarker of human cognitive dysfunction", Nat Med., vol. 25, pp. 270-276, 2019.

[7] Kauppi Karolina and M Dale Anders, "Combining Polygenic Hazard Score With Volumetric MRI and Cognitive Measures Improves Prediction of Progression from Mild Cognitive Impairment to Alzheimer's Disease", Frontiers in Neuroscience, 2018.

[8] Min Li Liu, Wei Lan, Fang-Xiang Wu, Yi Pan and Jianxin Wang, "Classification of Alzheimer's Disease Using Whole Brain Hierarchical Network", IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 15, no. 2, March/April 2018.

[9] Jeffrey Cummings, "The National Institute on Aging—Alzheimer's Association Frameworkon Alzheimer's disease: Application to clinical trials", Elsevier, 2019.

[10] SJ Van der Lee, CE Teunissen, R Pool, MJ Shipley, A Teumer and V Chouraki, "Circulating metabolites and general cognitive abilityand dementia: Evidence from 11 cohort studies", Alzheimer's Dement2018, vol. 14, pp. 707-22.