Analysis of Machine Learning Algorithms for MRI Based Alzheimer's Disease Dementia Classification

Arsya M, *CB.EN. P2DSC21006, 1st year MTech – Data Science, Batch* (2021-23),
Course:21DS602(21 − 22(Odd)), email: cb.en.p2dsc21006@cb.students.amrita.edu

Abstract

Worldwide 1 in 6 people are affected by brain diseases and it includes a wide spectrum of diseases and disorders caused from stroke and Alzheimer's to epilepsy, traumatic brain injury, and many more. Dementia is one such condition caused by damage to or loss of nerve cells in the brain that affects memory and several cognitive abilities. There is often less awareness and understanding about dementia resulting in stigmatization and barriers to diagnosis and care. Most of the time this illness is subjected to misclassification and result in over-delaying which increases the possible prognosis of dementia to a more severe stage. This work focuses on using different Machine Learning algorithms such as Logistic Regression, K-Nearest Neighbor(KNN), Decision Tree, Gaussian Naive Bayes, SVM(Support Vector Machine), and ensemble algorithms like stacking, Bagging(Random Forest), and Boosting (AdaBoost & XGBoost) on the longitudinal collection of T1-weighted Magnetic Resonance Imaging (MRI) scan features available in Open Access Series of Imaging Studies (OASIS) dataset to predict and classify the patients as demented or non-demented. In this work, the above-mentioned algorithms were evaluated to find which algorithms perform best with good accuracy and less misclassification that is high F1-score. As a result, it is been observed that XGBoost performed best at 10-fold stratified cross-validation with 95% test accuracy and a 0.95 test F1 score and also Logistic Regression, Gaussian Naïve Bayes, and Stacked Model algorithms had 95% test accuracy F1-score of 0.94.

Index Terms

Accuracy, Bagging, Boosting, Dementia, Ensemble Stacking, F1-Score, Gaussian Naive Bayes, KNN, Logistic Regression, OASIS Dataset, PCA, Stratified Cross-Validation, SVM.

I. INTRODUCTION

DEMENTIA is a chronic or progressive condition that leads to deterioration in cognitive function beyond what might be expected from normal aging. According to the World Health organization, every year there are nearly 10 million new cases of dementia. It does not affect consciousness but it affects memory, thinking capability, learning capacity, decision making, language, and other cognitive functions which deteriorate emotional control and social behavior. Disability and dependency among older people are one of the major effects of dementia. There will be a physical, psychological, social, and economic impact on family and society. Apart from aging, a variety of diseases and brain injuries can primarily or secondarily affect the brain and result in dementia. In that way, 60-70% of dementia cases were the result of Alzheimer's disease, making it the most common form of dementia.[1]

In recent years, Machine Learning has evolved as a contemporary approach in predicting, classifying, making decisions, and identifying without any human interaction or involvement. ML algorithms are playing

Dr.Sowmya.V,
Assistant Professor (Sr. Grade),
Center for Computational Engineering and Networking, Amrita School of Engineering,
Coimbatore, Amrita Vishwa Vidyapeetham, India, e-mail: sowmiamrita@gmail.com

a key role in the medical world ranging from diagnosis to visualization of diseases and the study of disease transmission by analyzing the vital parameters or from medical images. In this paper, different ML algorithms' performance has been evaluated based on how well the algorithm is classifying a set of features given to the model and classifying them as demented or non-demented.

To improve sensitivity for disease detection in medical studies, longitudinal studies are widely applied to observe within-subject changes over a particular period in groups of subjects [6]. Here in this work longitudinal collection of data where for each subject multiple MRI scans (3 to 4) were obtained to collect the various measures. Some of the measures that are closely related to dementia identification are CDR score and MMSE. To quantify the severity of symptoms of dementia Clinical Dementia Rating (CDR) scale is used [11]. It is a five-point scale in which CDR-0 implies no cognitive impairment, and then the remaining four points connote various stages of dementia which is shown in Table.1.

TABLE 1
CDR SCORE AND ITS CONDITIONS

CDR Score	Annotation
CDR-0.5	very mild dementia
CDR-1	mild
CDR-2	moderate
CDR-3	severe

To test cognitive function among the elderly the Mini-Mental State Exam (MMSE) is widely used and it includes tests of orientation, attention, memory, language, and visual-spatial skills [9]. MMSE scores of subjects whose education levels are shown in Table 2.

TABLE 2
MMSE GRADE AND EQUIVALENT SCORES

Grade in MMSE	MMSE Score
7th grade or lower	22 or below
8th grade or some high school (but not a graduate of)	24 or below
high school graduate	25 or below
some college or higher	26 or below

The dataset used in this work is the OASIS (Open Access Series of Imaging Studies) dataset available at Mendeley data [2]. The dataset is a longitudinal collection of 150 subjects. The subjects include both men and women and all were right-handed and are of age group 60-96 years. Each subject was scanned on two or more visits with a time gap of at least 1 year contributing to a total of 373 imaging sessions. For each session and for an individual 3 or 4 T1-weighted MRI scans were taken. Out of 150 subjects, 72 of them were non-demented throughout the study, 64 were categorized as demented at their initial visits, and in which 51 individuals were in the state of mild to moderate Alzheimer's. The remaining 14 were characterized as non-demented at initial visits and subsequently identified with dementia at later visits. The dataset includes 14 features and is listed below:

1) ID - Identification
2) Group - Demented or Nondemented
3) Visit - The visit number

4) M/F – Gender
5) Hand - Dominant Hand
6) Age - Age in years
7) Educ - Years of Education
8) SES - Socioeconomic Status
9) MMSE - Mini-Mental State Examination
10) CDR - Clinical Dementia Rating
11) Etiv - Estimated Total Intracranial Volume
12) nWBV - Normalize Whole Brain Volume
13) ASF - Atlas Scaling Factor
14) Delay - Delay

II. LITERATURE REVIEW

The OASIS dataset was available in Mendeley data and is cited in "Data for Machine Learning In Medicine: Classification And Prediction Of Dementia By Support Vector Machines (SVM)"[3]. Choi et al. used the OASIS pre-processed dataset and applied Logistic Regression, Support Vector Machine, Decision tree Classifier, Random Forest Classifier, and Adaboost algorithms to predict AD [5]. Several studies have been done on this dataset and in [3] the authors have implemented an SVM algorithm for dementia prediction and it performed best at low gamma (1.0E-4) and high regularized (C = 100) values. This approach achieved accuracy and precision of 68.75% and 64.18%. Here the missing entries were filled-up by averaging particular attribute values. The features were selected based on the correlation values. Attributes like Subject ID, CDR, MMSE, Age, MR Delay, and n WBV were chosen as input to SVM.

Another work proposed by P. Kishore et al. on detection and analysis of Alzheimer's disease using various machine learning algorithms [4]. For preprocessing data, that is to replace missing values, imputation had been performed. They evaluated based on accuracy obtained after 5-fold cross-validation on algorithms such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), and three different Support Vector Machines (SVMs). It was observed that SVM using Linear Kernel with 'C' value 2 obtained the best accuracy of 0.95. Several works using deep learning approaches were also done on to this data and one such work is [7] by using Long Short-Term Memory (LSTM) for Alzheimer's Disease Dementia Classification by Sneha Mirulalini Gnanasegar et al. For feature selection they have used an R package called Boruta and using those features trained an LSTM model which resulted in an accuracy of 94%.

III. OBJECTIVE

The goal is to:

1) Pre-process the data and convert that into a form suitable for model training. It includes:
   - Exploratory data analysis
   - Imputation
   - Feature selection
   - Feature scaling
2) Apply different ML algorithms such as:
   - Logistic Regression,

- KNN,
- Gaussian Naïve Bayes,
- Decision Tree,
- SVM,
- Ensemble algorithms:
    - Stacking (Logistic Regression, KNN, Gaussian Na¨ıve Bayes, Decision Tree, SVM)
    - Bagging (Random Forest)
    - Boosting (AdaBoost, XGBoost)

to the data and to analyze and evaluate each algorithm based on their accuracy and f1 score. The workflow is shown in figure 1.

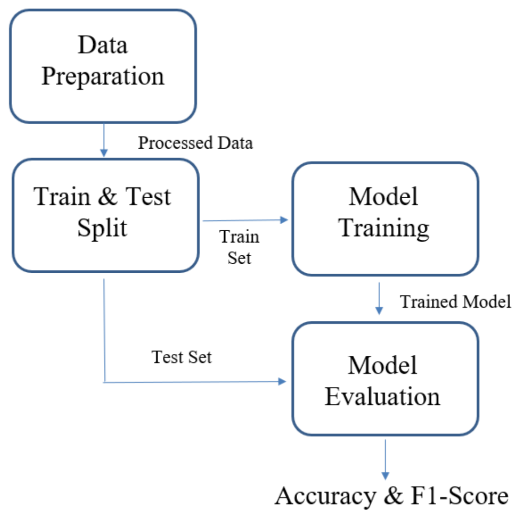

Figure 1. Workflow

IV. THEORETICAL BACKGROUND

A. Stratified K Fold Cross-Validation

In k fold cross-validation, the training dataset is split into k folds. The first k-1 folds are used for training, and the remaining fold is held for testing wherein each fold, the samples from classes are not equally distributed. In stratified k fold validation, for each fold's samples from each class are taken equally rather than random sampling as shown in figure 2.

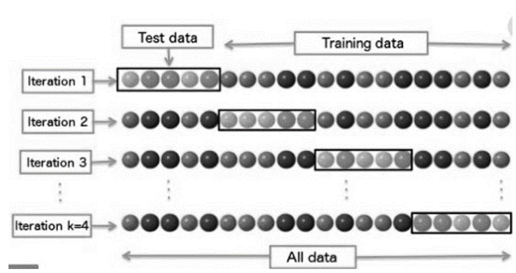

Figure 2. Stratified K-Fold

B. BORUTA

It's an R package used for feature selection and it falls under wrapper method subset selection. It ranks the feature based upon its importance which means it scores all features based upon the predictive power and returns only the minimal optimal features. It is built upon random forest and the base idea is to compare the relevance of the real feature with the shadowversion of the same (i.e., randomized). For each feature randomized shadow feature is created and the Z-score of the original variable is compared with the maximum Z-score of shadow variables. Real features that have low scores compared to the shadow features best score are considered unimportant.[11]

C. PCA (Principal Component Analysis)

Principal Component Analysis (PCA) is a statistical procedure used for dimensionality reduction that uses an orthogonal transformation that converts a group of correlated variables to a group of uncorrelated variables which are called principal components in eigenspace. The direction of maximum variance within the input space happens to be equivalent because of the principal eigenvector of the covariance matrix. Let 'X' be Nxn matrix with columns as

$$x_1 - \bar{x}, x_2 - \bar{x}, \ldots x_n - \bar{x}$$

and

$$Q = X^T X$$

The PCA theorem is given by: Each x_j can be written as shown below, where e_i are the 'n' eigenvectors of Q with non-zeroeigenvalues.

$$x_j = \bar{x} - \sum_{i=1}^{n} g_{ji} e_i$$

D. Logistic Regression

Logistic regression is a supervised classification technique used when the target is categorical. It helps to predict the probability of an outcome. It is similar to linear regression but instead of probability natural logarithm of the "odds" of the target is used to construct the curve. The sigmoid function is employed to map the anticipated values to the possibilities. The mapped values are in the range of 0 to 1. The sigmoid function is given by

$$S(z) = \frac{1}{1+e^{-z}}$$

Where:

S(z) = (probability estimate) output between 0 and 1

z = input to the function (b0 + b1*x)

The maximum likelihood is used to find the values of b0 and b1 such that the resultant probabilities are close to either 1or 0 and to calculate these coefficients, the following equation is updated,

b = b + alpha (y-prediction) * prediction *(1-prediction) * x

E. K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a supervised algorithm used mainly for classification. It is used when there are two classes (foreg. Class A and Class B) and when a new point arrives to find to which class it falls. falls. Steps followed to implement KNN are:

Step-1: Select the 'K' value which denotes the number of the neighbors. K value should not be:

- a very low value such as K=1 or K=2, which can be noisy and it will have the effect of outliers.
- a very large value, because it is likely that the decision may be skewed.

Step-2: Calculate the distance between the test point with all train points. The distance measure can be Euclidean/Minkowski/Chebyshev/Manhattan but the commonly used distance measure is Euclidean which is the default. For 'd' dimensional space, the Euclidean distance formula is:

$$\sum_{i=1}^{d}(x_i - y_i)^2$$

Step-3: Sort the distance in ascending order.

Step-4: Take the first K values.

Step-5: Among the k neighbors, count the number of data points in each class.

Step-6: Assign the new data point to that category for which the number of neighbors is maximum.

F. Naive Bayes

The Naive Bayes algorithm is a supervised classification algorithm, which is based on the Bayes theorem. It is a probabilistic classifier, which means it predicts based on the probability of an object. It assumes that the occurrence of 1 feature is independent of the occurrence of other features. The Bayes classifier:

$$arg_y max\ P(Y|X_1, X_2, ..., X_n)$$

$$P(Y|X_1, X_2, ..., X_n) = \frac{P(X_1, X_2, ..., X_n|Y)P(Y)}{P(X_1, X_2, ..., X_n)}$$

where:

$P(Y|X_1, X_2, \ldots, X_n)$ – Posterior probability
$P(X_1, X_2,, X_n|Y)$ – Likelihood probability
P (Y) – Prior Probability
$P(X_1, X_2, \ldots, X_n)$ is Marginal Probability or Normalization constant

G. Decision Tree

The decision tree is a supervised algorithm for classification and regression. It starts with a single node called the root node, which branches into possible outcomes. To decide the root node and which node has to be split is decided based on some measures. A few of those measures are Entropy, Gini index, and Information gain. Entropy is the measure of randomness and entropy at a given node 't' is given by:

$$Entropy(t) = -\sum_{j} p(j|t)\ log_2 p(j|t)$$

where p(j — t) is the relative frequency of class j at node t. The low entropy feature is made as to the root node and for splitting, a feature with high entropy is split. After the split, the information gain should be high and entropy after the split should be low compared to before the split. Information gain is given by

Information gain = entropy before the split – entropy after the split

$$Gain_{split} = Entropy(p) - \sum_{i=1}^{k} \frac{n_i}{n}\ Entropy(i)$$

where parent node p, is split into k partitions, and n_i is the number of records in the partition i. Gini index is the measure of homogeneity. So, if the Gini value is low then the node is pure. If the Gini value is high, the node has high randomness and it is impure. Gini Index for a given node 't' is given by :

$$GINI(t) = 1 - \sum_j [p(j|t)]^2$$

$$GINI_{split} = \sum_{i=1}^{k} \frac{n_i}{n} Gini(i)$$

H. SVM

Support Vector Machine is a supervised learning algorithm used both for regression and classification but primarily usedto perform classification. The goal of SVM is to create a hyperplane/decision boundary that segregates n-dimensional spaceinto classes so that when new data arrives it can be correctly categorized in the future. The points that help in creating the boundary are called support vectors. Only at support vectors constraints are active and the Lagrangian multiplier is non-zero. Figure 3 shows two different classes separated by a hyperplane. The SVM function tries to find the maximum margin and is given by

$$Maximize \frac{2}{\sqrt{w_1^2 + w_2^2}} \equiv Minimize \frac{\sqrt{w_1^2 + w_2^2}}{2}$$

$$(i.e)\ Minimize_{w,g}\ \frac{1}{2} w^T w, \quad with\ s.t\ constraints$$

For 'n' data points with 'm' features, constraints can be

$$w^T x_i - g <= -1\ for\ class\ -1$$

$$w^T x_i - g >= 1\ for\ class\ +1$$

where w is the weight vector, g is the intercept and x is the i^{th} data point where 'i' ranges from 1 to n. The number of weights corresponds to the number of features and the number of constraints is based on the number of data points. SVM formulation with class included,

$$Minimize_{w,g}\ \frac{1}{2} w^T w, \quad s.t\ d_i(w^T x_i - g) = 1, \quad where\ i = 1\ to\ n$$

where'd' is the class label of the i^{th} data point in the dataset containing 'n' data points and 'm' features.

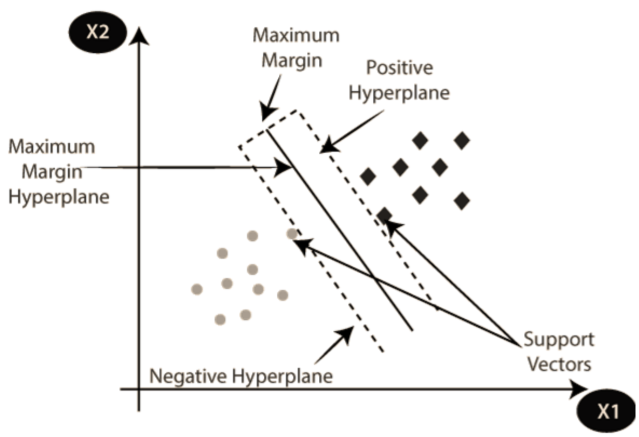

Figure 3. SVM Classifier

Types of SVM:

- Linear SVM: Used for linearly separable data where a single straight line can separate two classes.
- Non-linear SVM: Used for non-linearly separable data, for such data kernel should be used to project data to higher dimension for that kernel are used (to map from input space to feature space) and then perform linear operations.

I. Ensemble Learning Techniques

Ensemble models are machine learning approaches that can be used to optimize the model by achieving a bias-variance trade-off. It combines multiple models called base estimators and the predictions from those models are aggregated by the meta-model /meta-algorithm. The three ensemble techniques are:

1) Stacking
2) Bagging
3) Boosting

1) ***Stacking:*** This is an ensemble approach that involves stacking heterogeneous models to achieve robust predictions. It has two levels: the first level is the base models where different ml algorithms are stacked and level 2 is the meta learner model where the predictions from base models are aggregated to give the final predictions. The three main aggregating methods that can be used to combine base model predictions are:

- Max Voting: The final prediction is based on a majority vote for classification problems.
- Averaging: In this technique, for regression problems, the predictions are averaged.
- Weighted Average: This technique is used for regression problems where while averaging the predictions some weights are given to each base model prediction. The model's prediction is given high weight if it has good evaluation metrics.

2) ***Bagging:*** In this approach, the training data is made available to an iterative process of learning which reduces variance and avoids overfitting. This approach makes use of homogeneous models. One example is Random Forest, where it is a bag of decision trees. Shallow tress-like Decision trees have low bias and high variance, this high variance can be reduced by bagging many decision trees called random forests. Thus, Random Forest aims to reduce variance. Figure 4 shows the outline of the bagging approach.

- Bootstrapping: Bagging is based on sampling technique. It creates multiple subsets of original training data with replacement of equal size and the subsets are given to train models parallelly.
- Random Forest: It builds multiple split trees by using the subset of samples as well as the subset of features and is implemented randomly. Each training set is fitted by multiple trees.

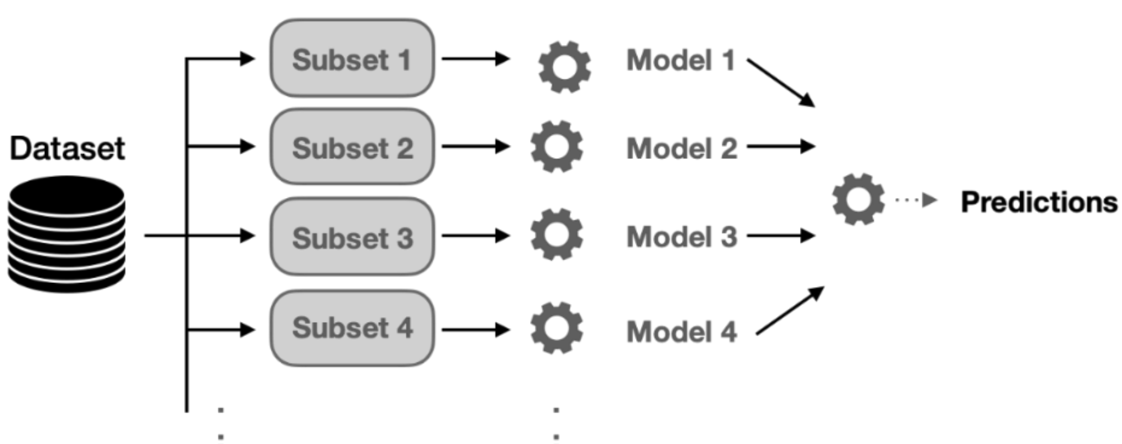

Figure 4. Bagging

*3) **Boosting:*** It is an ensemble of homogeneous algorithms/models where the model is built on top of weak learners. It isa sequential process where each tree is adjusting its weights based on prior knowledge of accuracies. Training and measuring error in estimates are done for a given number of iterations or until the error rate is not increasing significantly. Figure 5 shows the boosting approach. Some of the boosting techniques are:

1) AdaBoost(Adaptive Boosting)
2) XGBoost(Extreme Gradient Boosting)

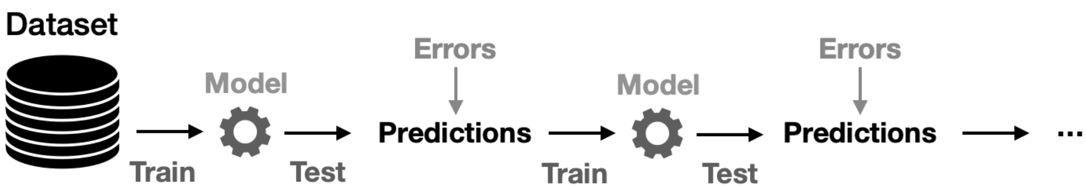

Figure 5. Boosting Approach

V. PROPOSED METHODOLOGY

The longitudinal dataset contains 373 data points with 15 features of 150 subjects as mentioned in section II. The proposed workflow as shown in figure 1 is applied to all the ml algorithms mentioned in the previous section. This section includes all the preprocessing and data exploration done on the dataset before being given to various ml models and also about parameters passed to the model to achieve the best results.

A. Exploratory Data Analysis

From the dataset exploration, it's been observed from figure 6 that there was not much class imbalance in the dataset where 51% data points belong to class non-demented and 49% data points to class demented. From the bar graph in figure 7, it can be noticed that there is a higher rate of dementia in males than in females.

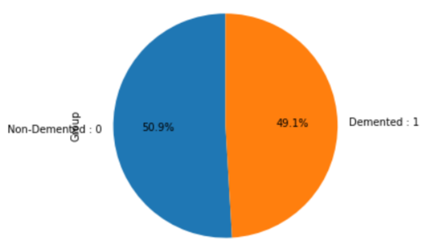

Figure 6. Class Distribution

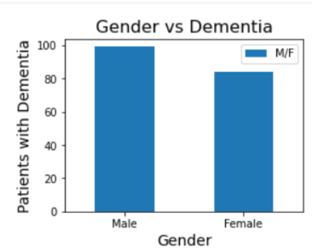

Figure 7. Gender vs Dementia

Normalized Whole Brain Volume (nWBV) has a negative correlation with age in general, however, from the scatter plot itcan be observed that this correlation seems more pronounced in dementia patients.

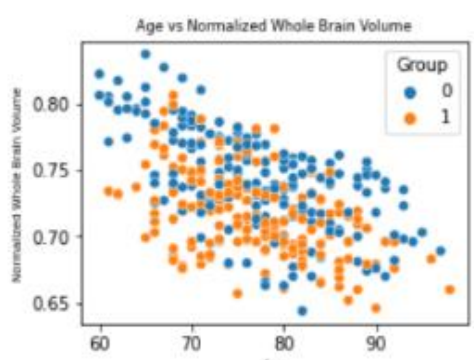

Figure 8. Relation of Age with Normalized Whole Brain Volume

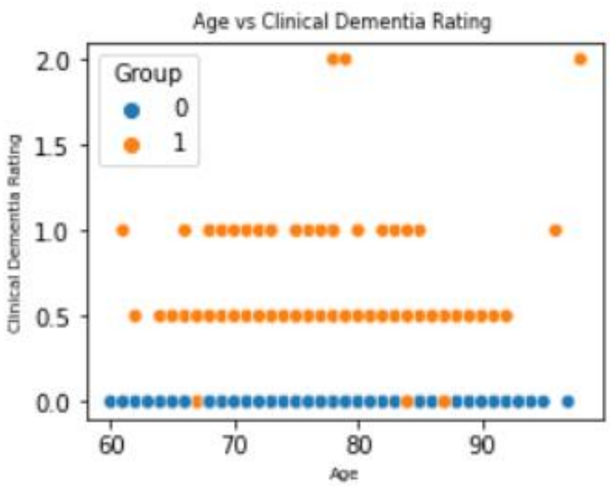

Figure 9. CDR Score Distribution

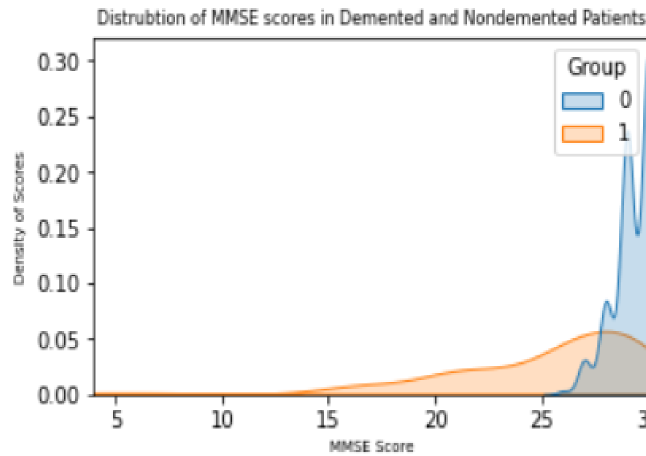

Figure 10. Distribution of MMSE Score

Regardless of age, Clinical Dementia Rating (CDR) shows clear distinctions between demented and nondemented patients where almost all dementia patients have a CDR Score >=0.5 and is displayed in the figure. Similarly, MMSE scores also show separation between the two groups where MMSE score for the demented group is spread out in the range 17 - 26 while non-demented MMSE scores are in the range 26-30. However, the CDR score seems to be more robust than the MMSE score.

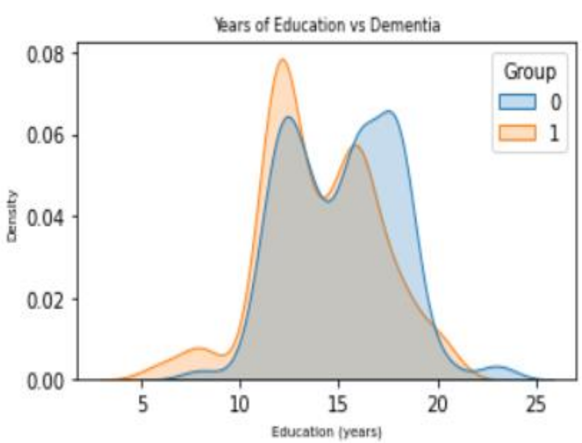

Figure 11. Educational Level Spread Over Data

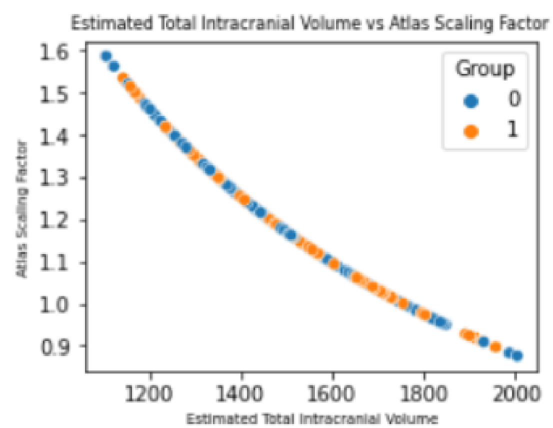

Figure 12. Etiv vs ASF

The relation between the years of education and the group shows that Demented patients had fewer years of education when compared to non-demented patients. The relationship between Atlas Scoring Factor (ASF) and estimated total intracranial volume (eTIV) was like 1-1. The reasonis ASF is the volume-scaling factor and it should be proportionate to TIV because atlas normalization depends on the headsize.

B. Pre-Processing

1) Typecasting & Imputation: The OASIS dataset was a mix of both numerical and categorical features, so the first stepwas to convert the gender and group columns to numerical values. In the gender column 'M' is replaced as 1 and 'F' as 0 andin the group column, non-demented is set as 0 and 'demented ' & ' converted' is set as 1 because converted also falls to the demented group. Here 'Group' column is the target column. As there are few missing values in the dataset, the missing valueswere imputed by the median value of the respective feature column.

2) Feature Selection: The important features that are highly linked in classifying a patient as demented or not are alone taken and given to model for training. To find those significant features, Boruta as mentioned in section IV is used to know the ranking of the features based on their relation to correctly classifying a data point. In addition to this PCA (Principal Component Analysis) to know how many components are required to achieve a variance of 0.95 is shown in figure 14. The selected features are listed in table 3 with BORUTA ranking :

TABLE 3
OPTIMAL PARAMETERS

SELECTED FEATURES	BORUTA RANKING
Group - 0 or 1	Target
Gender - 0 or 1	5
Age - Age in years	3
Educ - Years of Education	4
MMSE - Mini-Mental State Examination	1
CDR - Clinical Dementia Rating	1
eTIV - Estimated Total Intracranial Volume	1
nWBV - Normalize Whole Brain Volume	1
ASF - Atlas Scaling Factor	2

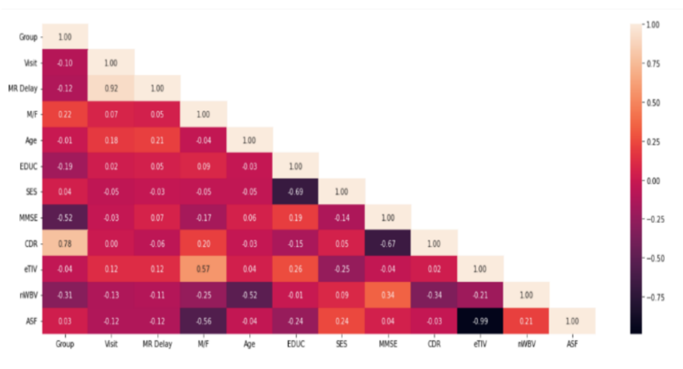

Figure 13. Correlation Heatmap

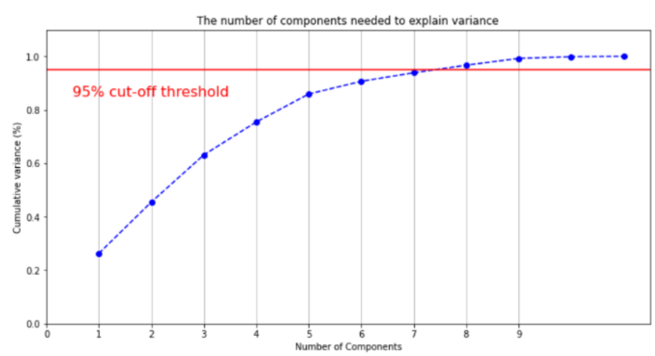

Figure 14. Covariance map of components

***3) Feature Scaling**:* Now the data points with the selected features are normalized to a standard form by applying z-score and it is given by the formula shown below, where μ is the mean and σ is the variance (standard deviation).

$$z = \frac{x - \mu}{\sigma}$$

C. Train-Test Split

The normalized dataset is then split into train and test set in the ratio 70:30. The split done here is stratified so the train and test set consists of the equal number of class 0 and class 1 data points. That is out of 373 samples with 8 features, the train set contains 261 samples (Class 0: 133, Class 1: 128) and the test set contains 112 samples (Class 0: 57, Class 1: 55).

D. Model Building & Training

Then the train and test sets are passed through the various models. For Logistic Regression, Decision Tree &Naive Bayes only random state as 42 is set. The other models:

KNN: Here the optimal value K is found and then that K value is used to train and test the model. The distance measure usedhere is Euclidean.

SVM: To find the optimal parameters like C, gamma, and kernel, grid search with scoring as accuracy was performed. The optimal parameters obtained after a 10-fold grid search were used to train and test the model.

Ensemble Approach:

Stacking: In this work, Logistic regression, naive Bayes, KNN (with the optimal K value), Decision tree, and SVM (with optimal parameters) were used as base models and the Metamodel here is Logistic regression.

Bagging - Random Forest: Grid Search with scoring as accuracy and cross-validation as 10 folds were performed to find theoptimal parameters like estimator, max depth, and criterion.

Boosting - AdaBoost & XGBoost: For both the boosting algorithms grid search with scoring as accuracy and cross-validation as10 folds were performed to find the optimal estimator.

VI. RESULTS & DISCUSSION

The grid search parameters for the algorithms are shown in table 4.

TABLE 4
OPTIMAL PARAMETERS

K Nearest Neighbor	
K -value	5
Support Vector Machine	
C	6
Gamma	0.5
Kernel	RBF (Radial Basis Function)
Random Forest	
Criterion	Gini
Max Depth	9
No. of Estimators	50
AdaBoost	
No. of Estimators	20
XGBoost	
No. of Estimators	20

From the confusion matrix of all the models as in figure 15, it can be depicted that Logistic Regression, Naïve Bayes and XGBoost models had less misclassification accounting for a total of 6 wrongly misclassed data points.

The 10 fold cross-validation results of the models are shown in table 5. From the results, it can be seen that moreover, all the models are performing well as the data given to the models are clean and processed. Comparatively, the XGBoost algorithm performed well on the data in classifying the patients as demented or non-demented with a test accuracy of 95% and f1-score of 0.95 on testing. Following XGBoost, the Logistic

Regression model had a testing accuracy of 0.95% and f1-score of 0.94 in testing, and also stacked heterogeneous models fetched a testing accuracy of 94% and 0.94 testing f1-score.

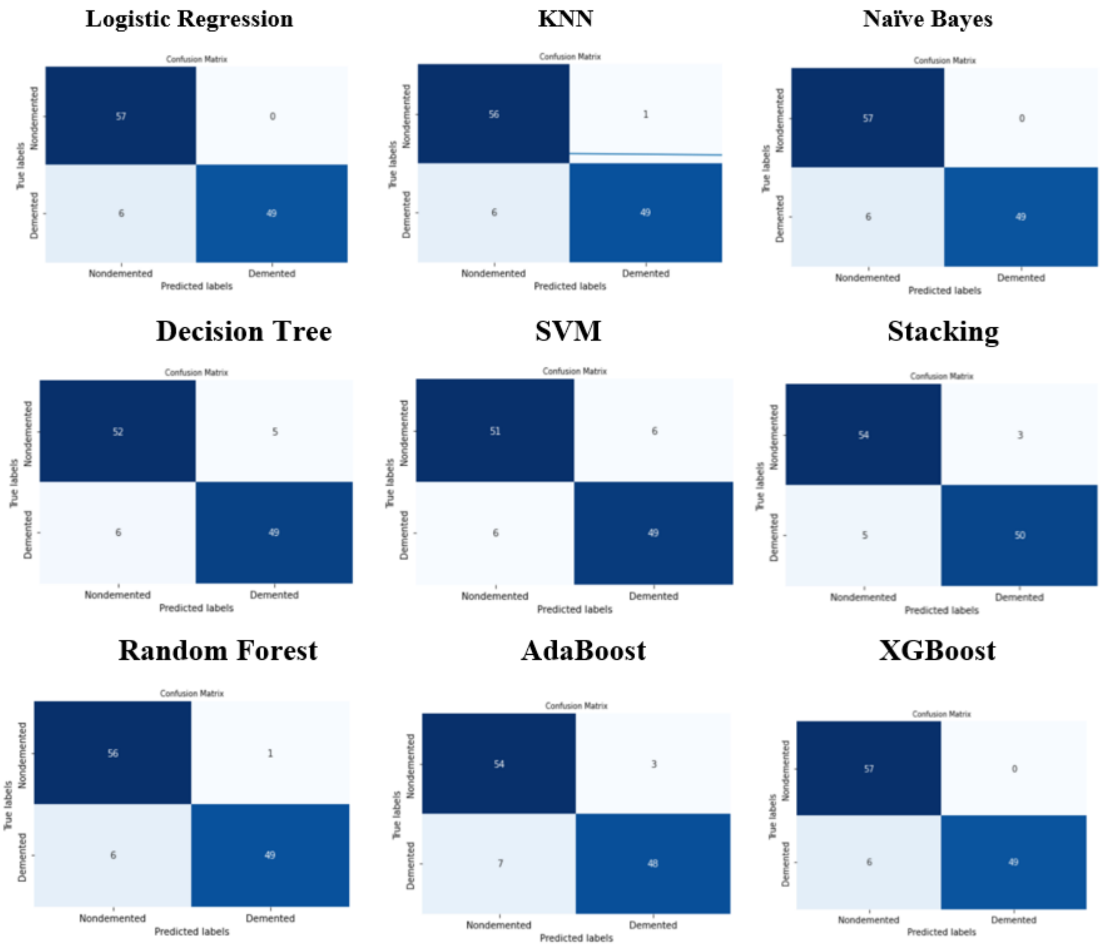

Figure 14. Confusion Matrix of All Models

TABLE 5
EVALUATION METRICS FOR ALL MODELS

MODEL	ACCURACY(Mean)		F1- SCORE(Mean)	
	TRAINING	TESTING	TRAINING	TESTING
Logistic Regression	0.95	0.95	0.94	0.94
KNN	1.0	0.88	1.0	0.88
Naïve Bayes	0.95	0.95	0.94	0.94
Decision Tree	1.0	0.88	1.0	0.89
SVM	0.97	0.89	0.97	0.88
Stacking	0.95	0.94	0.95	0.94
Random Forest	0.98	0.94	0.98	0.93
AdaBoost	0.96	0.93	0.96	0.93
XGBoost	0.96	0.95	0.95	0.95

VII. CONCLUSION

In this work, five individual Machine Learning algorithms and Ensemble Techniques were performed on the OASIS longitudinal Alzheimer's dementia classification dataset and their performance was evaluated based on accuracy and f1 score. Due to proper data preprocessing, all the models were able to achieve the best accuracy possible with the least misclassification. Selecting the optimal features made the model predict the respective class of the datapoint with a smaller number of features. Further, this work can be improved by doing a comparative study on ML and DL algorithms, and also this work can be done on MRI image data to classify the patient as demented or non-demented.

REFERENCES

[1] Brain Disease Information and statistics related to dementia (n.d.). https://www.who.int/news-room/fact-sheets/detail/dementia

[2] Dataset Available at https://data.mendeley.com/datasets/tsy6rbc5d4/1

[3] Gopi Battineni, Nalini Chintalapudi, Francesco Amenta (2019), Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM), https://doi.org/10.1016/j.imu.2019.100200

[4] P.Kishore, Ch. Usha Kumari, M.N.V.S.S.Kumar, T.Pavani (2020), Detection and analysis of Alzheimer's disease using various machine learning algorithms,https://doi.org/10.1016/j.matpr.2020.07.645

[5] Choi H, Kyuri, Saurin, Detecting Early Alzheimer's Using MRI Data and Machine Learning Internet. Kaggle. Available from: www.kaggle.com/hyunseokc/detecting-early-alzheimer-s.

[6] Leon M. Aksman, David J. Lythgoe, Steven C.R. Williams, Martha Jokisch, Christoph Mönninghoff, Johannes Streffer, Karl-Heinz Jöckel, Christian Weimar, Andre F. Marquand (2016) Making use of longitudinal information in pattern recognition, https://doi.org/10.1002/hbm.23317

[7] Gnanasegar SM, Bhasuran B, Natarajan J (2020) A Long Short-Term Memory Deep Learning Network for MRI Based Alzheimer's Disease Dementia Classification. J Appl Bioinforma Comput Biol 9:6. DOI: 10.37532/jabcb.2020.9(6).187

[8] Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C., Buckner, R. L. (2010). Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. Journal of cognitive neuroscience, 22(12), 2677–2684. https://doi.org/10.1162/jocn.2009.21407

[9] MMSE SCORE details are available at:https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?id=phd001525.1

[10] BORUTA concept explained for feature selection (n.d.). https://medium.com/analytics-vidhya/feature-importance-explained-bfc8d874bcf

[11] CDR SCORE details for dementia patients are available at website: https://knightadrc.wustl.edu/cdr/pdfs/cdroverviewtranscript − revised.pdf