International Journal of Intelligent Communication and Computer Science
Vol. 2, No. 2, 2024, Pages 1-16.

A comparative analysis of Machine Learning based algorithms for diagnosis of Alzheimer's Disease

Tamoghna Mukherjee[1], Anirban Mitra[1], Pronaya Bhattacharya[1], Debolina Nath[1] and Pragya Mukherjee[1]

[1]Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University, Kolkata-700135, West Bengal, India

Research Paper

Email: pbhattacharya@kol.amity.edu,
Received: 11 Jul 2024, revised: 1 Aug 2024 Accepted: 1 Aug 2024

Abstract:

Alzheimer's disease (AD) is one of the most prevalent types of dementia, a term used to describe a deterioration in memory along with various other mental and behavioural skills in people. The rate of the affected individuals is increasing at an alarming rate and is ultimately affecting a huge section of the population. The OASIS (Open Access Series of Imaging Studies) longitudinal dataset has been used in this study's comparative analysis of various classification algorithms for the diagnosis of AD. Since the accuracy of the obtained results is of utmost importance in the case of highly sensitive medical data, the aim is to achieve high prediction accuracy. Among the applied classification algorithms such as Random Forest (RF), Logistic Regression (LR), Gradient Boosting Classifier (GBC), Naive Bayes Classifier (NBC), and Support Vector Machine (SVM), it was found that NBC provides the most stable results with a high accuracy of 97.33% on our dataset. Hence, NBC was chosen for the creation of our prediction model and deployed to operate as the underlying algorithm for the UI of our web application. The developed UI accepts patient-oriented data inputs from the user and displays whether the patient is demented or non-demented.

Keywords: Alzheimer's disease, Logistic Regression, Random Forrest, Gradient Boosting, Naive Bayes, Support Vector Machine.

1. Introduction

Dementia [1] is the general term associated with the loss of memory, language and other thinking capabilities that severely affects daily functioning in human beings. As reported by World Health Organization (WHO), dementia is currently the seventh leading cause of death and one of the major causes of disability and dependency among the aged population. With 60–70% of cases, Alzheimer's disease (AD) [2] is the most prevalent type of dementia. At present, more than 55 million people across the globe are affected by AD and the number is on continuous rise. Statistics indicate that it will increase to 78 million by 2030 and 139 million by 2050 [3]. The disease also has a major economic impact on the nations, and it is accessed that its total estimated cost is over 1.3 trillion USD across the globe. Though there is no such curative method for this neurodegenerative disease, its early detection can slow down the progressive degeneration in the patient through supporting medication. The initiation of proper medication at an early stage can prevent the neurons from getting damaged to a greater extent within a shorter time frame.

The World Alzheimer Report [4], released every year, suggests the different developments with respect to the disease diagnosis and its prevention and post diagnosis treatment process. The early detection of AD using modern technology is one of its main focuses. Several studies are being conducted in this domain and efforts are being made to predict as well as identify symptoms of AD with accuracy. The detection of AD requires thorough medical analysis under the supervision of an expert professional. This makes it a costly as well as time-consuming affair. The diagnosis of the disease is mostly based on the analysis of MRI scans conducted over a period on the patients. As the cost of the medical assessments is continuously on rise, often it is beyond the capacity of the masses to afford it. Moreover, the demand for such medical assessments is so high that it leads to long waiting periods before undergoing the assessment and obtaining the test results. This leads to unwanted delays in the diagnosis and initiation of the treatment. Thus, automating the detection process can reduce both the time and the cost of treatment. Artificial Intelligence and ML have emerged as one of the most used approaches for the detection of AD[5]. In this study, we aim to apply different supervised learning algorithms to classify patients as demented or non-demented and use the best-performing algorithm to develop a prediction model for AD. For this purpose, the OASIS longitudinal dataset was chosen since it is a labeled dataset suitable for the application of classification algorithms. The following were applied after data preprocessing as discussed in detail under the proposed mechanism: LR [6], RF [7], GBC [8], NBC [9], and SVM [10]. Each of them performed well on the pre-processed data, however, NBC showed the best and most stable results with an accuracy of 97.33%. Thus, it was chosen to develop the prediction model. The working model seemed to classify the patients with great accuracy. A UI interface was also created by deploying the NBC-based prediction model at the backend to make it easier to predict the results by giving the patient-specific data as input.

Motivation, contribution and limitations

The motivation behind this study is rooted in the significant impact that dementia, particularly Alzheimer's disease (AD), has on the global population. Dementia is a major cause of disability and dependency among the elderly, with AD being the most prevalent form. The current methods of diagnosing AD are costly, time-consuming, and often lead to delays in treatment initiation due to the high demand and associated expenses of medical assessments like MRI scans. This study aims to address these challenges by leveraging Artificial Intelligence (AI) and Machine Learning (ML) to develop an automated and cost-effective method for early detection of AD. Early detection can significantly slow the progression of the disease through timely intervention, thus improving patient outcomes and reducing the economic burden on healthcare systems.
The methods for analysis was chosen meticulously based on certain traits matching with the desired outcome. Logistic Regression is a straightforward and interpretable algorithm that was applied to classify patients as demented or non-demented. Its use helps in understanding the influence of individual features on the probability of a patient being demented, providing valuable insights into the relationships between the input variables and the outcome. In terms of Random Forest, an ensemble learning method, was used to improve classification performance by combining multiple decision trees. It helps in handling a large number of input variables and managing missing data effectively. A equally competitive method Gradient Boosting Classifier was applied to enhance predictive accuracy by sequentially building models to correct the errors of previous models. This method is effective in capturing complex patterns in the data. To judge more we used Naive Bayes Classifier, based on Bayes' theorem. It was found to perform best with an accuracy of 97.33%. It was selected for the final prediction model due to its simplicity, efficiency, and surprisingly robust performance despite its strong independence assumptions. Lastly Support Vector Machine was used to classify patients by finding the optimal hyperplane that separates the data into different classes. SVM is effective in high-dimensional spaces and with datasets where the number of dimensions exceeds the number of samples.
The accuracy and performance of the model are highly dependent on the quality and comprehensiveness of the OASIS longitudinal dataset. Any biases or limitations in the dataset can directly affect the model's reliability and generalizability. While NBC performed best in this study, its effectiveness is contingent on the assumption that features are independent. In real-world scenarios, this assumption may not always hold true, potentially affecting the model's accuracy in different contexts. The model's ability to generalize to diverse populations or datasets outside of the OASIS dataset is not established. The effectiveness of the model across different demographic and clinical settings needs further validation. Developing a prediction model and user interface is one aspect, but integrating this tool into existing healthcare systems and ensuring its adoption by healthcare

providers presents additional challenges. Issues related to interoperability, user training, and system acceptance need to be addressed. Using patient data for training and deploying AI models raises ethical and privacy concerns. Ensuring data security, patient consent, and compliance with regulations such as HIPAA is crucial.

Each method contributed uniquely to the study, offering insights into their strengths and limitations. The Naive Bayes Classifier's outstanding performance led to its selection for the final prediction model, which was integrated into a user-friendly interface for practical use in early detection of AD. Overall, while the study makes significant contributions towards automating the early detection of AD using AI and ML, further work is needed to address these limitations and ensure the model's broader applicability and integration into healthcare practice.

The remainder of the paper is organized as follows: Section II discusses the previous studies conducted in this domain, Section III focuses on the Proposed Mechanism, Section IV contains the Experimentation Results, and Section V which suggests the Conclusion and Future Works, followed by the set of References used in this paper.

2. Literature Review

Several studies have been conducted over the years using different datasets as well as different approaches for developing an efficient clinical system to predict Alzheimer's disease. Numerous strategies have been applied to detect and predict the disease, particularly in the fields of ML and deep learning. Some of the pre-existing works conducted in this field have been discussed below.

Both supervised and unsupervised learning algorithms have been extensively employed in ML for disease diagnosis as well as prediction. In [11], generalized linear models (GLM), decision trees (DT), Rule induction, NBC, k-nearest neighbors (k-NN), and deep learning techniques were all applied to the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, collected through the TADPOLE (The Alzheimer's Disease Prediction Of Longitudinal Evolution) challenge [12]. The objective was to categorize the five distinct phases of AD and to determine the most distinctive characteristic for each stage of AD within the ADNI dataset. The obtained results showed that GLM outperformed the rest with the highest accuracy of 88.24%. In another study [13], ADNI was used as a part of the early detection analysis of AD where a sophisticated hybrid cognitive classification mechanism was applied using 2-layer model stacking. The procedure outperformed other popular classification techniques. The algorithm showed 3 experiments with the ultimate highest accuracy of 95.12% using hybrid modeling. In [14], a neural network architecture was proposed along with a novel preprocessing algorithm for the prediction of AD. Data from the ADNI was preprocessed using a novel technique named as "All-Pairs" technique to produce the training dataset. In the "All-Pairs" technique the comparison of all possible pairs of temporal data points for each patient was considered. The trained model was capable of correlating clinical data obtained from patients at a particular instance of time with the progression of AD in the future. The model was found to be effective at predicting AD with an average mAUC score of 0.866. An in-depth examination of various deep learning approaches under the generative and discriminative architecture of deep learning has been performed in [15]. Data from OASIS has been used in [16] for the prediction of AD by SVM, LR, DT, and RF. After running the developed model before as well as after fine-tuning, it was found that SVM yielded the best results with an accuracy of 91.8% among the applied algorithms. A similar approach was used for detection in [17] by employing LR, SVM, RF, Extra Trees, and GBC. Extensive research work has been done using the data obtained from OASIS concentrating more on feature selection and feature extraction using ML algorithms in [18].

In [19], SVM, NBC, XGBoost, DT, LR, RF, Bagging and AdaBoost were applied on The Korean Brain Aging Study for the Early diagnosis and prediction of Alzheimer's disease dataset to classify patients under examination into the following three categories: individuals with cognitively normal control, individuals with mild cognitive impairment and individuals suffering from AD. XGBoost performed best on this dataset showing an accuracy of 82.09%.

Since RF has been proven to be effective at reducing high-dimensional and multi-source data, it has been widely employed for the prediction of AD. For instance, [20] examines the most recent RF implementations on single and multimodal neuroimaging data for the diagnosis and prognosis of AD. The use of Next Generation Sequencing (NGS) techniques has been applied in [21] to develop high-throughput screening methods for identifying the biomarkers and variants that assist in the early diagnosis of a disease. It suggests a model called VEPAD that relies on the use of cross-validation-based recursive feature elimination, which is then followed by a forward feature selection to choose key features to distinguish between deleterious and neutral variants.

Table 1: Summarized literature survey

Study	Dataset	Methods	Key Findings
Shahbaz et al. [11]	ADNI (TADPOLE)	GLM, DT, NBC, k-NN, Deep Learning	GLM outperformed other methods in classifying AD stages
Khan et al. [13]	ADNI	Hybrid Cognitive Classification (2-layer model stacking)	Hybrid model stacking outperformed popular classification techniques
Soliman et al. [14]	ADNI	Neural Network with "All-Pairs" technique for preprocessing	Novel preprocessing technique improved prediction accuracy
Shastry et al. [15]	Not specified	Various Deep Learning Approaches	In-depth examination of generative and discriminative DL architectures
Antor et al. [16]	OASIS	SVM, LR, DT, RF	SVM performed best after fine-tuning
Varun et al. [17]	OASIS	LR, SVM, RF, Extra Trees, GBC	Emphasis on feature selection and extraction
Hosseinzadeh Kasani et al. [19]	Korean Brain Aging Study	SVM, NBC, XGBoost, DT, LR, RF, Bagging, AdaBoost	XGBoost performed best among applied algorithms
Sarica et al. [20]	Various	Random Forest	Systematic review of RF implementations for AD diagnosis
Rangaswamy et al. [21]	NGS data	VEPAD Model, cross-validation-based recursive feature elimination	High-throughput screening for biomarkers and variants
Kishore et al. [22]	Not specified	SVM	SVM with linear kernel outperforms other ML algorithms
Khan et al. [23]	MRI data	Transfer Learning (VGG architecture)	Layer-by-layer fine-tuning with real MRI data
Liu et al. [24]	MRI images	Multiple kernels combining edge and node features	Effective categorization of AD
Huang et al. [25]	Not specified	Longitudinal monitoring, hierarchical classification	Addressed high feature dimensionality and added spatial information for better accuracy

The study conducted in [22] shows that the SVM algorithm with linear kernel outperforms other ML algorithms in detecting AD, although initial research studies proved lesser accuracy with the SVM algorithm in the detection of the said disease. In [23], transfer learning using a VGG architecture has been developed where the network has been fine-tuned layer-by-layer while also feeding a predefined group of such layers with real MRI data. The use of multiple kernels to combine edge characteristics and node features for the categorization of AD, as well as the evaluation of 710 MRI images using 10-fold cross-validation, have been proven to be effective [24]. Longitudinal monitoring of MCI brain imaging and a hierarchical classification strategy for predicting AD have been applied in [25] to deal with high feature dimensionality difficulties and adding spatial information for increasing prediction accuracy. Table1 describes the summarized literature survey.

These are a few to name and several other notable works have been conducted in this field. Hence, gaining insights from the above-discussed studies, our aim in this study is to apply classification techniques on our

chosen dataset post-preprocessing and develop a predictive model that can predict AD with high accuracy. A supporting UI will also be developed to make the developed model easy to use by all.

3. Proposed Mechanism

To develop a model for predicting and analyzing Alzheimer's Disease (AD), we followed a systematic workflow as illustrated in Fig. 1. This workflow outlines the step-by-step process used to build and evaluate each machine learning (ML) algorithm applied in our study. The process begins with data collection, where we gather and prepare datasets that include relevant clinical and imaging features associated with AD. This step ensures that the data used for training and testing the models are comprehensive and representative of various aspects of the disease. Next, we perform data preprocessing, which involves cleaning the data, handling missing values, and normalizing or standardizing features to improve model performance. Feature selection is then conducted to identify and retain the most informative variables, reducing dimensionality and enhancing the relevance of the data for the model. Following preprocessing, we split the data into training and testing sets. The training set is used to build and tune the ML models, while the testing set is reserved for evaluating their performance. This split helps ensure that the models are both trained on a representative subset and validated on unseen data. We then apply a range of machine learning algorithms, including supervised and unsupervised learning techniques, as specified in our study. Each algorithm is trained using the training data, and hyperparameters are optimized to achieve the best possible performance.

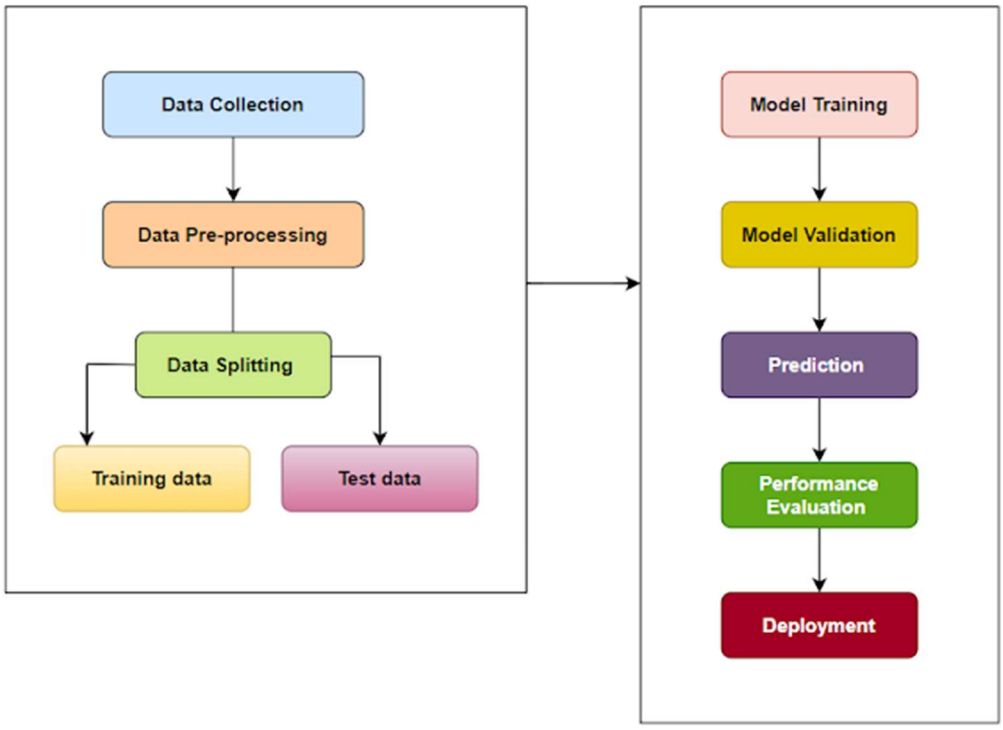

Figure 1: Overall workflow of the proposed mechanism

3.1 Dataset Description

In this study, the data used has been obtained from the OASIS longitudinal dataset available in Kaggle [37]. This dataset which was released in 2010 contains details of MRI scans of 150 participants, elderly males as well as females, in the age group of 60 to 96 collected over a course of 373 imaging sessions. Thus, OASIS-LD is a labelled dataset on which supervised learning algorithms are applicable for prediction purposes, Table 2.

Table 2: OASIS-LD Attributes

Sl.No.	Attributes	Description
1	Subject ID	Subject Identification
2	MRI ID	MRI Exam Identification
3	Group	Class or type of patients
4	Visit	Visit Order
5	MR Delay	MR delay time
6	M/F	Gender
7	Hand	Hand
8	Age	Age of patients
9	EDUC	Years of Education
10	SES	Socioeconomic Status
11	MMSE	Mini Mental State Examination
12	CDR	Clinical Dementia Rating
13	eTIV	Estimated total intracranial volume
14	nWBV	Normalize Whole Brain Volume
15	ASF	Atlas Scaling Factor

3.2 Data Preprocessing

Data preprocessing is an important step before evaluating any ML algorithm. Several steps have been taken into consideration to remove data inconsistency and redundancy from the raw dataset, such as data scaling, removal of features having lesser to least relation to the target variable, handling missing values, conversion of textual data to numeric data, etc. The chosen dataset required preprocessing without which the algorithms could not be applied it satisfactorily. Hence the following preprocessing steps were carried out to make the dataset ready for further processing:

1. The columns namely -**Subject ID, MRIID, Hand,** and **Visit** were dropped from the dataset as they were not significant parameters for model creation.
2. The null values present in the **SES** and **MMSE** fields were replaced by the mean values of the respective columns.
3. The column labelled **M/F** was renamed as **Gender**.
4. Since it is easier to deal with numeric values compared to characters, the values **M** and **F** present in the **Gender** column, signifying male and female respectively, were converted to 0 and 1.
5. Similarly, the values **Non-Demented** and **Demented** were replaced by 0 and 1 respectively.

Some ML algorithms show better performance when the input variables with numeric values are scaled to a standard range. Standardization scales the individual input variable by subtracting the mean from the data point and dividing by the standard deviation. This ensures the mean of the data values to be 0 and the standard deviation to be 1. As a part of our analysis, Standard scalar has been used to scale the input data. It was found that the algorithms considered, especially SVC significantly improved the performances and showed higher accuracy once the dataset was scaled using the standardization technique. The dataset was divided into a training dataset (containing 80% of the data) and a testing dataset (containing 20% of the data) following pre-processing and scaling. The testing data was used to evaluate the model's accuracy after it had been trained using ML techniques utilizing the training data.

3.3 Classification Algorithms

As a part of the comparative study of several classification techniques of ML algorithms for the prediction of AD in an individual, this paper focuses on the following algorithms:

3.3.1: Logistic Regression (LR):

Logistic regression is a fundamental and widely-used statistical method for binary classification tasks, where the goal is to predict the probability of an outcome belonging to one of two distinct categories. The relationship between the independent features and the dependent variable, which has two alternative outcomes (e.g.: 0/1), is shown via logistic regression [26] [27]. The general function for logistic regression, used for the predictive analysis of the dependent binary variable (in our case it is *demented/non-demented*) is sigmoid function denoted by-

$$f(x) = 1/1+e^{-x} \quad (1)$$

The function $f(x)$ in Eq. 1 converts a real value to a probability between 0 and 1. The class probability is taken to be 0 if it lies below the threshold value (0.5), else 1 in Fig. 2.

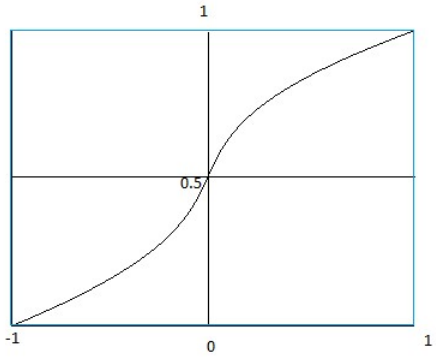

Figure 2: Logistic function for deciding class label

Key Points:

1. **Binary Classification:** Primarily used for binary outcomes (0/1, true/false, yes/no), but can be extended to multiclass classification.
2. **Linear Model:** Assumes a linear relationship between the input features and the log-odds of the output.
3. **Sigmoid Function:** Uses the logistic sigmoid function to convert the linear combination of inputs into a probability.
4. **Interpretability:** The coefficients (weights) provide insight into the influence of each feature on the prediction.
5. **Regularization:** Can incorporate L1 (Lasso) or L2 (Ridge) regularization to prevent overfitting.

3.3.2: Random Forest Classifier (RF):

Random Forest is a versatile and robust ensemble learning technique primarily used for classification and regression tasks. It builds upon the concept of decision trees by creating a "forest" of multiple decision trees, which collectively contribute to the final prediction. The process begins with generating a multitude of decision trees using a technique called bootstrap aggregating, or "bagging." In this method, different subsets of the

training data are sampled with replacement to train each tree, which introduces diversity among the trees and helps reduce overfitting.
In nutshell RF classifier [28] [29] takes the average of the outcomes from every decision tree and the most voted prediction result is chosen. This provides accurate and more precise prediction results.
The algorithm goes like –

1. Start by selecting random instances from the training data.
2. Build decision trees with the selected data.
3. Choose the number of decision trees.
4. Repeat steps 1 and 2
5. For testing data points, assign the category with majority votes from the created sub trees.

The visual flow of the algorithm has been demonstrated below in Fig. 3.

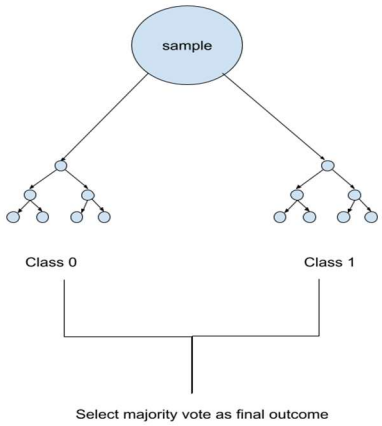

Figure 3: Majority voting in RF

The first decision tree for the RF model trained on the considered dataset has been visualized in Fig.4.

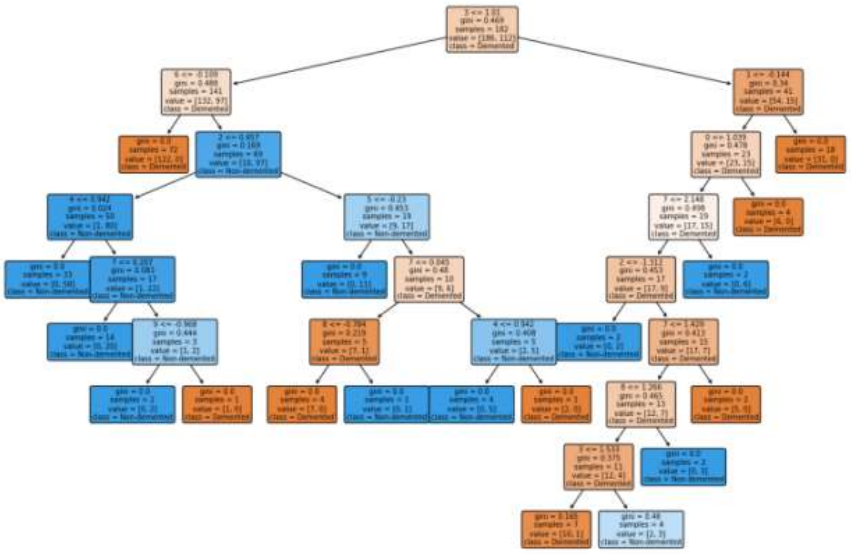

Figure 4: DT structure in classifying demented and non-demented individual.

Key Points:

1. **Ensemble Method**: Combines multiple decision trees to improve predictive performance and control overfitting.
2. **Bootstrap Aggregation (Bagging)**: Each tree is trained on a random subset of the data (with replacement) to create diversity among the trees.
3. **Random Feature Selection**: At each split in the tree, a random subset of features is considered for splitting, which helps reduce correlation between trees.
4. **Non-linear**: Capable of capturing complex interactions between features.
5. **Robustness**: Generally robust to overfitting and performs well on a variety of tasks.
6. **Feature Importance**: Provides an estimate of feature importance by measuring the impact of each feature on the accuracy of the model.

3.3.3: Gradient Boosting Classifier (GBC):

The GBC is a powerful and flexible ensemble learning technique used for classification tasks. It builds models incrementally by combining the outputs of several weak learners, typically decision trees, to create a strong predictive model. The fundamental idea behind GBC is to improve the performance of the model by addressing the errors made by previous models in the ensemble. GBC [30] moves ahead with repeated functions that keep on minimizing the loss function. In other words, it is a type of ensemble learning mechanism which makes sure to improvise at every iteration [31].

The skeleton of the GBC algorithm works as follows-

1. Initialize the model with log(odds) as the constant value.
2. Calculate the residuals of the predictor.
3. Create the decision trees.
4. Update the residual errors in the next iteration.
5. Continue steps 2-4 until the loss function is minimized.

Key Points:

1. **Ensemble Method**: Builds an ensemble of weak learners (usually decision trees) sequentially.
2. **Boosting**: Each subsequent tree is trained to correct the errors of the previous trees by focusing on the residual errors.
3. **Additive Model**: Combines the predictions of multiple trees by adding them together, typically with a learning rate to control the contribution of each tree.
4. **Gradient Descent**: Uses gradient descent to minimize a loss function by iteratively adding trees that reduce the overall error.
5. **Flexibility**: Can optimize various loss functions and is suitable for both classification and regression tasks.
6. **Hyperparameters**: Involves several hyperparameters such as the number of trees, tree depth, and learning rate, which need careful tuning to avoid overfitting and achieve optimal performance.

3.3.4. Naive Bayes Classifier (NBC):

The Naive Bayes Classifier (NBC) is a probabilistic machine learning algorithm based on Bayes' Theorem with an assumption of independence among features. It is widely used for classification tasks due to its simplicity, efficiency, and effectiveness in handling large datasets. NBC uses a probabilistic framework for solving classification problems [32] [33]. It is based on the use of the Bayes Theorem (Eq. 2) [34] which follows:

$$P(C / A) = (P(A / C).P(C)) / P(A) \quad (2)$$

where P(C) is the prior probability and P(C|A) is the posterior probability with respect to the occurrence of event A. In NBC, the relationship between input features and class expressed as probabilities and a naïve assumption is made the features/ attributes are independent of each other.

The algorithm follows as –

1. Get the frequency table based on the target class.
2. Calculate the respective likelihood.
3. Calculate posterior probability using Bayes Theorem [34].

3.3.5 Support Vector Machine (SVM)/ Support Vector Classifier (SVC)

The SVM is a powerful supervised learning algorithm used for classification and regression tasks. It is known for its effectiveness in high-dimensional spaces and its ability to create complex decision boundaries. The foundation of SVC[35] [36] is based on the idea of building a decision boundary or hyperplane that can divide n-dimensional space into classes so that fresh input points can be classified correctly. To create the hyperplane, SVM selects the extreme points or vectors, also referred to as support vectors.

The SVM works as follows:

1. SVM tries to find the best fit boundary to separate the classes.
2. The distance between the hyperplane and the support vector is known as margin, and the main motive is to maximize the margin for linear data points.
3. In case of non-linear data points, the concept of 3D is implemented.

4. Result and Discussion

The models have been developed and implemented using Python 3.9.1 using various standard libraries available. The results obtained after the implementation of the discussed mechanism are presented in this section.

4.1 Dataset Visualization

A thorough visual analysis of the dataset reveals several significant conclusions that provide valuable insights into its characteristics and underlying patterns. This analysis involves using various visualization techniques to explore and understand the data more deeply.

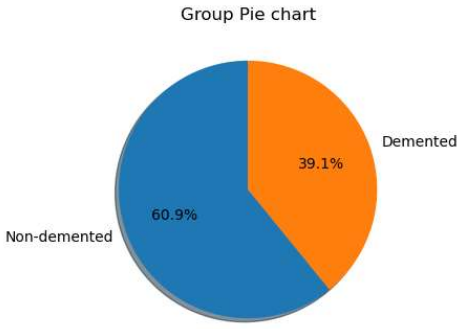

Figure 5: Group distribution

The OASIS longitudinal dataset reveals that among 373 of the total individuals under consideration, 227 of them are found to be non-demented (60.9%) while 146 of the total are demented (39.1%) according to Fig. 5. This indicates that the dataset is somewhat imbalanced, with a higher proportion of non-demented cases.

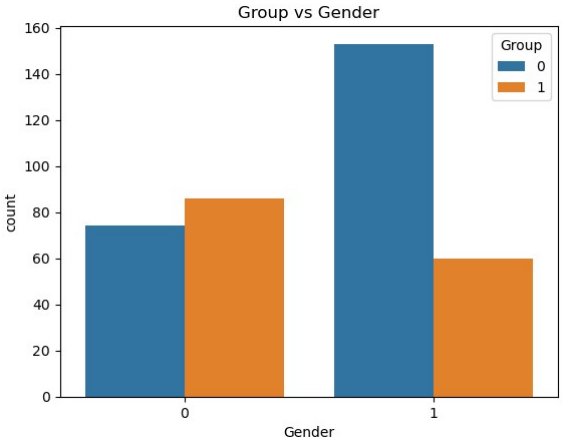

Figure 6: Group vs. Gender distribution

Men have higher chances of dementia as compared to female (0 for male and 1 for female as demonstrated in Fig. 6). This distribution provides additional context for evaluating model performance, as it shows the imbalance not only in the target variable (Demented vs. Non-demented) but also in the gender distribution.

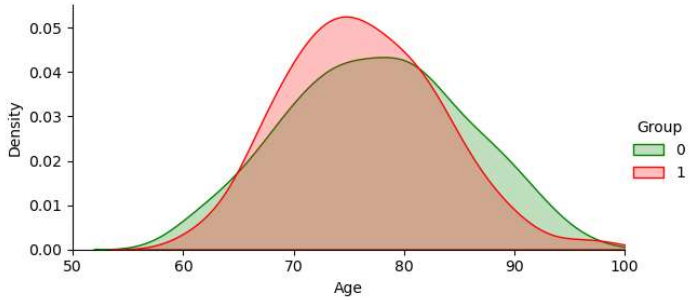

Figure 7: Group vs. Age distribution

Dementia is highly probable within the age span of 70-80 years according to Fig. 7. The plot suggests that the Demented group tends to have a higher peak density around the age of 75. The Non-demented group has a more spread out age distribution, with a peak density slightly lower than the Demented group.

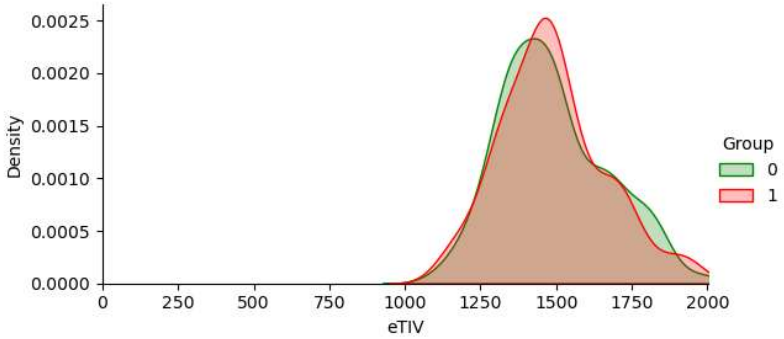

Figure 8: Group vs. eTIV distribution

Fig. 8 shows both groups have a similar distribution pattern, with a peak around the same eTIV value. Group 1 (red line) appears to have a slightly higher peak density than Group 0 (green line), indicating a higher concentration of data points around the peak eTIV value.The distribution tails off similarly for both groups on either side of the peak, but Group 1 shows a bit more spread towards higher eTIV values.

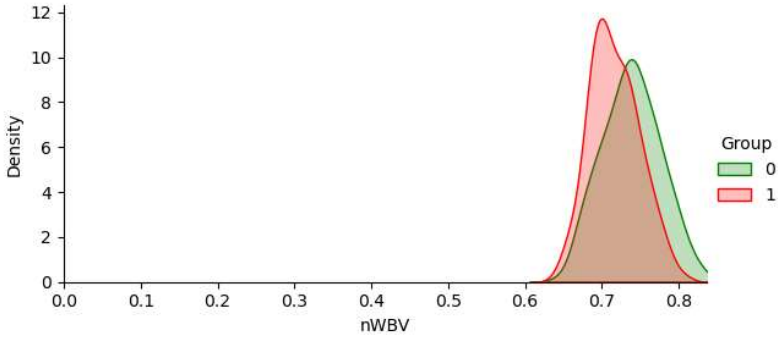

Figure 9: Group vs. nWBV distribution

Fig. 9 shows the distributions of both groups are narrower and more peaked compared to the eTIV plot. Group 1 (red line) has a higher peak density than Group 0 (green line), indicating a higher concentration of data points around the peak nWBV value. The peak for Group 1 is slightly shifted to the left of the peak for Group 0, indicating that Group 1 tends to have lower nWBV values compared to Group 0. Both groups have a very tight distribution, with most of the values falling between 0.6 and 0.8.

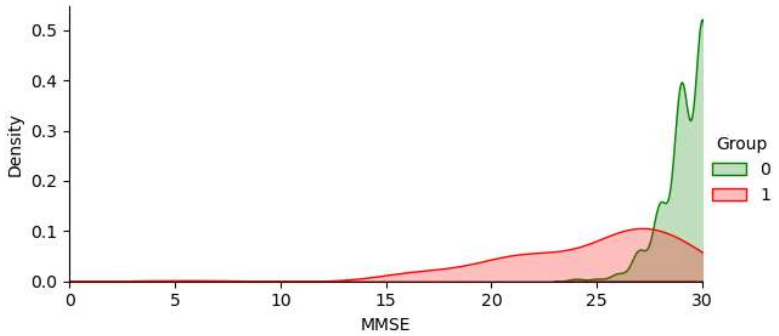

Figure 10: Group vs. MMSE distribution

In Fig. 10. Group 0 (green line) shows a significant peak at an MMSE score of 30, indicating a high concentration of individuals with the highest possible MMSE score. Group 1 (red line) has a more spread-out distribution, with a notable peak around an MMSE score of 25, but also showing a wider range of lower scores. The density for Group 0 increases sharply around the maximum MMSE score, while Group 1 shows a gradual increase and decrease in density across the score range.
This suggests that individuals in Group 0 generally perform better on the MMSE, clustering around the highest score, whereas individuals in Group 1 have more variability in their MMSE scores, with many scoring below the maximum.

4.2 Model Results

The confusion matrices for the four classification models—Random Forest (RF), Logistic Regression (LR), Gradient Boosting Classifier (GBC), and Support Vector Classifier (SVC)—reveal detailed insights into their performance in predicting positive and negative classes.
For the Random Forest (RF) model, the confusion matrix shows a total of 484 true positives (TP) and 16 false negatives (FN). This indicates that the model effectively identified positive instances while misclassifying a relatively small number. The false positive (FP) count is also 16, with 484 true negatives (TN), demonstrating a high level of accuracy in distinguishing negative cases (Table 3). The RF model performs robustly with a high number of correct predictions and a low rate of misclassifications.

Table 3: Confusion Matrix for Random Forest (RF)

	Predicted Positive	Predicted Negative
Actual Positive	484 (TP)	16 (FN)
Actual Negative	16 (FP)	484 (TN)

In the case of Logistic Regression (LR), the model correctly predicted 480 positives (TP) but missed 20 positive instances (FN). It also incorrectly classified 20 negatives as positives (FP) while correctly identifying 480 negatives (TN) (Table 4). Compared to the Random Forest model, the Logistic Regression model shows a slight increase in both false negatives and false positives, though it still maintains a strong overall performance.

Table 4: Confusion Matrix for Logistic Regression (LR)

	Predicted Positive	Predicted Negative
Actual Positive	480 (TP)	20 (FN)
Actual Negative	20 (FP)	480 (TN)

The Gradient Boosting Classifier (GBC) exhibits similar performance to Logistic Regression, with 480 true positives and 20 false negatives. It also recorded 20 false positives and 480 true negatives (Table 5). The results from GBC are consistent with those of the Logistic Regression model, indicating its effectiveness in classification tasks.

Table 5: Confusion Matrix for Gradient Boosting Classifier (GBC)

	Predicted Positive	Predicted Negative
Actual Positive	480 (TP)	20 (FN)
Actual Negative	20 (FP)	480 (TN)

The Support Vector Classifier (SVC) shows some variations in its performance metrics. It achieved 473 true positives and 27 false negatives, suggesting that it missed a higher number of positive instances compared to the other models. The SVC also had 26 false positives and 474 true negatives (Table 6). While the model still performs well, it demonstrates slightly more difficulty in balancing precision and recall compared to the others.

Table 6: Confusion Matrix for Support Vector Classifier (SVC)

	Predicted Positive	Predicted Negative
Actual Positive	473 (TP)	27 (FN)
Actual Negative	26 (FP)	474 (TN)

The Naive Bayes Classifier (NBC) has a high rate of true positives (43) and very low false positives (0), suggesting it is effective at identifying positive cases without misclassifying negatives as positives (Table 7). However, its overall performance is lower compared to other classifiers, such as Random Forest, Logistic Regression, Gradient Boosting Classifier, and Support Vector Classifier, which exhibit higher accuracy and better overall classification metrics. The Naive Bayes approach might be suitable in scenarios where computational simplicity is prioritized, but it does not match the robustness and accuracy of the more complex models like RF, LR, GBC, and SVC.

The performance analysis of various algorithms applied to the pre-processed data reveals that the Naive NBC outperforms all other models with an impressive accuracy of 97.33%. This accuracy surpasses that of the RF classifier, which achieved an accuracy of 96.8%, and both LR and GBC, which both recorded accuracies of 96.00%. The SVC lagged slightly behind with an accuracy of 94.67%. Given these results, NBC was selected for developing the prediction model due to its superior performance.

Table 7: Confusion Matrix for Naive Bayes Classifier (NBC)

	Predicted Positive	Predicted Negative
Actual Positive	43 (TP)	2 (FN)
Actual Negative	0 (FP)	29 (TN)

The Table 4, presents a comprehensive comparison of various machine learning models based on several performance metrics, including Accuracy, Precision, Recall, and F1 Score.

Table 4: Comparative analysis of the results

Model	Accuracy	Precision	Recall	F1 Score
RF [16]	96.8%	96.8%	96.8%	96.8%
LR [17]	96.00%	96.0%	96.0%	96.0%
GBC [17]	96.00%	96.0%	96.0%	96.0%
SVC [22]	94.67%	94.67%	94.67%	94.67%
NBC (Proposed)	**97.33%**	**94.1%**	**100%**	**97.0%**

Accuracy: The NBC model achieved the highest accuracy at 97.33%, surpassing the RF, LR, GBC, and SVC models. This indicates that NBC is more effective in making correct predictions overall compared to the other classifiers. RF, LR, and GBC followed closely with accuracies of 96.8% and 96.00%, respectively, while SVC had the lowest accuracy at 94.67%.
Precision: Precision measures the proportion of true positive predictions among all positive predictions made by the model. The NBC model exhibited a precision of 94.1%, which is lower than that of RF, LR, and GBC, each achieving 96.0%. However, this trade-off in precision is compensated by NBC's perfect Recall.
Recall: Recall indicates the model's ability to correctly identify all relevant positive cases. NBC outperformed all other models with a perfect recall of 100%. This demonstrates NBC's superior ability to minimize false negatives and accurately detect all positive instances, an essential attribute for applications where missing a positive case could be critical.
F1 Score: The F1 Score, which balances precision and recall, was highest for the NBC model at 97.0%, reflecting its overall robustness in handling positive cases. The RF, LR, and GBC models had an F1 Score of 96.0%, while the SVC model's F1 Score was the lowest at 94.67%.

Overall, the NBC model not only provides the highest accuracy and F1 Score but also excels in Recall, making it the most effective model among those evaluated for the task at hand.

5. Conclusion and Future works

AD is a major health concern affecting a large part of the aged population across the globe. Thus, it is more important to diagnose early symptoms of the disease accurately to reduce risk and provide early medical intervention in cases of such diseases that do not have a cure. As seen in previous works, several ML models have been developed to diagnose AD at an early stage; however, the challenge lies in achieving accurate results. In this study, we have addressed this challenge and aimed to achieve high prediction accuracy with our developed model. We presented a comparative analysis of the performance among five state-of-art classification algorithms on the chosen dataset and developed the prediction model using the one with the highest accuracy. Following the application of each algorithm to the preprocessed data, the findings show that NBC has the best performance with the maximum accuracy of 97.33%, followed by other applied algorithms like - RF (96.8%), LR (96.00%), GBC (96.00%), and SVC (94.67%). In comparison to previous works, we

achieved an appreciably high accuracy. As part of our originality, we attempted to fit our models with as many relevant features from the dataset as feasible after thorough data cleaning and pre-processing. Apart from that, it was evident that efficient data cleaning and pre-processing resulted in better performances on the considered algorithms. Such high prediction accuracy is of immense significance, especially in cases of medical data and disease prediction where a patient's life is at stake.

This research can be expanded in the future by applying neural networks for prediction methods that will aid in disease detection and analysis of new features in the dataset. In addition, we will focus on improving the accuracy of the prediction in the future scope of the study, which will boost its significance in medical disciplines.

References

1. Geldmacher, D.S., and P.J. Whitehouse. "Evaluation of Dementia." The New England Journal of Medicine (1996). https://doi.org/10.1056/NEJM199608013350507.
2. Scheltens, Phillip, Bart de Strooper, Miia Kivipelto, Helene Holstege, Gael Chételat, C.E. Teunissen, J. Cummings, and W. Flier. "Alzheimer's Disease." The Lancet 397, no. 10284 (2021): 1577-1590.
3. Wang, X., J. Qi, Y. Yang, and P. Yang. "A Survey of Disease Progression Modelling Techniques for Alzheimer's Diseases." IEEE International Conference on Industrial Informatics (2019): 1237-1242.
4. Flemming, Richard, John Zeisel, and Kirsty Bennett. World Alzheimer Report 2020. Alzheimer's Disease International, September 2020.
5. Yang, Kuo, and Emad A. Mohammed. "A Review of Artificial Intelligence Technologies for Early Prediction of Alzheimer's Disease." Electrical Engineering and Systems Science (2020). https://doi.org/10.48550/arXiv.2101.01781.
6. Wright, R. E. "Logistic Regression." In Reading and Understanding Multivariate Statistics, edited by L. G. Grimm and P. R. Yarnold, 217–244. Washington, DC: American Psychological Association, 1995.
7. Chaudhary, Archana, Savita Kolhe, and Raj Kamal. "An Improved Random Forest Classifier for Multi-Class Classification." Information Processing in Agriculture 3, no. 4 (2016): 215-222.
8. Bahad, P., and P. Saxena. "Study of AdaBoost and Gradient Boosting Algorithms for Predictive Analytics." In International Conference on Intelligent Computing and Smart Communication 2019, edited by G. Singh Tomar, N. S. Chaudhari, J. L. V. Barbosa, and M. K. Aghwariya, 22. Algorithms for Intelligent Systems. Singapore: Springer, 2020. https://doi.org/10.1007/978-981-15-0633-8_22.
9. Rish, I. "An Empirical Study of the Naive Bayes Classifier." In IJCAI, vol. 3, 2001.
10. Evgeniou, T., and M. Pontil. "Support Vector Machines: Theory and Applications." In Machine Learning and Its Applications: Advanced Lectures, 249-257. Berlin: Springer, 2001. https://doi.org/10.1007/3-540-44673-7_12.
11. Shahbaz, Muhammad, Shahzad Ali, Aziz Guergachi, Aneeta Niaz, and Amina Umer. "Classification of Alzheimer's Disease using ML Techniques." In ICDS19, vol. 2, 296-303. https://doi.org/10.5220/0007949902960303.
12. Marinescu, Razvan V., Neil P. Oxtoby, Alexandra L. Young, Esther E. Bron, Arthur W. Toga, Michael W. Weiner, Frederik Barkhof, Nick C. Fox, Stefan Klein, Daniel C. Alexander, and the EuroPOND Consortium. "TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease." Quantitative Biology (2018).
13. Khan, Afreen, and Swaleha Zubair. "Development of a Three Tiered Cognitive Hybrid ML Algorithm for Effective Diagnosis of Alzheimer's Disease." Journal of King Saud University - Computer and Information Sciences 34, no. 10, Part A (2022).
14. Soliman, Sarah A., El-Sayed A. El-Dahshan, and Abdel-Badeeh M. Salem. "Predicting Alzheimer's Disease with 3D Convolutional Neural Networks." International Journal of Applications of Fuzzy Sets and Artificial Intelligence 1 (2020): 125-146.
15. Shastry, K.A., V. Vijayakumar, M.K.M. V, M. B A, and C. B N. "Deep Learning Techniques for the Effective Prediction of Alzheimer's Disease: A Comprehensive Review." Healthcare 10 (2022): 1842. https://doi.org/10.3390/healthcare10101842.
16. Antor, Morshedul Bari, A. H. M. Shafayet Jamil, Maliha Mamtaz, Mohammad Monirujjaman Khan, Sultan Aljahdali, Manjit Kaur, Parminder Singh, and Mehedi Masud. "A Comparative Analysis of ML

Algorithms to Predict Alzheimer's Disease." Journal of Healthcare Engineering (2021). https://doi.org/10.1155/2021/9917919.
17. Varun, Krishna Kumar, Shankar Hamritha, and Mavuthanahalli Channabasavegowda Vinay. "Prediction of Alzheimer's Disease Using ML." In Proceedings of Third International Conference on Communication, Computing and Electronics Systems, Lecture Notes in Electrical Engineering, vol. 844. Singapore: Springer. https://doi.org/10.1007/978-981-16-8862-1_50.
18. Sivakani, R., and G. A. Ansari. "ML Framework for Implementing Alzheimer's Disease." 2020 International Conference on Communication and Signal Processing (ICCSP) (2020): 588-592. https://doi.org/10.1109/ICCSP48568.2020.9182220.
19. Hosseinzadeh Kasani, Payam, Sara Hosseinzadeh Kasani, Yeshin Kim, Cheol-Heui Yun, Seong Hye Choi, and Jae-Won Jang. 2021 International Conference on Information and Communication Technology Convergence (ICTC). Jeju Island, Republic of Korea, October 2021.
20. Alessia, Sarica, Antonio Cerasa, and Aldo Quattrone. "Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review." Frontiers in Aging Neuroscience 9 (2017). https://doi.org/10.3389/fnagi.2017.00329.
21. Rangaswamy, Uday, S. Akila Parvathy Dharshini, Dhanusha Yesudhas, and M. Michael Gromiha. "VEPAD - Predicting the Effect of Variants Associated with Alzheimer's Disease Using ML." Computers in Biology and Medicine 124 (2020): 103933. https://doi.org/10.1016/j.compbiomed.2020.103933.
22. Kishore, P., Ch. Usha Kumari, M.N.V.S.S. Kumar, and T. Pavani. "Detection and Analysis of Alzheimer's Disease Using Various ML Algorithms." Materials Today: Proceedings 45, Part 2 (2021).
23. Khan, N. M., N. Abraham, and M. Hon. "Transfer Learning With Intelligent Training Data Selection for Prediction of Alzheimer's Disease." IEEE Access 7 (2019): 72726-72735. https://doi.org/10.1109/ACCESS.2019.2920448.
24. Liu, J., J. Wang, B. Hu, F.-X. Wu, and Y. Pan. "Alzheimer's Disease Classification Based on Individual Hierarchical Networks Constructed With 3-D Texture Features." IEEE Transactions on NanoBioscience 16, no. 6 (2017): 428-437. https://doi.org/10.1109/TNB.2017.2707139.
25. Huang, M., W. Yang, Q. Feng, et al. "Longitudinal Measurement and Hierarchical Classification Framework for the Prediction of Alzheimer's Disease." Scientific Reports 7 (2017): 39880. https://doi.org/10.1038/srep39880.
26. Austin, J. T., R. A. Yaffee, and D. E. Hinkle. "Logistic Regression for Research in Higher Education." In Higher Education: Handbook of Theory and Research, vol. 8, 379–410. New York: Agathon Press, 1992.
27. Cabrera, A. F. "Logistic Regression Analysis in Higher Education: An Applied Perspective." In Higher Education: Handbook of Theory and Research, vol. 10, 225–256. New York: Agathon Press, 1994.
28. Biau, Gerard. "Analysis of a Random Forests Model." Journal of Machine Learning Research 13 (2012): 1063-1095.
29. Kulkarni, Vrushali Y., and Pradeep K. Sinha. "Effective Learning and Classification using Random Forest Algorithm." International Journal of Engineering and Innovative Technology (IJEIT) 3, no. 11 (2014).
30. Friedman, Jerome H. "Greedy Function Approximation: A Gradient Boosting Machine." Annals of Statistics 29, no. 5 (2001): 1189-1232. https://doi.org/10.1214/aos/1013203451.
31. Natekin, A., and A. Knoll. "Gradient Boosting Machines, a Tutorial." Frontiers in Neurorotics 7 (2013).
32. Yang, F.-J. "An Implementation of Naive Bayes Classifier." 2018 International Conference on Computational Science and Computational Intelligence (CSCI) (2018): 301-306. https://doi.org/10.1109/CSCI46756.2018.00065.
33. Nagarani, V., A. Karpagam, B. Vasuki, and N. Sundarakannan. "Real Life Applications of Bayes Theorem." JAC: A Journal Of Composition Theory XIV, no. II (2022).
34. Puga, J. L., and M. Krzywinski. "Bayes' Theorem." Nature Methods (2015): 277–278. https://doi.org/10.1038/nmeth.3335.
35. Vishwanathan, S. V. M., and M. Narasimha Murty. "SSVM: A Simple SVM Algorithm." In Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02, vol. 3, 2393-2398. Honolulu, HI, USA, 2002. https://doi.org/10.1109/IJCNN.2002.1007516.
36. Pisner, Derek A., and David M. Schnyer. "Support Vector Machine." In Machine Learning, Methods and Applications to Brain Disorders, 101-121. 2020. https://doi.org/10.1016/B978-0-12-815739-8.00006-7.
37. Oasis longitudinal dataset : https://www.kaggle.com/datasets/ninadaithal/imagesoasis