Prediction of Alzheimer's Disease (AD) using Machine Learning Classifier Models: A Comparative Analysis

Avishkar Behera[1]

Abstract— **Alzheimer's Disease or AD is a neurological brain disorder that destroys memory and mental functions. It is one of the most common causes of dementia and is usually found in older patients. It is a progressive disease affecting millions worldwide and very few people who are suffering from AD are diagnosed correctly and in a timely manner. Hence, early and accurate prediction of AD is largely beneficial in battling dementia and delaying symptoms of the disease.**

This study aims to accurately predict Alzheimer's Disease (AD) in the early stages by making use of machine learning techniques. Specifically, we will be using Tree Based Classifiers: Decision Tree, Random Forest, AdaBoost, Gradient Boost, & XGBoost Classifiers as well as Non-Tree Based Classifiers: SVM, Logistic Regression, k-Nearest Neighbours, Neural Network, & Voting Classifiers to classify test subjects into demented or nondemented groups. We will compare the performance of these models based on their Accuracy, Precision, Recall, AUC, F1 Score and Matthews Correlation Coefficient (MCC). The dataset used is the Open Access Series of Imaging Studies (OASIS) dataset, in which the Soft Voting Classifier gave the best results having an accuracy of 0.947, AUC score of 0.953 and MCC of 0.898. We conclude that the involvement of ensemble modeling machine learning classifiers like the Voting Classifier can predict with better accuracy the development of AD at an early stage.

I. INTRODUCTION

Dementia is an umbrella term for a set of symptoms, including memory loss, confusion, and difficulty with problem-solving, caused by damaged brain cells. Alzheimer's disease or AD is the most common form of dementia, accounting for approximately 60-70% of cases. Alzheimer's disease is marked by the accumulation of abnormal protein deposits, known as beta-amyloid plaques and tau tangles, in the brain. These deposits interfere with communication between nerve cells, leading to the gradual deterioration of brain function. The early stages of Alzheimer's are often characterized by mild forgetfulness, but as the disease progresses, individuals may experience confusion, disorientation, and difficulty with language and problem-solving.

Even with the advancements in medicine and technology, there is currently no cure for Alzheimer's Disease (AD). Scientists don't yet fully understand the cause for AD and research is still ongoing in this area. In 2020, close to 7 million people were diagnosed with AD in the US alone and this number is expected to double every 20 years. It is the 6th leading cause of death among US adults and about 1 in 3 people over 65 years die with AD or another form of dementia, hence, it is a major global issue.

[1]Undergraduate Student, Electrical Engineering, IIT Hyderabad ee21btech11005@iith.ac.in

Early detection of AD provides an opportunity for timely medical intervention, which can help manage symptoms and slow the progression of the disease. While there is no cure for Alzheimer's, certain medications and therapies can alleviate symptoms and enhance cognitive function, especially in the early stages. Initiating these interventions promptly can contribute to maintaining independence and preserving cognitive abilities for a more extended period.

Early detection also plays a crucial role in advancing research and treatment development for Alzheimer's disease. By identifying individuals in the early stages of the disease, researchers can gain valuable insights into the disease process and test potential interventions more effectively. This can accelerate the development of more effective treatments and even pave the way for a cure in the future.

There is no simple model to screen for Alzheimer's disease, partly because the diagnosis of Alzheimer's disease itself is complex—typically involving expensive and sometimes invasive tests not commonly available outside highly specialised clinical settings.

Hence, as part of this study, we attempt to employ machine learning techniques for early, inexpensive and accurate prediction of Alzheimer's disease. Our goal is to find the optimal machine learning model for classifying a patient into demented or nondemented groups, which is a classification problem. Therefore, we will be comparing various popular Tree and Non-Tree Based Classifiers to achieve our objective of finding the best classifier model.

II. LITERATURE REVIEW

Kevin de Silva et al.[1] employed convolutional neural networks on magnetic resonance imaging data from one single central slice of the brain to predict Alzheimer's disease and the best performance obtained was MCC: 0.77, Accuracy: 0.89, F1: 0.89, AUC: 0.92. Aniverthy Amrutesh et al.[2] used transfer learning models on MRI images to predict Alzheimer's disease. Raveendra Reddy et al.[3] achieved 0.98 accuracy using Visual Geometry Group (VGG)-16 et Improved Faster Recurrent Convolutional Neural Network (IFRCNN) method on CT scan images. Martinez-Murcia et al.[4] used deep convolutional autoencoders to explore data analysis of Alzheimer's disease. K.R. Kruthika et al.[5] employed multi-stage classifiers along with feature selection on the Alzheimer's Disease Neuroimaging Institute (ADNI) dataset and obtained an accuracy of 0.963 ± 0.012. Francesco Amenta et al.[6] used wrapping technique for feature selection on four ML models: Naive Bayes, Artificial Neural Network, K-Nearest Neighbour and SVM and

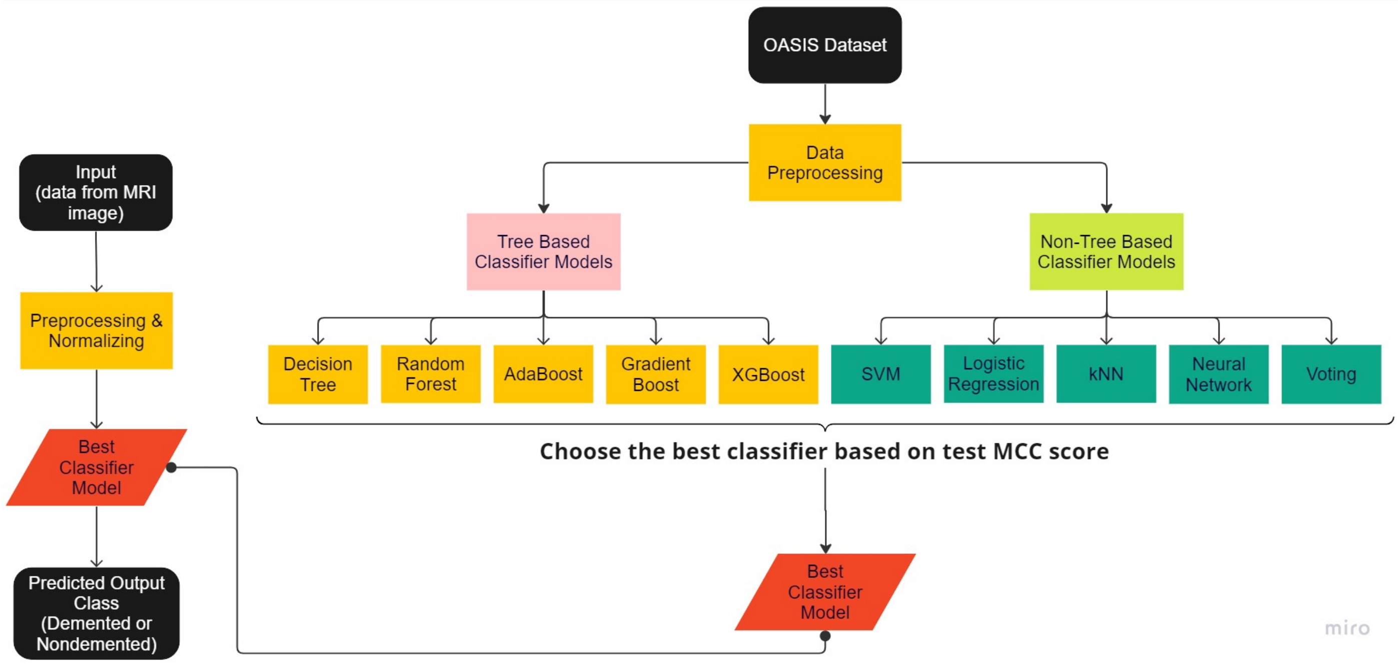


Fig. 1: Proposed Workflow

achieved accuracy of 0.980 with AUC of 0.991.

III. Materials and Methods

The dataset is publicly available from Open Access Series of Imaging Studies (OASIS) website (https://www.oasis-brains.org/#data). The OASIS-2 dataset consists of longitudinal MRI data in Nondemented and Demented Adults.

A. Proposed Workflow

Fig.1 summarises the proposed workflow. First, we will perform an initial analysis of our OASIS dataset and preprocess the data accordingly. After preprocessing, we will split the dataset in 80:20 ratio for training (supervised learning) and testing our classifier models respectively. We will then choose the model which gives the best performance (best MCC score) and output its predictions on the sample test set.

We will do a comparative study of the following ML Classifier models -

a) Tree Based Models - Decision Tree, Random Forest, AdaBoost Classifier, Gradient Boosting Classifier, XGBoost Classifier.

b) Non-Tree Based Models - Support Vector Machine, Logistic Regression Classifier, k-Nearest Neighbours Classifier, Neural Network, Voting Classifier (both soft and hard).

B. Data Description

The dataset consists of a longitudinal collection of 150 subjects aged 60 to 96. Each subject was scanned on two or more visits, separated by at least one year for a total of 373 imaging sessions. For each subject, 3 or 4 individual T1-weighted MRI scans obtained in single scan sessions are included.

The subjects are all right-handed and include both men and women. 72 of the subjects were characterized as nondemented throughout the study. 64 of the included subjects were characterized as demented at the time of their initial visits and remained so for subsequent scans, including 51 individuals with mild to moderate Alzheimer's disease. Another 14 subjects were characterized as nondemented at the time of their initial visit and were subsequently characterized as demented at a later visit.

The data from the MRI image was extracted as follows: The MRI images were corrected for interscan head rotation and wrapped spatially into atlas space. The transformation outcome placed the brains in a correlated coordinate system, with the bounding box as the actual atlas. With this procedure, every image was turned out as a unique, high contrast, averaged MP-RAGE image in an atlas-space. The insight explanation on image acquisition and postprocessing steps are detailed in [7].

The estimated total intracranial volume (eTIV) was defined manually across intracranial volume on an atlas. Normalized whole-brain volume (nWBV) was computed with the FAST program of the FSL software suite. Image segmentation was done to classify brain tissue as spinal fluid or white or gray matter. This segmentation process was iteratively assigned as voxels to tissue classes based on high probability estimates of hidden Markov random field models. In the end, nWBV was calculated as the proportion of accumulated voxels across the brain mask, and the normalized volume was expressed in a percentage of total gray and white matter voxels of eTIV. [7]

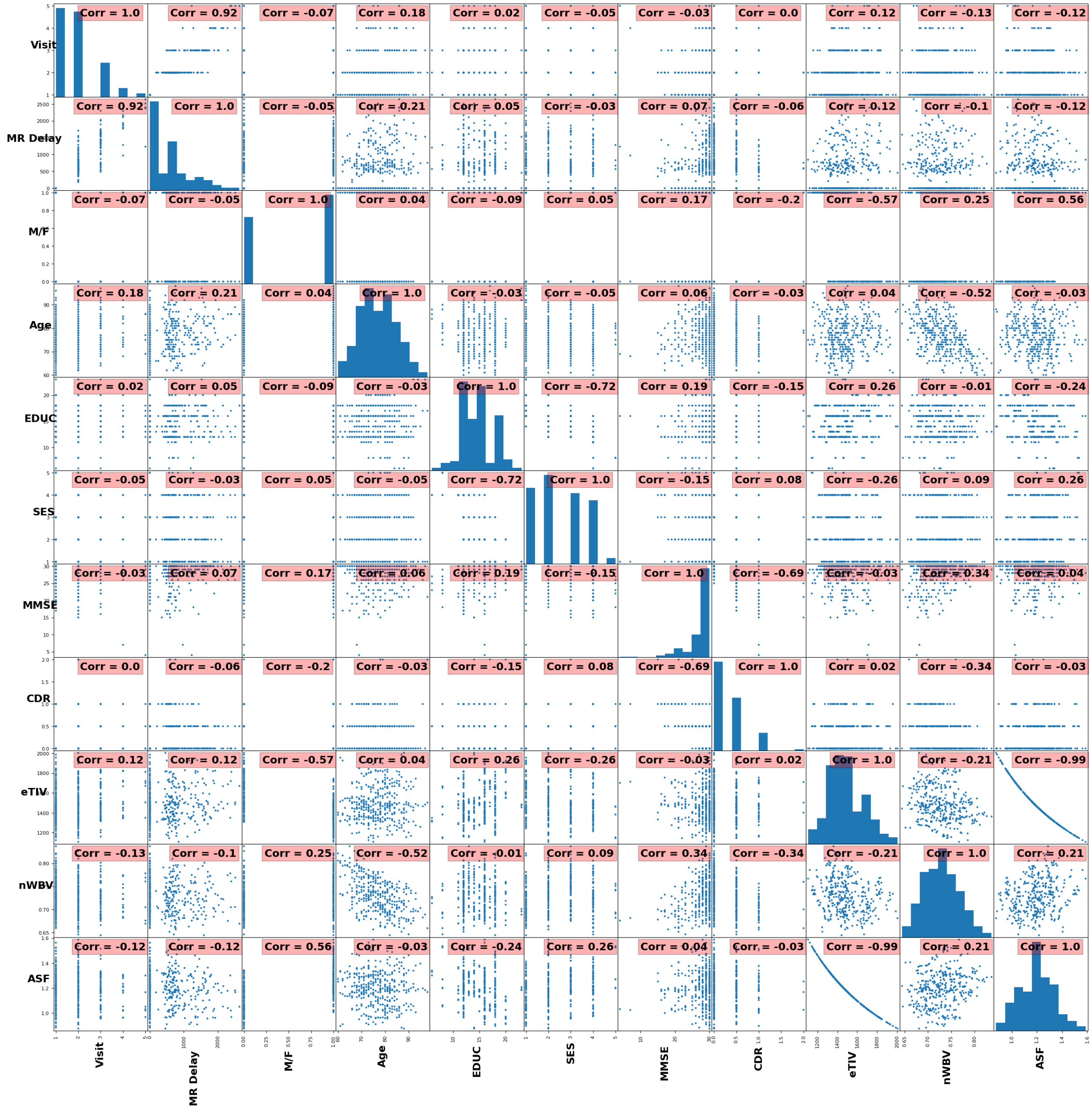


Fig. 2: Scatter matrix of correlation between variables

The variables in the dataset are:

MR Delay : Delay time prior to the image procurement

M/F : Gender, Male or Female

Hand : Right Handedness or Left Handedness

EDUC : Years of Education

SES : Socioeconomic Status ranging from 1 (highest status) to 5 (lowest status)

MMSE : Mini Mental State Examination score ranging from 0 (worst) to 30 (best), 30-point questionnaire that has been shown to be valid and reliable in identifying dementia

CDR : Clinical Dementia Rating (0 = normal, 0.5 = very mild Dementia, 1 = mild Dementia, 2 = moderate Dementia, 3 = severe Dementia)

eTIV : Estimated Total Intracranial Volume

nWBV : Normalized Whole Brain Volume, reflecting the percentage of the intracranial cavity occupied by brain

ASF : Atlas Scaling Factor, allows for comparison of the estimated total intracranial volume (eTIV) based on differences in human anatomy

Group : Patient is Demented, Converted or Nondemented

C. Data Analysis

We conduct an initial analysis of the data to find relations between any variables. Fig.2 shows the correlation matrix between all the variables in the dataset. We observe that ASF (Atlas Scaling Factor) and eTIV (Estimated Total Intracranial

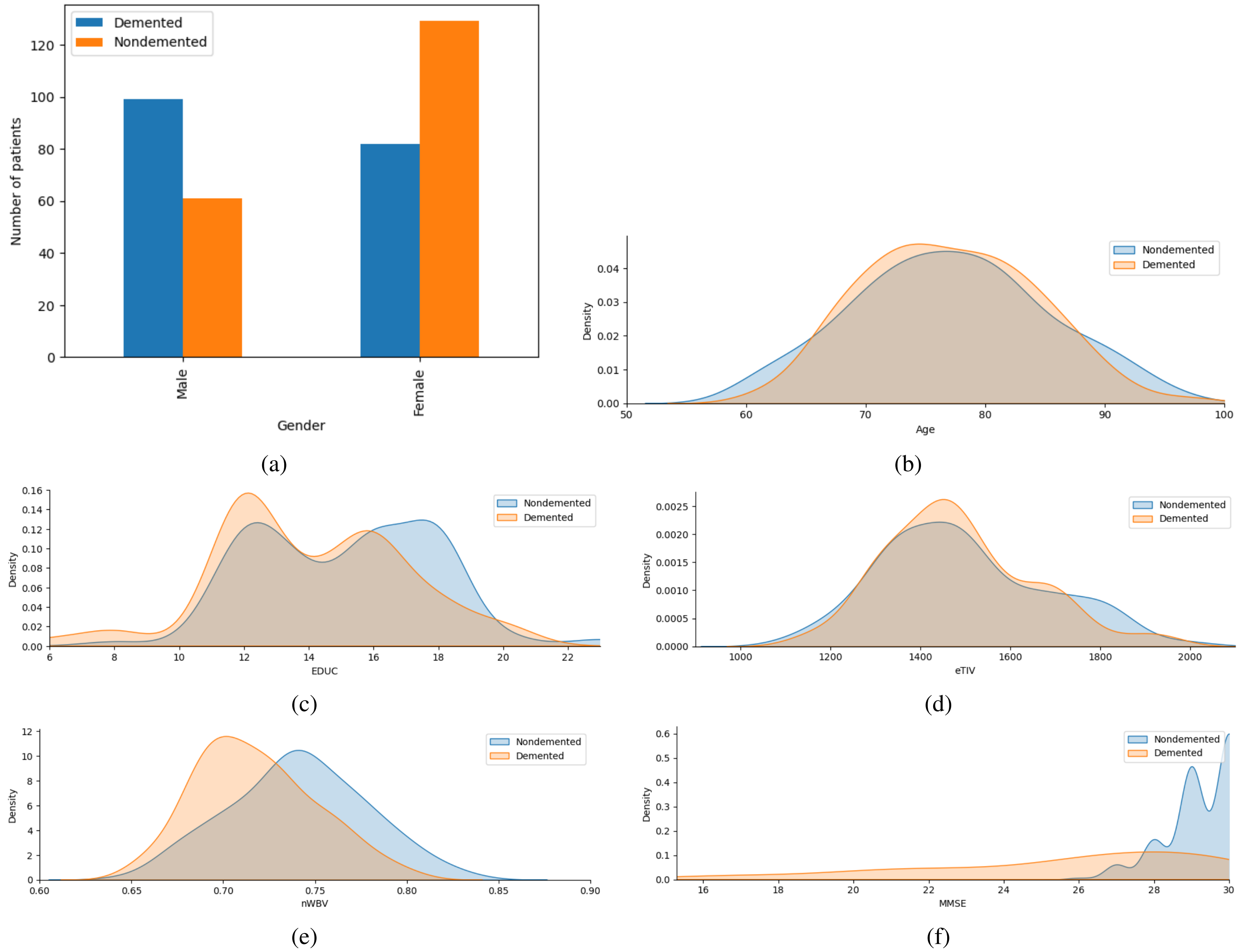


Fig. 3: (a) Gender vs Dementia; (b) Age vs Dementia; (c) EDUC vs Dementia; (d) eTIV vs Dementia; (e) nWBV vs Dementia; (f) MMSE vs Dementia

Volume) have very high correlation (≈ -1). Hence, we drop the ASF column from our dataset as it could bias our machine learning models. Also, in the Hand column, all the patients are Right-Handed (marked R in the dataset). Hence, we can drop this column too.

Fig.3 shows the distribution of patients with the variables in our dataset. We observe that dementia/AD is more prevalent in Males than Females from (a). From (c), we can see that demented patients tend to have lesser years of education. Demented patients also tend to have lesser brain volume as seen in (e). There is no obvious relation between age/eTiv and dementia as observed in (b) and (d). There is a very clear connection in the MMSE scores and dementia as seen in (f). Demented patients have lower and widespread scores (15-30) in the MMS Examination whereas, nondemented patients' MMSE scores are concentrated in the range 27-30. Hence, the MMSE score is an important feature for training our ML models.

The dataset contains 19 rows with missing values in the SES column and 2 rows with missing values in the MMSE column. We cannot delete these 19 missing entries because our dataset is already small, hence, we perform imputation for these values by using the median of the available SES values. However, for the 2 missing MMSE values, we have no choice but to delete these entries as MMSE is an important feature for prediction as seen above.

We will split this dataset in 80:20 ratio, where 80% will be used for training (supervised) our ML classifier models and 20% will be used for testing these models.

D. Machine Learning Classifier Models (Tree Based)

1) Decision Tree Classifier: Decision Tree is a supervised machine learning technique, commonly used for classification problems. It is binary tree-structured classifier where a decision is taken at each internal node. The internal nodes represent the features of the dataset, the branches represent the decision rules and the leaf nodes represent the decision outcomes. The attributes for each decision node are chosen by Information Gain and Gini Impurity.

$$Information\ Gain = Entropy_{parent} - avg(Entropy_{children})$$

$$Entropy = -\sum_{i=1}^{N} p_i \, log_2 p_i$$

$$Gini\ Impurity = \sum_{i=1}^{N} p_i \,(1 - log_2 p_i)$$

where p_i is the probability of randomly selecting a data point belonging to class i and N is the number of classes.

2) Random Forest Classifier: Random Forest is a popular supervised machine learning algorithm that aggregates the output of many decision trees on various subsets of the dataset to give a single accurate prediction. Greater number of decision trees in the forest gives better accuracy and makes the model less prone to overfitting. The random forest algorithm utilizes bagging as its ensemble technique to generate random subsets from the dataset.

3) AdaBoost Classifier: AdaBoost or Adaptive Boosting Classifier is an iterative ensemble boosting classifier which combines multiple weak classifiers to get a strong classifier which gives best accuracy. It begins by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases.

4) Gradient Boost Classifier: Gradient Boost Classifier is an additive classifier model in which instead of fitting a predictor on the data at each iteration, it actually fits a new predictor to the residual errors made by the previous predictor. The predictors usually used are decision trees (called decision stumps). At each stage, N (number of classes) regression trees are fit on the negative gradient of the loss function. The Loss Function in gradient boosting is of the form:

$$L = -y * log(odds) + log(1 + e^{log(odds)})$$

where y are the observed values.

The Residual Output is calculated by:

$$Residual = Observed\ Output - Predicted\ Output$$

The output of the leaves is transformed by:

$$\gamma = \frac{\sum Residual}{\sum PreviousProb * (1 - PreviousProb)}$$

The formula for making predictions is as follows:

$$OldTree + Learning\ Rate * NewTree$$

5) XGBoost Classifier: XGBoost or Extreme Boosting Classifier is an optimized gradient boosting classifier model. It creates decision trees sequentially and weights are assigned to all the independent variables which are then fed into the decision tree which predicts results. The weight of variables predicted wrong by the tree is increased and these variables are then fed to the second decision tree. These individual classifiers/predictors then ensemble to give a strong and more precise model.

E. Machine Learning Classifier Models (Non-Tree Based)

1) Support Vector Machine: Support Vector Machine (SVM) is a powerful supervised machine learning model used widely for classification problems. The objective of SVM is to find the optimal hyperplane (with the help of support vectors) in an N-dimensional space that best separates the two classes. The algorithm tries to maximise the margin the between two support vectors i.e. the margin between the two closest points of different classes.

The SVM algorithm minimises $\mathbf{w}, b, \zeta$ according to

$$||\mathbf{w}||^2 + C\sum_{i=1}^{n} \zeta_i$$

subject to the following constraints

$$y_i(\mathbf{w}^T \mathbf{x}_i - b) \geq 1 - \zeta_i \ \ and \ \ \zeta_i \geq 0 \ \ \forall i \in \{1, 2, ..., n\}$$

2) Logistic Regression Classifier: Logistic Regression is a classification technique used in machine learning primarily to model the dependent variable, which is dichotomous in nature (2 classes) like in our dataset. It uses the logistic function or sigmoid function for modeling.

The cost function in LR Classifier is:

$$Cost = -\frac{1}{N}\sum_{i=1}^{N} y_i \, log(h\theta(y_i)) + (1 - y_i)\, log(1 - h\theta(y_i))$$

where y_i are the observed values and $h\theta$ is the Sigmoid function i.e.

$$h\theta(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

We make use of Gradient Descent to minimise our cost function which estimates the parameters β_0 and β_1 of our model.

3) k-Nearest Neighbours Classifier: The k-Nearest Neighbors (kNN) algorithm is a simple yet robust machine learning technique that is used to solve both regression and classification issues. The kNN algorithm locates the k closest neighbors, using a distance metric like Euclidean distance, to a given data point. The majority vote or the average of the k neighbors is then used to establish the class or value of the data point.

4) Neural Network: Neural Network is a deep learning model that is used to model complicated patterns and give accurate predictions. The Neural Network we will be using is a Multi-Layer Perceptron (MLP) Classifier. It consists of interconnected nodes (neurons) organized into three layers - input layer, hidden layer and output layer. Information flows through these nodes, and the network adjusts the weights during training to learn from data, enabling it to recognize patterns and make predictions. A node/neuron in the MLP Classifier is modeled as:

$$f(b + \sum_{i=1}^{n} x_i w_i)$$

where x_i are the input, w_i are their corresponding weights, b is the bias and f is the activation function.

5) Voting Classifier: Voting Classifier is a machine learning model that trains on an ensemble of numerous models and predicts an output (class) based on their highest probability of the chosen class as the output. It simply aggregates the findings of each classifier passed into Voting Classifier and predicts the output class based on the highest majority of voting. The idea is instead of creating separate dedicated models and finding the accuracy for each of them, we create a single model which trains by these models and predicts output based on their combined majority of voting for each output class. For our Voting Classifier model, we will feed all the above models (both tree and non-tree based) as estimators.

Soft Voting is when the output class prediction is chosen based on the average probability given to that class. Whereas, Hard Voting is when the output class prediction is chosen based on the highest majority of votes given to that class by the classifiers.

F. Outcome Prediction

We will find the best parameters for each model using the training set and also perform 5-fold cross-validation which ensures that there is no overfitting in any model. After choosing the best parameters, the models will then be evaluated on Accuracy, Precision, Recall, AUC, F1 Score, and MCC to find the best classifier model among these.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$AUC = Area\ under\ the\ ROC\ curve$$

$$F1\ Score = 2 * \frac{Precision\ *\ Recall}{Precision\ +\ Recall}$$

$$MCC = \frac{TP*TN - FP*FN}{\sqrt{(TN+FN)(TP+FP)(TN+FP)(TP+FN)}}$$

where TP : True Positive, TN : True Negative, FP : False Positive and FN : False Negative.

Matthews Correlation Coefficient or MCC is a single-value performance metric that summarizes the confusion matrix, which is usually used for binary classification problems. It ranges from -1 to +1, where 0 corresponds to random classification, +1 corresponds to perfect classification and -1 corresponds to perfect opposite classification. Metrics like F1 Score ignore the count of True Negatives, whereas MCC extends to all four entries of the confusion matrix. A high MCC always corresponds to a high AUC, but the converse is not true. MCC is high only if the classifier is doing well on both the negative and the positive elements, hence, it gives a more balanced assessment of the classifier[8]. We will be using MCC as our main performance metric while comparing various classifiers.

IV. Experimental Results

We begin by training all our classifier models on the train split of the dataset and we find the best parameters for each model based on their accuracies and MCC. The following were the best parameters obtained:

Decision Tree: max depth = 6

Random Forest: trees=20, max features=1, max depth=7

AdaBoost: estimator=DecisionTree, trees=30, learn rate=1

Gradient Boost: trees = 60, learn rate = 1, max depth = 5

XGBoost: trees = 90, learn rate = 0.1, max depth = 7

Support Vector Machine: kernel = RBF, C = 1000

Logistic Regression: C = 100

k-Nearest Neighbours: k = 1

Neural Network: activation function=ReLU, hidden layers=2, nodes per layer=250

Soft Voting: estimators = all the above models

Hard Voting: estimators = all the above models

Using the above parameters, we obtain the performance metrics as seen in the graph and table in Figs. 4(a), 4(b) respectively. The Soft Voting Classifier obtained the highest accuracy of 0.947, perfect precision of 1, highest recall of 0.907, highest AUC score of 0.953, highest F1 score of 0.951 and highest MCC of 0.898.

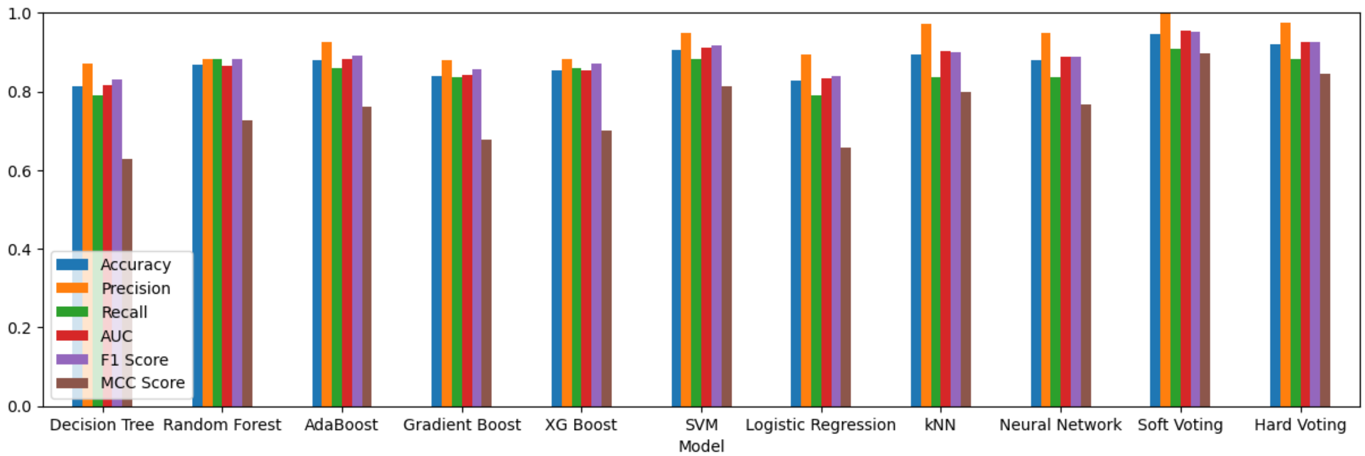
The SVM, Soft and Hard Voting Classifiers gave the top 3 highest performance metrics with MCC greater than 0.8, whereas the Decision Tree, Gradient Boost and Logistic Regression Classifiers gave the 3 lowest performance metrics with less than 0.7 MCC. Also, the Non-Tree Based Classifiers tend to give better performance than the Tree Based Classifiers in general.

Fig.5 shows the predictions for a few samples of the test data by the Soft Voting Classifier model.

V. Discussion

ML techniques have been applied to various types of data, including clinical and neuropathological research data, MRI brain imaging, and even pathological speech of AD patients[9]. Here, we have attempted to predict AD using basic ML Classifier models. Based on our study, we can conclude that Non-Tree Classifiers perform better than Tree Classifiers. Moreover, the best results are obtained by taking an ensemble of numerous models like the Voting Classifier in our study. We have achieved an outstanding accuracy of 0.947 and MCC of 0.898 using just basic data from an MRI scan (without using the MRI images) and psychological parameters like MMSE. Out of all the reviewed research papers using this same dataset, we have obtained the highest accuracy, AUC and MCC.

The Soft Voting Classifier takes around 10 seconds to train and around 1 second to generate predictions on a standard computer without any GPU or graphics card. Hence, we have developed a fast, inexpensive and accurate way to predict AD in older patients. Revisiting our objective of finding the optimal ML model, we have found that the Soft Voting Classifier with its ensemble of estimator Tree and Non-Tree models is the optimal ML model for classifying a patient into demented or nondemented groups, but more accurate



(a)

Type	Model	Accuracy	Precision	Recall	AUC	F1 Score	MCC
Tree	Decision Tree	0.81333333	0.87179487	0.79069767	0.81722384	0.82926829	0.62808941
	Random Forest	0.86666667	0.88372093	0.88372093	0.86373547	0.88372093	0.72747093
	AdaBoost	0.88	0.925	0.86046512	0.88335756	0.89156627	0.76011486
	Gradient Boost	0.84	0.87804878	0.8372093	0.84047965	0.85714286	0.67654858
	XG Boost	0.85333333	0.88095238	0.86046512	0.85210756	0.87058824	0.70167006
Non-Tree	SVM	0.90666667	0.95	0.88372093	0.91061047	0.91566265	0.81415147
	Logistic Regression	0.82666667	0.89473684	0.79069767	0.83284884	0.83950617	0.65855734
	kNN	0.89333333	0.97297297	0.8372093	0.90297965	0.9	0.79731451
	Neural Network	0.88	0.94736842	0.8372093	0.88735465	0.88888889	0.7663997
	Soft Voting	**0.94666667**	**1**	**0.90697674**	**0.95348837**	**0.95121951**	**0.89788727**
	Hard Voting	0.92	0.97435897	0.88372093	0.92623547	0.92682927	0.8439277

(b)

Fig. 4: (a) Graph of model performance metrics; (b) Table showing model performance metrics (the maximum value for each metric has been highlighted in green)

and complex (but slower) Deep Learning models exist which is beyond the scope of this study and can be explored in the future.

In our study, we have used only data from the MRI images and psychological parameters and not the MRI images themselves. But as seen in the literature review, using the MRI images along with ML techniques gives better accuracies of around 0.96 to 0.98. Hence, future work can be done by combining both brain MRI images as well as psychological parameters to predict the disease with higher accuracy in the earlier stage itself. Additionally, future research could focus on the integration of ML techniques with other diagnostic tools and biomarkers to create a more comprehensive and holistic approach to AD prediction.

Our study does not consider younger patients because the OASIS dataset contains data of only older adults of age greater than 60 years, hence, future research could also include data from the younger population which would significantly enhance the ability of the model in early detection of AD. In conclusion, early and accurate prediction has the potential to aid clinicians and researchers in better understanding the disease and developing personalized treatment plans for patients, and our study contributes a significant step towards this research.

All the materials and code used in this study are available on my GitHub using the following clickable URL: AvishkarB/Alzheimers_Disease_Prediction

References

[1] K. de Silva and H. Kunz, "Prediction of alzheimer's disease from magnetic resonance imaging using a convolutional neural network," *Intelligence-Based Medicine*, vol. 7, p. 100091, 2023.

[2] A. Amrutesh, C. G. Gowtham Bhat, A. Amruthamsh, K. Asha Rani, and S. Gowrishankar, "Alzheimer's disease prediction using machine learning and transfer learning models," in *2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, pp. 1–6, 2022.

[3] R. Reddy and R. Krishna, "Alzheimer' disease prediction and classification using ct images through machine learning," *Bratislavske lekarske listy*, 2023.

[4] F. Martinez-Murcia, A. Ortiz, J. Gorriz, J. Ramírez, and D. Castillo-Barnes, "Studying the manifold structure of alzheimer's disease: A deep learning approach using convolutional autoencoders," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1–1, 06 2019.

[5] K. Kruthika, Rajeswari, and H. Maheshappa, "Multistage classifier-based approach for alzheimer's disease prediction and retrieval," *Informatics in Medicine Unlocked*, vol. 14, pp. 34–42, 2019.

M/F	Age	EDUC	SES	MMSE	eTIV	nWBV	Test Values	Predicted Values
1	84	13	2	28	1402.1	0.694679	1	1
0	74	18	2	30	1636.08	0.679806	0	0
0	87	14	2	27	1986.55	0.696106	0	0
0	62	12	4	17	1525.34	0.73225	1	1
0	76	16	3	30	1832.1	0.768647	0	0
0	66	16	1	21	1707.73	0.702767	1	1
0	85	12	4	30	1699.269	0.705081	0	0
1	78	8	5	29	1382.69	0.755658	0	0
0	88	12	4	26	1482.59	0.709476	1	1
1	76	12	2	27	1316.37	0.72685	1	1
1	88	16	3	30	1294.81	0.743776	0	0
1	78	8	5	23	1462.4	0.690761	1	1
1	72	11	4	21	1489.19	0.685979	1	1
1	84	12	2	27	1389.82	0.727641	1	1
0	78	12	4	29	1505.65	0.714846	0	1
0	82	16	1	28	1692.88	0.693926	1	1
0	77	16	3	16	1589.79	0.696332	1	1
1	94	23	1	29	1474.35	0.696004	0	0
1	61	16	3	30	1312.78	0.804868	0	0
0	73	12	2	23	1661.27	0.697563	1	1
1	68	18	2	29	1289.679	0.795234	0	0
1	87	18	1	24	1275.18	0.682665	1	1
1	80	16	1	30	1450.29	0.698444	0	0
1	85	12	4	29	1224.664	0.70955	0	0
0	85	17	1	30	1724.32	0.703522	0	0
0	68	15	2	30	1556.42	0.713287	1	1
1	85	15	2	30	1487.646	0.740513	0	0
1	73	16	3	29	1286.992	0.770652	1	0
0	86	12	3	27	1813.215	0.760823	0	0

Fig. 5: Table showing predicted and actual class for a test sample by Soft Voting Classifier model where 1 means Demented and 0 means Nondemented (correct predictions are highlighted in green and wrong predictions are highlighted in red)

[6] F. Amenta, G. Battineni, N. Chintalapudi, and E. Traini, "A comprehensive machine-learning model applied to magnetic resonance imaging (mri) to predict alzheimer's disease (ad) in older subjects," *Clin Med.*, July 2020.

[7] D. Marcus, A. Fotenos, J. Csernansky, J. Morris, and B. RL, "Open access series of imaging studies: longitudinal mri data in nondemented and demented older adults," *Cogn Neurosci.*, Dec 2010.

[8] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation," *BMC Genomics 21*, 2020.

[9] L. Hason and S. Krishnan, "Spontaneous speech feature analysis for alzheimer's disease screening using a random forest classifier," *Frontiers in Digital Health*, vol. 4, 2022.