iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.44978

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Diagnosis of Alzheimer's Disease using Machine Learning Algorithms

Dr. CH. Dhanunjaya Rao[1], G. Hari Charan[2], G. Mounika[3], K. Kavya Sri[4], P. Naveen Kumar Reddy[5], S. Leela Rama Krishna[6]

[1]*Professor, Department of CSE,* [2, 3, 4, 5, 6]*Students from Final year B. Tech, Aditya Institute of Technology and Management, Tekkali, Andhra Pradesh*

Abstract: ***Alzheimer's disease (AD) is one of the most common neurodegenerative diseases and is considered to be the main cause of cognitive impairment in elderly people. It is a progressive disease that destroys memory and other important mental functions and causes problems with memory, thinking and behavior. Symptoms usually develop slowly and worsen over time Symptoms may become severe enough to interfere with daily life, and lead to death. In 2022, 55 million people worldwide suffered from this disease. AD is predicted to affect 1 in 85 people globally by 2050, and at least 43% of prevalent cases need a high level of care. Alzheimer's Disease Neuroimaging Initiative (ADNI) give datasets that can be utilized for different Alzheimer's Disease related examinations. The dataset consists of a longitudinal MRI data of 150 subjects aged 60 to 96.72 of the subjects were grouped as 'Nondemented' throughout the study.64 of the subjects were grouped as 'Demented' at the time of their initial visits and remained so throughout the study.14 subjects were grouped as 'Nondemented' at the time of their initial visit and were subsequently characterized as 'Demented' at a later visit. These fall under the 'Converted' category.***
In our project, we propose some machine learning models to detect the Alzheimer's disease in earlier stage by finding the accuracy levels and determining the attributes that helps us to find the maximum accuracy rate.

I. INTRODUCTION

Alzheimer's disease is an irreversible, degenerative brain illness that gradually erodes memory and thinking skills, as well as the ability to do even the most basic tasks. Symptoms occur in the mid-60s in the majority of patients with the disease (those with the late-onset variety). Early-onset Alzheimer's disease is extremely rare and occurs between the ages of 30 and 60. The most common cause of dementia in elderly people is Alzheimer's disease. Memory issues are usually one of the early signs of Alzheimer's disease, though the severity of the symptoms varies from person to person. Other areas of thinking, such as finding the proper words, vision/spatial difficulties, and impaired reasoning or judgement, may also indicate Alzheimer's disease in its early stages. Mild cognitive impairment (MCI) is a disorder that can be a precursor to Alzheimer's disease, but not everyone with MCI will progress to the disease Alzheimer's patients have difficulty executing simple tasks such as driving a car, cooking a meal, or paying bills. They may repeatedly ask the same questions, become easily disoriented, misplace or misplace items, and find even simple things perplexing. Some people grow concerned, furious, or violent as the condition worsens. According to a 2022 World Health Organization survey, Alzheimer's disease affects an estimated 55 million people globally, with over 10 million new cases diagnosed each year. Alzheimer's disease is difficult to diagnose clinically, especially in its early stages. We want to improve diagnosis efforts with the use of classification tools. This research investigates some of the strategies for classifying Alzheimer's disease patients based on MRI and demographic data. Our study focused on using MRI biomarkers, demographic data, and cognitive tests gathered from subjects to train Machine Learning models and classify patients as either having AD or not. MRI biomarkers were largely obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database while some MRI biomarkers were taken from UC Berkeley Biomarkers. We used SVM, Decision Tree, Random Forest, Extra Tree Classifier, LGBM to output a predicted label of AD or Normal (Not AD) as well as finding the accuracy of machine learning algorithms and determining the attribute that helped us in giving a maximum accuracy rate

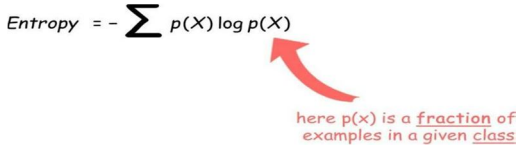
II. METHODOLAGY

There has been considerable research in field of Alzheimer's disease. A comprehensive algorithm has been implemented to predict Alzheimer's from available dataset that is extracted from the Kaggle Platform. The algorithms Support Vector Machine, Decision Tree, Extra Tree Classifier, Random Forest and LGBM are implemented. Deterioration is defined as "the scientific problem of the cognitive energy is symbolized by the serious global decrease in mental work is especially not because of adjustment in the carefulness.

©IJRASET: All Rights are Reserved

A. Implementation Of Machine Learning Algorithms

1) *Decision Tree:* Decision Tree is a widely used classification as well as Regression problem. It is a Supervised learning technique. In decision tree internal nodes represent the features of a dataset, branches depict the decision rules and each leaf node says the outcome.



2) *Support Vector Machine:* In machine learning algorithms, a support vector machine is a supervised learning algorithm. If you give support vector machine algorithm some labelled multiple groups. In the one-dimensional (1D) space, this classifier is called a point.in two-dimensional (2D) space, this classifier is called a line. In three-dimensional (3D) space, this classifier is called a plane. In four dimensional(4D) or more space, this classifier is called hyperplane.
3) *Random Forest:* Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.
4) *Extra Trees Algorithm:* Extremely Randomized Trees, or Extra Trees for short, is an ensemble machine learning algorithm. Spec9ifically, it is an ensemble of decision trees and is related to other ensembles of decision trees algorithms such as bootstrap aggregation (bagging) and random forest. The Extra Trees algorithm works by creating a large number of unpruned decision trees from the training dataset. Predictions are made by averaging the prediction of the decision trees in the case of regression or using majority voting in the case of classification. The predictions of the trees are aggregated to yield the final prediction, by majority vote in classification problems and arithmetic average in regression problems. Unlike bagging and random forest that develop each decision tree from a bootstrap sample of the training dataset, the Extra Trees algorithm fits each decision tree on the whole training dataset. Like random forest, the Extra Trees algorithm will randomly sample the features at each split point of a decision tree. Unlike random forest, which uses a greedy algorithm to select an optimal split point, the Extra Trees algorithm selects a split point at random. The Extra-Trees algorithm builds an ensemble of unpruned decision or regression trees according to the classical top-down procedure. Its two main differences with other tree-based ensemble methods are that it splits nodes by choosing cut-points fully at random and that it uses the whole learning sample (rather than a bootstrap replica) to grow the trees.
5) *Light Gradient Boosting Machine:* Light GBM is a more upgraded version of the Gradient boosting machine due to its efficiency and fast speed. Unlike GBM and XGBM, it can handle a huge amount of data without any complexity. On the other hand, it is not suitable for those data points that are lesser in number. Instead of level-wise growth, Light GBM prefers leaf-wise growth of the nodes of the tree. Further, in light GBM, the primary node is split into two secondary nodes and later it chooses one secondary node to be split. This split of a secondary node depends upon which between two nodes has a higher loss Hence, due to leaf-wise split, Light Gradient Boosting Machine (LGBM) algorithm is always preferred over others where a large amount of data is given.

B. Data Preprocessing

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always the case that we come across clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put it in a formatted way. So, for this we use data preprocessing tasks. Due to the amount of missing feature values in the original dataset, a preprocessing of the data was required. Firstly, any sample that has missing values is removed. Also, the features that have text labels are converted numerical labels. After removing these missing values, there are some very few (less than 10) sample points with additional labels. These are the cases where the labeling was changed in the dataset, for example labeling the character data into numeric form. In the data there are scan of subjects are 3 times from the dataset we have taken out the one visited scan. And there is and non-Converted people in the initial stage later on those people who are non-converted people converted to dementia.

C. Model Evaluation – Confusion Matrix

The performance of the classifier is usually evaluated using a confusion matrix. It is a specific table which is used to presents the true classes and classifier

predicted classes and illustrates the type of errors made by the classifier. Confusion matrix for binary classification (class '0' and class '1') is shown in Fig. 2.

There are four different terminologies used in the confusion matrix which are given as follows:

True Positive (TP) means classifier predicted positive ('1') and it is true ('1').

True Negative (TN) means classifier predicted negative ('0') and it is true ('0').

False Positive (FP) means classifier predicted positive ('1') and it is false ('0').

False Negative (FN) means classifier predicted negative ('0') and it is false ('1').

Figure 2: Confusion matrix for binary classification.

Calculated from a confusion matrix for a classification model. Accuracy means how often the classification model correctly classifies the data samples?

$$\text{Accuracy}=\frac{TP+TN}{ToTal} \quad (1)$$

Precision is the number of TP classes over the sum of both the TP classes and FP classes.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall is the number of TP classes over the sum of both the TP classes and FN classes.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

III. RESULTS AND DISCUSSIONS

Classifier	Accuracy
Decision Tree	75
Super Vector Machine	77
Random Forest	83
Extra Tree Classifier	87
Light Gradient Boosting Machine	80

IV. CONCLUSION

Alzheimer Disease which is also called as Senile Dementia. In a sentence, we can conclude that the ability of an individual to function independently is a continuous deterioration in thought, behavioral and mental abilities. Hence early detection of Alzheimer disease is necessary. The outcomes of this project will help us to detect the disease in earlier stages by finding the accuracy of machine learning algorithms and determining the attribute that helped us in giving a maximum accuracy rate. Some studies have suggested that characteristics of MRI may predict the rate of decline in AD and may guide future therapy. However, clinicians and researchers will need to use machine learning techniques that can accurately predict a patient's progress from mild cognitive impairment to dementia in order to reach that stage.

REFERENCES

[1] K.R.Kruthika, Rajeswari, H.D.Maheshappa, "Multistage classifier-based approach for Alzheimer's Disease prediction and retrieval", Informatics in Medicine Unlocked, 2019.

[2] Ronghui Ju , Chenhui Hu, Pan Zhou , and Quanzheng Li, "Early Diagnosis of Alzheimer's Disease Based on Resting-State Brain Networks and Deep Learning", IEEE/ACM transactions on computational biology and bioinformatics, vol. 16, no. 1, January/February 2019.

[3] Ruoxuan Cuia, Manhua Liu "RNN-based longitudinal analysis for diagnosis of Alzheimer's disease", Informatics in Medicine Unlocked, 2019.

[4] Fan Zhang , Zhenzhen Li , Boyan Zhang , Haishun Du , Binjie Wang , Xinhong Zhang, "Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease", NeuroComputing, 2019.

[5] Chenjie Ge , Qixun Qu , Irene Yu-Hua Gu , Asgeir Store Jakola "Multi-stream multi-scale deep convolutional networks for Alzheimer's disease detection using MR images", NeuroComputing, 2019.

[6] Tesi, N., van der Lee, S.J., Hulsman, M., Jansen, I.E., Stringa, N., van Schoor, N. et al, "Centenarian controls increase variant effect sizes by an average twofold in an extreme caseextreme control analysis of Alzheimer's disease", Eur J Hum Genet. 2019;27:244–253

[7] J. Shi, X. Zheng, Y. Li, Q. Zhang, S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease", IEEE J. Biomed. Health Inform., vol. 22, no. 1, pp. 173- 183, Jan. 2018.

[8] M. Liu, J. Zhang, P.-T. Yap, D. Shen, "Viewaligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data", Med. Image Anal., 2017 vol. 36, pp. 123- 134.

[9] Hansson O, Seibyl J, Stomrud E, Zetterberg H, Trojanowski JQ,Bittner T, "CSF biomarkers of Alzheimer's disease concordwith amyloid-bPET and predict clinical progression: A study of fullyautomated immunoassays in BioFINDER and ADNI cohorts". Alzheimers Dement 2018;14:1470–81.

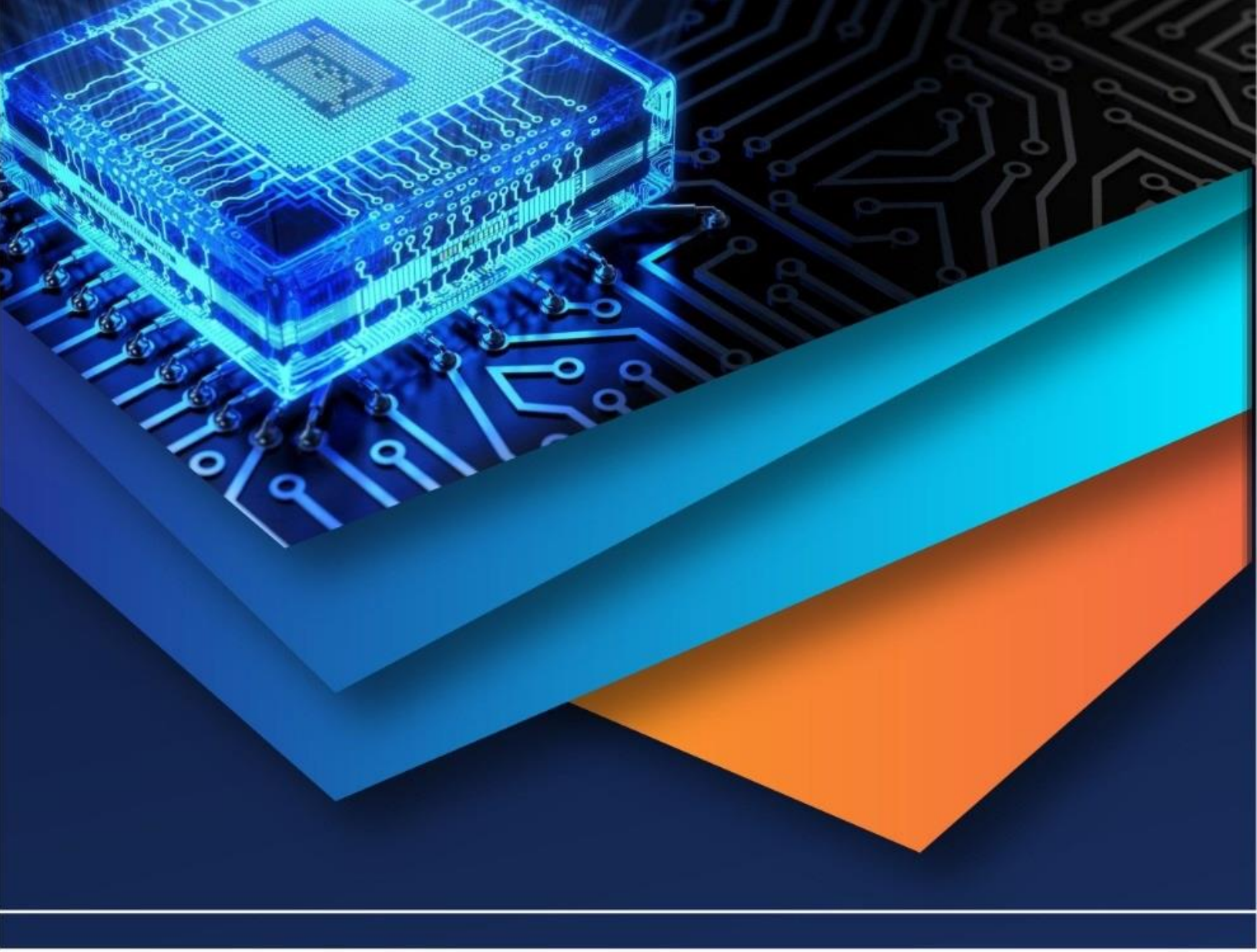
[10] Van der Lee SJ, Teunissen CE, Pool R, Shipley MJ, Teumer A,Chouraki V, "Circulating metabolites and general cognitive abilityand dementia: Evidence from 11 cohort studies", Alzheimer's Dement2018;14:707–22

[11] Kauppi Karolina, Dale Anders M, "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.

[12] Grassi M, Loewenstein DA, Caldirola D, Schruers K, Duara R, Perna G, "A clinically-translatable machine learning algorithm for the prediction of Alzheimer's disease conversion: further evidence of its accuracy via a transfer learning approach", Int Psychogeriatr, 2018 14:1–9. doi: 10.1017/S1041610218001618.

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

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)