www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 04, April 2023, pp : 327-332

e-ISSN : 2583-1062

Impact Factor : 5.725

# ALZHEIMER'S DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS

**Farooq Sunar Mahammad[1], P. Samba Siva Rao[2], A. Supriya[3], D. Usha[4], P. Asha Preethi[5], B. Bhuvaneswari Devi[6], T. Geetha[7]**

[1,2]Professor, Department of Computer Science and Engineering, Santhiram Engineering College, Nandyal

[3,4,5,6,7]Department of Computer Science and Engineering, Santhiram Engineering College, Nandyal

E-mail: hod.cse@srecnandyal.edu.in

DOI: https://www.doi.org/10.58257/IJPREMS30899

## ABSTRACT

Alzheimer's disease is a neurodegenerative disorder that causes a progressive decline in cognitive function and memory loss. It is the most common cause of dementia in older adults, accounting for 60-80% of all dementia cases. The disease is characterized by the accumulation of abnormal protein deposits in the brain, including beta-amyloid plaques and tau protein tangles, which disrupt normal brain function and ultimately lead to the death of brain cells. The exact cause of Alzheimer's disease is not fully understood, but it is thought to be the result of a combination of genetic, environmental, and lifestyle factors. In this paper, we build a model to predict Alzheimer's disease and know whether the person is cured, normal, or has Alzheimer's Disease. By taking the dataset, preprocessing the dataset, and using Machine Learning Algorithms we classify whether the person comes under which category by considering the parameters such as group, number of visits, MR Delay, MMSE, Education, etc.

**Keywords:** Alzheimer's disease, moderate cognitive impairment, machine learning methods, psychosocial characteristics.

## 1. INTRODUCTION

Alzheimer's disease is a progressive and irreversible brain disorder that affects memory, thinking, and behavior. Early detection and accurate prediction of this disease can lead to better treatment outcomes and improve the quality of life for patients and their families.Machine learning algorithms offer a promising approach to predicting Alzheimer's disease by analyzing large datasets of patient information, such as medical histories, brain imaging, and genetic data. These Machine Learning algorithms can identify potential risk factors associated with the disease, and generate predictive models that can help clinicians make more informed decisions about patient care.

## 2. LITERATURE REVIEW

### A. Machine Learning:

Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed. In other words, the algorithms are designed to learn from data and improve their performance over time through a process of trial and error. The goal of machine learning is to enable machines to learn from data, recognize patterns, and make decisions based on that knowledge. Machine learning algorithms can automate repetitive tasks, freeing up human time and increasing efficiency in many industries. Machine learning models can analyze vast amounts of data to identify patterns and make accurate predictions, leading to better decision-making and improved outcomes.

### B. Multistage classifier-based approach for Alzheimer's Disease prediction and retrieval:

The most prevalent and common type of dementia is Alzheimer's disease (AD). However, it is notable that very few people who are suffering from AD are diagnosed correctly and in a timely manner. The definite cause and cure of the disease are still unavailable. The symptoms might be more manageable and its treatment can be more effective when the impairment is still at an earlier stage or at MCI (mild cognitive impairment). AD can be clinically diagnosed by physical and neurological examination, so there is a need for developing better and more efficient diagnostic tools for AD. In recent years, content-based image retrieval (CBIR) systems have been widely researched and applied in many medical applications. Combining an automated image classification system and the radiologist's professional knowledge, to increase the accuracy of prediction and diagnosis, were the main motives. In this paper, machine learning classifiers, including Decision Trees, Random Forests, and Gradient Boosting, were used to classify Alzheimer's disease more acceptably and efficiently.

## 3. METHODOLOGY

In this section, the methodology was adopted in order to predict Alzheimer's disease. More specifically in section A, the dataset information is described. And Section B consists of evaluation metrics.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 04, April 2023, pp : 327-332

e-ISSN : 2583-1062

Impact Factor : 5.725

**A. Dataset Information**

It consists of 15 attributes which are described as follows :

• Subject.ID - Unique Id of the patient

• MRI.ID - Unique Id generated after conducting MRI on patient

• Group - It is a group of Converted (Previously Normal but developed dementia later), Demented, and Non-demented (Normal Patients)

• Visit - Number of visits to detect dementia status

• MR Delay – The time of delay

• M.F - Gender

• Hand-Handedness (actually all subjects were right-handed so I will drop this column)

• Age - Age in years

• EDUC - Years of education

• SES - Socioeconomic status as assessed by the Hollingshead Index of Social Position and classified into categories from 1 (highest status) to 5 (lowest status)

• MMSE - Mini-Mental State Examination score (range is from 0 = worst to 30 = best)

• CDR - Clinical Dementia Rating (0 = no dementia, 0.5 = very mild AD, 1 = mild AD, 2 = moderate AD)

• eTIV - Estimated total intracranial volume, mm3

• nWBV - Normalized whole-brain volume

• ASF - Atlas scaling factor (unit-less).

**B. Evaluation Metrics**

• Accuracy: It is the percentage of examples correctly classified.

$$\text{accuracy} = \frac{Number\ of\ correct\ predictions}{Number\ of\ total\ values}$$

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

• Recall: It is the percentage of actual positives that were correctly classified.

$$\text{recall} = \frac{TP}{TP+FN}$$

• Precision: is the percentage of predicted positives that were correctly classified.

$$\text{precision} = \frac{TP}{TP+FP}$$

•F1-score: a combination of recall and precision to get a single measure, which falls between these two metrics

$$\text{F1-score} = 2\frac{precision.recall}{precision+recall}$$

## 4. IMPLEMENTATION AND ANALYSIS

In this section, the implementation details are mentioned to detect Alzheimer's disease. It contains the model selection, and the analysis that has done, and its accuracy is shown.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 04, April 2023, pp : 327-332

Impact Factor : 5.725

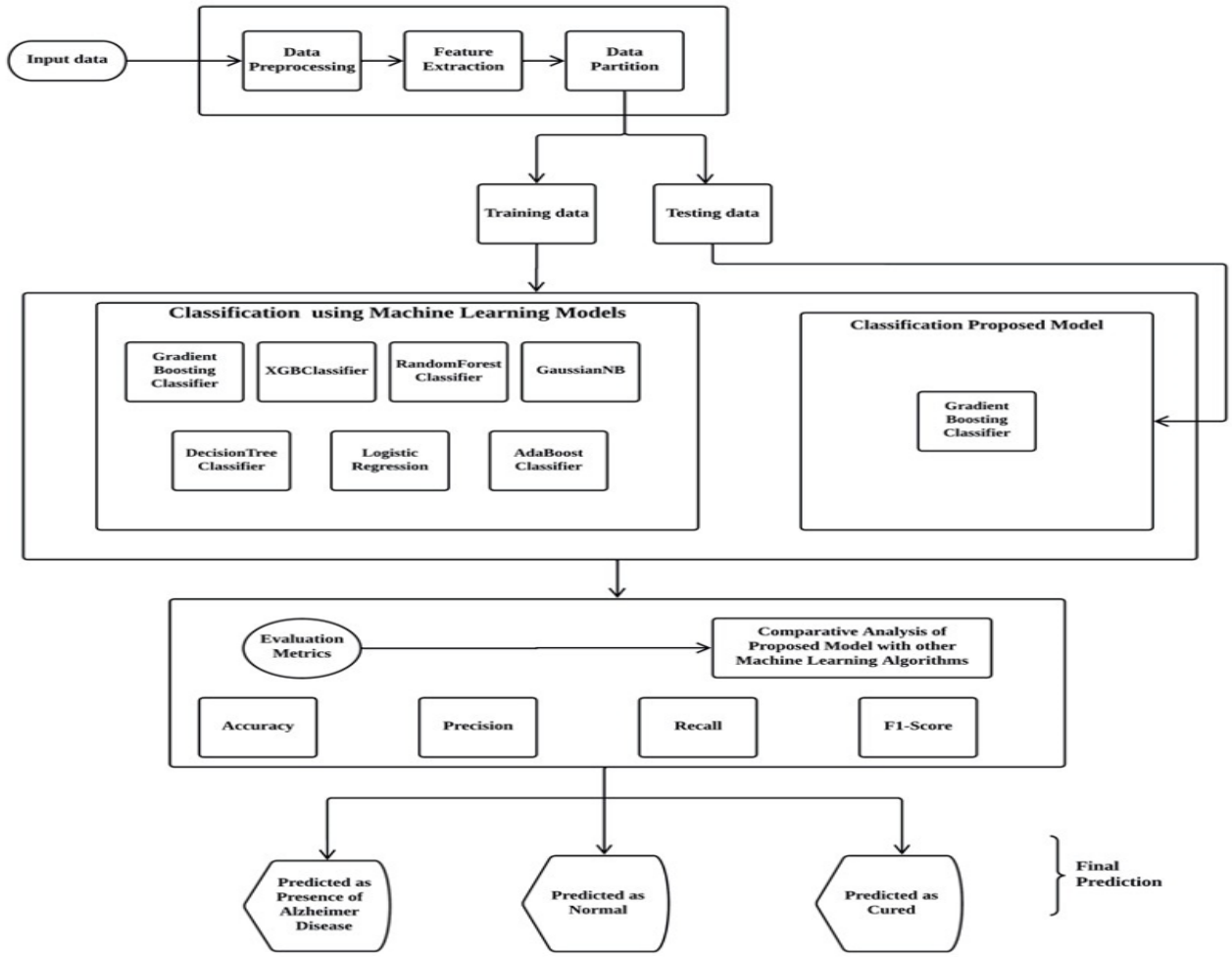


**Figure1 :** System Architecture

**Model - 1: Decision Tree Classifier:**

A decision tree classifier is a machine learning algorithm that is commonly used for classification tasks. It works by recursively partitioning the data into subsets based on the values of different input features, using a series of binary decisions. Each decision node in the tree represents a test of a specific feature, while each leaf node represents a predicted class label. The goal of the algorithm is to build a decision tree that can accurately predict the class label of new, unseen data. Decision tree classifiers are relatively easy to interpret and visualize, making them a popular choice for many applications.

**Model-2: Random Forest Classifier**

-Random Forest Classifier is a machine learning algorithm used for classification problems. It is a type of ensemble learning algorithm that combines multiple decision trees to create a strong learner. In Random Forest Classifier, each decision tree is trained on a random subset of the input data, and the final prediction is made by averaging the predictions of all the trees.

**Model-3: Gradient Boosting Classifier**:

In Gradient Boosting Classifier, the model is trained sequentially, where each new tree is trained to correct the errors of the previous tree. The algorithm starts by training a single decision tree on the input data, and then iteratively adds new trees to the ensemble. In each iteration, the algorithm calculates the gradient (or the derivative) of the loss function with respect to the predictions of the current ensemble. The new tree is then trained to predict the negative gradient, so that the overall ensemble can minimize the loss function.

**Model-4: Naïve Bayes Classifier**:

Naive Bayes classifier is a popular probabilistic machine learning algorithm used for classification tasks. Gaussian Naive Bayes is a variant of the Naive Bayes algorithm that assumes the input data follows a Gaussian or normal distribution. The algorithm works by calculating the probability of each class given the input features, using Bayes' theorem. It assumes that each feature is independent of the others, hence the term "naive". Despite this simplifying assumption, Naive Bayes classifiers can perform well in many real-world applications, particularly in text

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 04, April 2023, pp : 327-332

e-ISSN : 2583-1062

Impact Factor : 5.725

classification tasks. Gaussian Naive Bayes is particularly suited to continuous data where the feature values can be modeled using a normal distribution.

**Model-5: Logistic Regression**:

Logistic Regression is a statistical method used for binary classification problems, where the goal is to predict a binary outcome (e.g., yes/no, true/false). It is a type of supervised learning algorithm that learns a linear decision boundary between two classes based on a set of input features.The logistic regression model is based on the logistic function, which maps any input value to a probability value between 0 and 1. This probability value represents the likelihood of the input belonging to a particular class. The logistic function is also known as the sigmoid function because of its S-shaped curve. The logistic regression model is trained using a labeled dataset, where each data point is associated with a label indicating its class. The model learns the weights of the input features that maximize the likelihood of the observed data, given the class labels. These weights are then used to predict the class of new, unseen data points.

**Model-6: Ada Boost Classifier:**

AdaBoost (Adaptive Boosting) is a popular machine learning ensemble algorithm that is used for classification tasks. It works by combining several simple base models, often decision trees, to create a powerful predictive model. In AdaBoost, each base model is trained on a random subset of the training data, and the algorithm assigns weights to the data points based on the errors made by the base model. In the subsequent iteration, the algorithm assigns higher weights to the data points that were misclassified in the previous iteration and trains a new base model on the updated weighted data. This process is repeated several times, and the final prediction is made by taking a weighted average of the predictions made by all the base models. AdaBoost has several advantages, including its ability to handle high-dimensional data, its low tendency to overfit, and its versatility in working with different base models.

**Model-7: XG Boost Classifier-**

XG Boost (Extreme Gradient Boosting) is a machine learning algorithm that uses gradient boosting to build a more powerful ensemble model. It is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. XGBoost has gained popularity in the machine learning community due to its scalability, speed, and accuracy in various competitions, such as the Kaggle data science competitions. It uses decision trees as base learners and combines their results to improve the overall accuracy of the model. XGBoost is particularly useful when dealing with large datasets with many features and can handle missing values in the input data. It also provides built-in methods for regularization to prevent overfitting and early stopping to avoid training for too long. Overall, XGBoost is a powerful and flexible tool for classification and regression problems and is widely used in industry and academia.

## 5. RESULTS

Analysis of data is done by using different Machine Learning Models which helps us to choose the best model in order to predict Alzheimer Disease. We have used different Machine Learning Models such as XGB Boost Classifier, Gradient Boosting Algorithm, Decision Tree Algorithm, Random Forest Algorithm, Logistic Regression, Navies Bayes Algorithm, and Ada Boost Classification Algorithm.

In the bar plot, the accuracy of each model has been plotted.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 04, April 2023, pp : 327-332

e-ISSN : 2583-1062

Impact Factor : 5.725

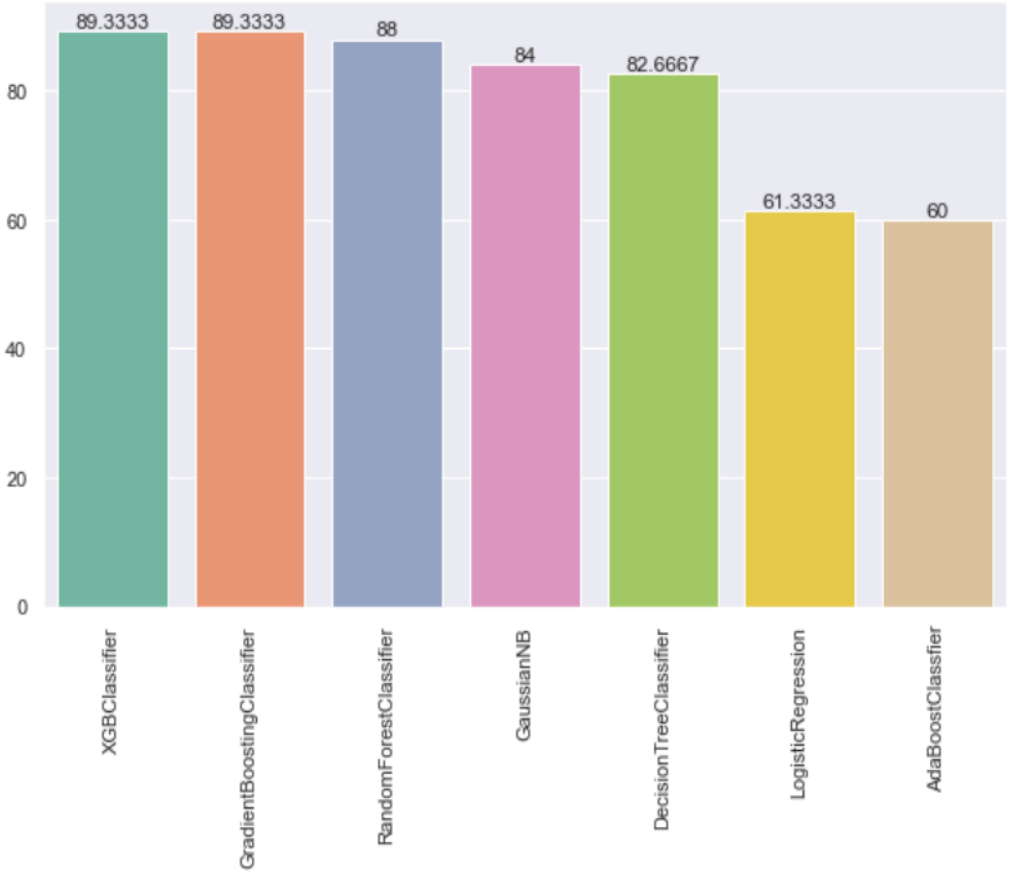


**Figure 2:** Accuracy scores on test data

- The Accuracy score for testing data using XGBoostClassifer is 89.3%.
- The Accuracy score for testing data using Gradient Boosting is 89.3%.
- The Accuracy score for testing data using RandomForestClassfier is 88.0%.
- The Accuracy score for testing data using DecisionTreeClassifier is 84.0%.
- The Accuracy score for testing data using NaiveBayes(GaussianNB) is 82.7%.
- The Accuracy score for testing data using Logistic Regression is 61.3%.
- The Accuracy score for testing data using AdaBoostClassifier is 60.0%.

| Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| GradientBoostingClassifier | 89.3 | 90.6 | 89.3 | 87.5 |
| XGBClassifier | 89.3 | 90.6 | 89.3 | 87.5 |
| RandomForestClassifier | 88.0 | 89.6 | 88.0 | 85.4 |
| GaussianNaviesBayes | 84.0 | 81.2 | 84.0 | 81.2 |
| DecisionTreeClassifier | 82.7 | 80.5 | 82.7 | 81.2 |
| LogisticRegression | 61.3 | 56.2 | 61.3 | 58.1 |
| AdaBoostClassifier | 60.0 | 86.6 | 60.0 | 56.2 |

**Table I:** Performance Comparison of All Algorithms Implemented

## 6. CONCLUSION

In our paper, we used an algorithm that predicts Alzheimer's Disease. Gradient Boosting classifier and XGBoostClassfier algorithm was found to be the best learning model with a test accuracy of 89.3% and a recall score of 87.5%.

## 7. REFERENCES

[1] Jyothi, V., and M. V. Subramanyam. "An enhanced routing technique to improve the network lifetime of the cognitive sensor network." Wireless PersonalCommunications 127.2 (2022): 1241-1264.

[2] Sreelatha, Tammineni, M. V. Subramanyam, and MN Giri Prasad. "Early detection of skin cancer using melanoma segmentation technique." Journalof medical systems 43.7 (2019): 190.

[3] Rao, Y. Mallikarjuna, M. V. Subramanyam, and K. Satya Prasad. "Cluster based hybrid routing protocol for wireless mesh networks." Wireless Personal Communications 103 (2018): 3009- 3023.

[4] Sreelatha, Tammineni, M. Subramanyam, and MN Giri Prasad. "Early detection of skin cancer using melanoma segmentation technique." Journalof medical systems 43.7 (2019): 190.

[5] Bhaskar, P., Farooq Sunar Mahammad, A. Hemanth Kumar, D. Raj Kumar, SM Abdul Khadar, P. Moin Khan, and P. Veer Sekhar Reedy."Machine Learning Based Predictive Model for Closed Loop Air Filtering System." JOURNAL OF ALGEBRAIC STATISTICS 13, no. 3 (2022): 609-616.

[6] Devi, M. Sharmila, Farooq Sunar Mahammad, D. Bhavana, D. Sukanya, TV Sai Thanusha, M. Chandrakala, and P. Venkata Swathi. "Machine Learning Based Classification and Clustering Analysis of Efficiency of Exercise Against Covid-19 Infection." JOURNAL OF ALGEBRAICSTATISTICS 13, no. 3 (2022): 112-117.

[7] Naik, S. Md Riyaz, et al. "Crowd Prediction at Various Public Places for Covid-19 Spread Prevention Employing LinearRegression." JOURNAL OF ALGEBRAIC STATISTICS 13.3 (2022): 08-16.

[8] Mahammad, Farooq Sunar, et al. "A comprehensive research on video imaging techniques." (2019).

[9] Ramasamy, Palanisamy, Karthik Balasubramanian, and Farooq Sunar Mahammad. "Comparative analysis of 3D-SVM and 4D-SVM for five-phase voltage source inverter." International Transactions on ElectricalEnergy Systems 31.12 (2021): e13138.

[10] Sunar, Mahammad Farooq, and V. Madhu Viswanatham. "A fast approach to encrypt and decrypt of video streams for secure channel transmission." World Review of Science, Technology and Sustainable Development 14.1(2018): 11-28.