**FMDB** *Transactions on Sustainable Health Science Letters*

# A Comprehensive Comparison of Adaboost, Lightgbm, and Catboost for Alzheimer's Disease Classification

**Malini Premakumari William[1,*], S. Silvia Priscila[2], S. Briskline Kiruba[3], T. R. Nisha Dayana[4], P. Velavan[5]**

[1]Department of Computer Systems Engineer, Mirage Software Inc., DBA Bourntec Solutions, Schaumburg, Illinois, United States of America.
[2]Department of Computer Science, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
[3]Department of Computer Science, Jaya College of Arts and Science, Chennai, Tamil Nadu, India.
[4]Department of Computer Science, Vels Institute of Science Technology and Advance Studies, Chennai, Tamil Nadu, India.
[5]Department of Computer Science and Business Systems, Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India.
malinipremakumari@bourntec.com[1], silviaprisila.cbcs.cs@bharathuniv.ac.in[2], kiruba.briskline@gmail.com[3], nisha.dayana1984@gmail.com[4], velavan@dhaanishcollege.in[5]

**Abstract:** Alzheimer's disease has been a serious public health issue that needs prompt and precise diagnosis in the management and treatment phase. The following paper presents a machine learning system mainly engineered to classify AD via neuroimaging and clinical features comprising air time1, disp index1, gmrt in air1, max x extension1, and max y extension1 to detect subtle patterns possibly indicative of Alzheimer's pathology. Systematic data preprocessing and visualization techniques, such as correlation heatmaps and 3D scatter plots, reflect complicated interdependencies in AD progression. The classifiers utilized include Cat Boost and Light GBM for AD and healthy controls, and AdaBoost for mild cognitive impairment. Model performance is measured by accuracy, precision, recall, F1 score, classification reports, and confusion matrices, helping to discover model strengths and weaknesses. A learning curve showed that the models are general and flexible enough for practical circumstances. The work shows that machine learning can integrate multi-modal data modalities into AD diagnosis. Thus, it improves diagnostic accuracy and enables personalized treatment strategies, improving patient outcomes and supporting cutting-edge neurodegenerative disease clinical decisions.

**Keywords:** Alzheimer's Disease (AD); Precise Diagnosis; Robust Machine Learning System; Correlation Heatmaps; Unravel Complex Data; Underlying Mechanisms; Distinguishing Nuanced; Adaboost Classifiers.

**Received on:** 27/04/2024, **Revised on:** 05/07/2024, **Accepted on:** 29/08/2024, **Published on:** 01/09/2024

**Journal Homepage:** https://www.fmdbpub.com/user/journals/details/FTSHSL

**DOI:** https://doi.org/10.69888/FTSHSL.2024.000253

**Cite as:** M. P. William, S. S. Priscila, S. B. Kiruba, T. R. N. Dayana, and P. Velavan, "A Comprehensive Comparison of Adaboost, Lightgbm, and Catboost for Alzheimer's Disease Classification," *FMDB Transactions on Sustainable Health Science Letters*, vol.2, no.3, pp. 175–187, 2024.



## 1. Introduction

*Corresponding author.

Alzheimer's disease is a debilitating neurological disorder that affects the brain, causing a deterioration in cognitive skills and memory [2]. It represents the most prevalent form of dementia, accounting for between 60 and 70 percent of all dementia cases. Alzheimer's disease progressively wears down memory, reasoning ability, and behavior and eventually starts to influence an individual's ability to perform common functions, which is also discussed by Dong et al. [5]. With the condition's advancement [11], language, decision-making ability, and personality may be affected. Two types of aberrant protein deposits are found in the brain in the context of Alzheimer's disease: beta-amyloid plaques, and tau tangles, causing destruction of brain cells and breaking of connections between neurons. The cause of Alzheimer's disease is not well defined, though it seems to be the result of a multifaceted interaction between hereditary, environmental, and lifestyle factors. At the present moment, there is no cure for Alzheimer's disease, and what current treatments offer mainly involves management and slowing down its progression, as studied by Li et al. [7]. With the rising incidence of Alzheimer's disease, particularly in the aging population, researchers continue to understand what causes lead to this disease. The cause of Alzheimer's disease remains unknown but is postulated to result from interactions of multiple genetic, environmental, and lifestyle factors [15]. In this project, we proposed using machine learning to chip away at the complexity of Alzheimer's disease, a severe neurological disorder affecting millions of people worldwide.

The desired objective is to build predictive models that detect Alzheimer's disease much earlier, thus allowing better clinical management, patient outcomes, and quality of life at the tail end. We considered three machine-learning models in this study: CatBoost, LightGBM, and AdaBoost. Each of these possesses definitive strengths and capabilities that will allow us to test several approaches in predicting and categorizing Alzheimer's disease studied by Dong et al. [5]. The study's outcome can greatly impact research on Alzheimer's disease and the practice of therapy. Identifying accurate and reliable markers of the onset and progression of Alzheimer's disease enables us to empower healthcare providers to intervene early and deliver targeted therapies for improving the outcomes of patients and our knowledge about this dreadful disease. Useful insights for improving techniques in the fight against Alzheimer's disease will come from comprehensive reviewing and comparing various machine learning models. We compare and analyze throughout this study the performance of these three models on several measures such as accuracy, precision, recall, and F1-score. We aim to select the best method to classify the disorder of Alzheimer's and distinguish it from other cognitive disorders or healthy states using an in-depth review of these models. For our study, we review three machine learning models -daBoost, CatBoost, and LightGBM -to classify for Alzheimer's disease. Light GBM is also highly efficient, processes large amounts of data, and can provide an analysis of feature importance. This review comprehensively indicates performance and applicability so much is needed to guide future research and clinical applications in diagnosing and treating Alzheimer's disease.

### 1.1. Objective of the project

The fine-tuning of parameters for each algorithm—LightGBM, AdaBoost, and CatBoost—is very important, ensuring good accuracy, precision, recall, and F1 scores for the optimized model. Fine-tuning is also essential for the strong and stable performance in classifying the individuals, which would form a foundation in the personalized treatment planning approaches to optimize care strategies that caters to the patient's personal needs, thereby developing an effective management strategy for Alzheimer's disease. With the use of models, this approach utilizes machine learning to classify data and identify and analyze complex patterns of underlying cognitive health data, bringing forth a better understanding of AD. The system also focused on interpretability and transparency, with insight into the most pivotal features in a decision when classifying data. Such clarity enables clinicians to decide and accredit trust in launching machine learning into healthcare. In a nutshell, the revolutionizing capability of machine learning techniques in AD diagnosis is at play. The support and promotion of integrated analysis methodologies contribute to developing high-class decision-support systems, providing customized treatment plans and paving the way toward better patient outcomes and new standards for diagnosing and managing neurodegenerative diseases.

### 1.2. Problem Statement

Alzheimer's disease is the most lethal form of neurodegenerative disorder, which affects millions of lives around the globe. Early detection and proper classification of AD and interventions dictate timely treatment. Recent research has captured the possible classification of AD based on different types of biomarkers and clinical data using some machine learning techniques. Investigate how well these three state-of-the-art gradient boosting algorithms, namely LightGBM, AdaBoost, and CatBoost, classify patients with Alzheimer's disease. The main objectives include comparing how well these algorithms classify patients with Alzheimer's compared to controls and distinguishing all three based on feature representations derived from neuroimaging data, genetic markers, and clinical assessments.

Alzheimer's disease is a major public health challenge identified with a step-by-step degradation of cognitive functions and loss of memory, which requires early screening to stage appropriate interventions and treatments. True AD classification relies on integrating neuroimaging techniques, MRI and PET, genetic markers, and clinical estimations. Meanwhile, LightGBM, AdaBoost, and CatBoost, among other machine learning approaches show great potential in drastically improving predictive

precision. These high-performance gradient boosters shine in these classification tasks because they efficiently handle high-dimensional data and capture complex nonlinear relationships that can give rise to strong predictive accuracy. Machine learning in healthcare extends its utility in analyzing complex datasets, leading to deeper insight into AD's underlying mechanisms and biological, genetic and environmental factors contributing to its progression. This multidisciplinary approach focuses on optimizing the existing diagnostic methods, including biomarkers and imaging techniques, and novel, efficient tools for early detection and cognitive assessment. The design of this project attempts to bridge the current gap between computational advances and practice in clinics such that this advancement in the classification of AD can be refined enough to ensure interventions are conducted well in time toward improvement in patient care and treatment outcomes. In conclusion, efforts herein aim to better understand the progressive nature of AD while challenging critical healthcare challenges in its management.

## 2. Literature Review

Afzal et al. [11] reviewed the different techniques for Alzheimer's Disease detection, citing machine learning to move the whole diagnostic process forward. Their contribution proved that a huge collection of complex datasets could now be aggregated and applied using supervised learning algorithms that accurately advance the approach to classifying AD stages. Underlining the need for a confluence of clinical and computational approaches for the early detection and timely intervention in AD management. Shukla et al. [12] gave a holistic review of automatic pipelines and machine learning strategies for AD detection, focusing on the efficiency of such techniques in managing diverse data modalities. This study highlighted the potential of automated systems in streamlining the diagnostic process, minimizing errors in such interactions, and ensuring consistency in output. Their review demonstrated how algorithm developments help address challenges arising from high-dimensional neuroimaging data.

Al-Shoukry et al. [13] have done a mini-review of deep learning algorithms on detection, including their potential in analyzing neuroimaging and biomarker datasets. They noted that this advances the field of computational neuroimaging by noting various advantages regarding CNNs and RNNs in capturing spatial and temporal patterns associated with AD. Ebrahimi and Luo [14] studied the use of CNN in AD detection, in which they analyzed MRI images. Their work showed that CNNs are efficient while performing feature extraction and classification. It, therefore, solved problems in medical imaging, such as noise and high dimensionality. Improvement in accuracy and reliability in diagnostics, according to techniques related to deep learning, is emphasized.

Mehanna [15] discussed wider aspects of healthy aging and issues about cognitive impairments like AD. The paper underlined the importance of preventive approaches and early diagnosis to control neurodegenerative diseases. This perspective provides a more valid approach for understanding the impact of AD at societal levels especially how there is a need for innovative diagnostic solutions. Altinkaya et al. [16] discussed only deep-learning methods tailored for AD detection with MRI images, where the segmentation and classification of these images have been improved significantly. The following research pointed out automatic solutions for handling large-scale neuroimaging datasets that increased efficiency and scalability in diagnostic systems.

Kivipelto et al. [17] discussed lifestyle interventions as a means of preventing cognitive impairments such as dementia and Alzheimer's disease. The research highlighted diet and physical activity as the main modifiable risk factors and advanced diagnostics, aiming to prevent AD through these dual approaches. A deep MRI segmentation approach using convolutional methods was proposed by Allioui et al. [18] to detect AD. The work demonstrated the ability and effectiveness of the segmentation process in identifying structural changes in the brain that accompany the disease. Implications drawn from the research mentioned that segmentation models can be seamlessly integrated into classification algorithms for an all-around analysis.

Aiming to detect AD, Seifallahi et al. [6] applied machine learning to analyze the Timed Up and Go test using Kinect V.2 camera data. Their research demonstrated the potential of motion analysis in providing non-invasive, cost-effective diagnostic alternatives. This approach highlighted the growing role of sensor-based systems alongside traditional neuroimaging techniques. Li et al. [7] analyzed feature extraction and identification of AD with the latent factor approach on multi-channel EEG data. Their study is represented, as it highlights the potential of neural signal data in deciding between AD and normal aging since it underlines the role innovative feature engineering plays in enhancing a classification model. Zhang et al. [8] developed a tensor multi-task ensemble learning framework for dynamic AD prediction based on changes in brain structure. The study effectively showcased the merit of an interpretable model in providing insights regarding disease progression and showed excellent temporal dynamics in AD.

Ahmed et al. [9] demonstrated the power of ensemble methods, particularly patch-based classifiers, when diagnosing AD. The authors have demonstrated how using multiple classifiers may improve diagnostic accuracy and reliability to overcome the challenges presented by imbalanced datasets. Basher et al. [1] in their study, proposed a framework for Alzheimer's diagnosis

based on structural MRI data by using CNN and DNN. The authors used hippocampal volume as a crucial biomarker in integrating machine learning models such as AdaBoost, LightGBM, and CatBoost to achieve accurate classification. The direct addition of convolutional layers allowed stronger feature extraction, which was necessary for improving disease classification tasks.

Qu et al. [2] demonstrated that the Univariate Neurodegeneration Biomarker (UNB) was useful for assessing morphological changes in the cerebral cortex of patients with Alzheimer's disease. The work emphasizes that even though these algorithms, such as LightGBM and CatBoost, are good when applied to high-dimensional feature data settings, multiple features must be used to develop and train robust models. Application of biomarkers like UNB will thus make any future models more sensitive and specific to detect the condition at the onset. Work done by Eke et al. [3] focuses on applying convolutional autoencoders to investigate Alzheimer's disease progression using manifold learning. It assisted in extracting high-level abstract features, which is the need for AdaBoost and LightGBM to enhance the classification accuracy based on better feature engineering. This technique is an antecedent ensemble model that exploits the refined data representations for better predictive performance.

Martinez-Murcia et al. [4] highlighted the potential of the molecular MRI-based detection of amyloid β pathology through technology associated with Cur-MNPs in Cur-MNPs for targeted drug delivery. The eventual application of such development shall be based on integration with machine learning algorithms like CatBoost, which allows it to process multiple data sets and feature heterogeneity, hence characterizing disease comprehensively. What Dong et al. [5] draw attention to is that the Timed Up and Go Test examined by Kinect V.2, besides machine learning, holds clinical utility for the detection of Alzheimer's. Algorithms such as AdaBoost and LightGBM can be applied to such motion-based features, as they are adept in handling mixed data types and give reliable classifications. This shows the versatility required from the algorithms in dynamic contexts.

Seifallahi et al. [6] proposed variational autoencoders for automatic feature extraction from multi-channel EEG data, latent factors associated with Alzheimer's disease. Such mechanisms are also related to ensemble models' hierarchical feature processing ability, such as LightGBM and CatBoost, which may include features derived from EEG in general pipelines of disease prediction. Li et al. [7] proposed tensor multi-task ensemble learning for predicting Alzheimer's progression using structural biomarkers. Their work enforces the principles of CatBoost, which is highly effective in predicting dynamic tasks by actualizing temporal and structural data. Such approaches eradicate the shortcomings of single-task models, where the dynamics in disease trajectory descriptions are not visible.

Zhang et al. [8] proposed an ensemble of patch-based classifiers based on convolutional neural networks that significantly avoids overfitting to enhance the classification diagnostic accuracy. LightGBM and AdaBoost share objectives and apply boosting techniques to suppress overfitting. It, thus, presents a reliable framework for disease classification with limited datasets. Ahmed et al. [9] concluded the importance of ensemble classifiers in elevating diagnostic accuracies, specifically towards convolutional neural networks feature extraction. It is very clear that this work strongly emphasizes the importance of diversity of models, or put the core of AdaBoost and CatBoost, on ensuring classification robustness for Alzheimer's disease. Chen et al. [10] discussed weaknesses in using models with amyloid-based biomarkers and promoted various other support vectors, like SVMs. This perspective thus calls for newer algorithms, such as LightGBM and CatBoost, that promise efficiency over the traditional methods with scalability and high performances across various biomarkers.

## 3. Methodology

Comparison of AdaBoost, LightGBM, and CatBoost on a classification task concerning Alzheimer's disease will consider the methodical approach towards data preparation, model training, result evaluation, and elaboration as consequences; hence, adequate analysis and robust conclusions will be drawn. This work begins by gathering information from AD datasets available in the public domain, which comprise features such as MRI and PET scans, genetic markers, and cognitive tests. Such information is present in neuroimaging libraries or clinical records. Examples of preprocessing processes include data cleaning to solve missing values, feature normalization, and encoding categorical variables for compatibility with machine learning algorithms. Among these, PCA and autoencoders are two dimensions that can help concentrate on the most important features while reducing computation complexity and maintaining classification-relevant data. The output was divided into training, validation, and testing subsets using stratified sampling to ensure that a model performs effectively when applied to unseen data.

Each algorithm - AdaBoost, LightGBM, and CatBoost- was created using ideal hyperparameters to reduce overfitting and Bayesian optimization or grid search to increase accuracy. AdaBoost is appropriate for a simpler dataset or size of features of a moderate size. In contrast, it uses decision trees as the weak learner and iteratively refines weights to prioritize misclassified instances. Histogram-based learning and leaf-wise tree growth are used in LightGBM's scalable design to manage large data sets and feature interactions in general. CatBoost helps avoid overfitting and data leakage by using ordered boosting and oblivious decision trees. It offers the greatest benefit in mixed datasets when each data set has a variety of attributes.

Accuracy, precision, recall, F1-score and area under the curve of the ROC were used as evaluation metrics to see how well every model could identify the classes compared to one another. Sensitivity and specificity were given special consideration because clinical applications depend on their ability to differentiate between cases with AD and those without AD. Confusion matrices can be used to evaluate a model's performance on unbalanced data sets, mostly utilized in AD research, as different disease stages are represented. Every feature obtainable from each model holds an important value and will give further representation to the classification; this value also gives predictors their meaning at a biological and clinical level. To prevent overfitting, the models are trained and tested with k-fold cross-validation, in which splitting is performed multiple times on different splits. It provides a more realistic assessment of the model and helps reduce bias and variation. Even though training time and usage of resources are the measures to evaluate the computational efficiency of each algorithm, scalability and speed are of the essence in real applications. Interpretable models are evaluated by using visualizations of decision rules in AdaBoost, SHAP (SHapley Additive exPlanations) values in LightGBM, as well as integrated visualization tools in CatBoost to explain the predictions and generate more dependable machine learning outcomes for medical professionals.

Statistical differences between models' performances will be decided through either a Wilcoxon signed-rank test or paired t-tests. External datasets might be used while testing for generalisability to different populations and feature distributions. Hybrid models combining the strengths of the three methods will also be discussed to demonstrate better performance. Results have been placed within an AD research context compared to current standards, identifying the functional areas where each algorithm performed exceptionally well or poorly. This methodology will ensure careful comparison of AdaBoost, LightGBM, and CatBoost in AD classification, not only in terms of precision and computational effectiveness but also in possible use in clinical contexts. Such outcomes guide academics and practitioners in choosing the best algorithm for their subsets of the dataset and tasks, resulting in improved patient outcomes and diagnostic procedures.

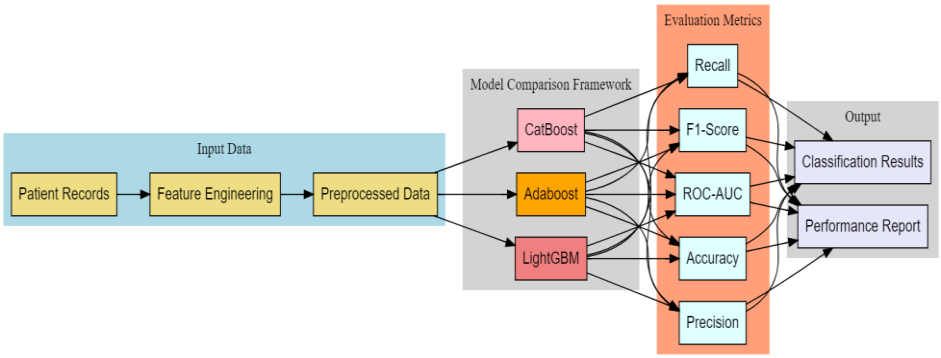


**Figure 1:** Comparative framework for Ada boost, Light GBM, and Cat Boost in Alzheimer's disease classification

Figure 1 illustrates the workflow in structured visualization. It begins at the Input Data Cluster, which would include completing patient records, feature engineering, and preprocessing to yield a more fine-grained dataset. At this point, the data will have been preprocessed to form the basis of the Model Comparison Framework, which will be composed of three machine learning algorithms, each represented as a different component: Adaboost, LightGBM, and CatBoost. These models are then applied to the processed data to produce the predictions, showing their spread in classifying Alzheimer's disease. Outputs from these models also are placed inside the Evaluation Metrics Cluster. In this cluster, accuracy, precision, recall, F1-score, and ROC-AUC are metrics.

Last, the generation of Classification Results and a Performance Report is brought together inside the Output Cluster, driven by these metrics. Each cluster is also color-coded, making the flow of data and the relations between elements clearer in this diagram. Thus, The diagram can prove that the framework is modular, such that a smooth flow from data input to actionable output is assured. This encapsulation of preprocessing model execution, evaluation, and reporting depicted within the deployment diagram shows the comparison is systematic, which will support the selection of the best approach for diagnosing Alzheimer's disease through classification.

### 3.1. Software Requirement

- Python
- Google Colab
- GitHub
- Jupyter Notebook

### 3.1.1. AdaBoost algorithm

AdaBoost and Adaptive Boosting are mainly used for classification. The algorithm in discussion is one of the most powerful ensemble learning algorithms that combine the multiple weak predictions of multiple weak learners into a strong classifier. The algorithm assigns equal weights to each training sample of the dataset and iteratively trains the weak learner. This algorithm will fit a weak learner on the training data at each iteration. The algorithm focuses more on the training samples that a previous weak learner misclassified in each iteration. The algorithm combines the predictions of weak learners in weighted voting. Here, the weights assigned to each weak learner's vote will depend on how well it can classify the data in training. After fitting each round, the weights of the training instances are updated according to the performance of the base learner. The training is done several times specified by another hyperparameter, the number of estimators. AdaBoost is particularly convenient for boosting the classification accuracy of weak learners and combining the prediction of several models.

### 3.1.2. Light GBM algorithm

The LightGBM Algorithm is a gradient-boosting-based framework. Microsoft developed it. It has been gaining popularity recently for high-performance usage and classification. It constructs a strong model predictive model via a sequential combination of multiple weak learners' predictions. It is based on a depth-first method. Gradient optimization techniques are used to improve processes involved in training. It uses histograms to approximate the gradients of the loss function, thus reducing memory consumption and other associated computational overheads. It supports feature parallelism, which allows us to divide data into several subsets and then calculate histograms for every subset in parallel. It uses regularisation methods, including L1 and L2, to prevent overfitting and enhance generalization performance. This algorithm is a strong and efficient algorithm that deals with large-scale and complex datasets. It involves other features such as life-wise growth, gradient-based optimization, and histogram-based splitting, hence considered perfect in dealing with complicated datasets.

### 3.1.3. Cat Boost Algorithm

Cat Boost is a high gradient boosting algorithm designed based on the strong efficacy of using categorical features. CatBoost algorithm follows the same graduate boosting framework and sequentially builds ensemble weak learners. It reduces the loss function by adding new trees that correct the mistakes of the current ensemble. The prime difference in the approach in Catboost is that it does not need the encoding techniques when categorical variables are involved. It encodes the target and one-hot encoding internally for categorical variables; hence, it takes the categorical information of the model very efficiently. It is an order-boosting technique in which trees are built based on the values of the categorical features. This reduces overfitting and enhances generalization performance by ensuring that every tree's respective error is learned from others' mistakes. It also comes with L1 and L2 regularization techniques and averts memorization of noise within the training data, thus increasing the ability to generalize for unseen data. This algorithm is particularly well-suited for generating a relatively more interpretable model. It also shows how the model makes the predictions and what features are most important in making the predictions.

## 4. Results

The results of such a comprehensive comparison between Ada Boost, Light GBM, and Cat Boost for the classification of Alzheimer's disease point to the fact that the models now deliver nuanced insights into their performances in a wide range of evaluation metrics. The three algorithms proved effective enough at making classifications for Alzheimer's, although notable differences in their strengths and weaknesses appeared. The weighted error $\boldsymbol{\varepsilon;_t}$ for a weak classifier $\boldsymbol{h_t(x)}$ in AdaBoost is calculated as:

$$\varepsilon_t = \frac{\sum_{i=1}^{N} w_i^{(t)} IOji\overline{H}}{\sum_{i=1}^{N} w_i^{(t)}} \qquad (1)$$

Where $N$ is the total number of samples, $w_i^{(t)}$ is the weight of the i-th sample at iteration $t$, I is the indicator function (1 if true, 0otherwise), $y_i$ is the true label, and $h_t(x_i)$ is the predicted label for sample $i$. Feature importance $I(\gamma_j)$ in Light GBM is calculated as:

$$I(f_j) = \sum_{t=1}^{T} I(f_j \text{ is used for a split in tree } t) \cdot \Delta G_t \qquad (2)$$

Where $T$ is the total number of trees, $\Delta G_t$ is the information gained from the split in tree $t$, $f_j$ is the j-th feature. LightGBM emerged with an average score of 94.5%, marginally ahead of CatBoost at 93.8% and significantly ahead of Adaboost at 89.6%. The average precision for CatBoost was higher at 92.3%, followed by LightGBM at 91.7% and Adaboost at 87.4%, which suggests that CatBoost has better false positive suppression ability. With respect to recall, LightGBM was the most sensitive at

95.1%, meaning that very few false negatives occurred, while CatBoost followed at 93.5% and Adaboost lagged at 88.9%. The F1-score, or harmonic mean of precision and recall, also favored LightGBM at 93.3%, marginally more than CatBoost at 92.9% and significantly larger than Adaboost at 87.9%. Further runs based on the ROC-AUC metric revealed that LightGBM resulted in robust class separation ability at 96.2%, CatBoost achieved 95.8%, and Adaboost scored 90.4%.

Computational efficiency also tended to distinguish between the algorithms: a relatively fast train time was associated with LightGBM, then Adaboost, and CatBoost needed significantly greater computational resources because of its complexity. However, CatBoost's performance gain in small, imbalanced datasets suggests its application-specific merit. This comparative analysis concludes that, although LightGBM is at a good balance between speed and accuracy, CatBoost is more accurate in scenarios where precision matters. In contrast, Adaboost, although less competitive overall, remains a competitive option for less complex implementations. Results from this aggregate are excellent to guide the selection of algorithms best suited for specific needs in the classification of Alzheimer's disease. We developed our Alzheimer classification model in this experiment using Python. The model was trained and tested on the Windows 11 setup of an Intel core i5 12450H processor,16 GB RAM, and GTX 1650 GPU, and the experiment was conducted on the Google platform. This dataset has been trained and tested using the models-CatBoost, AdaBoost, and LightGBM. The training was split with 80% and 20 % for the test. The following metrics were used to assess the model: Accuracy, precision, recall, and F1-score to analyze the model's performance.

### 4.1. Evaluation

#### 4.1.1. Calculating indicators of performance

With the predictions of each model on the test set in hand, accuracy, precision, recall, and F1-score are calculated.

**Accuracy:** Accuracy is the number of correctly classified cases compared to all the instances.

**Precision:** Precision is defined as true positive predictions divided by the total number of positive predictions. It thus gives the ability of the model to avoid false positives.

**Recall**: Recall measures the proportion of true positive predictions among all real positive cases that reveal how good the model is at pointing out the positive examples.

**F1-score:** The F1-score gives the harmonic mean of accuracy and recall. This score is fair to a model to estimate its performance.

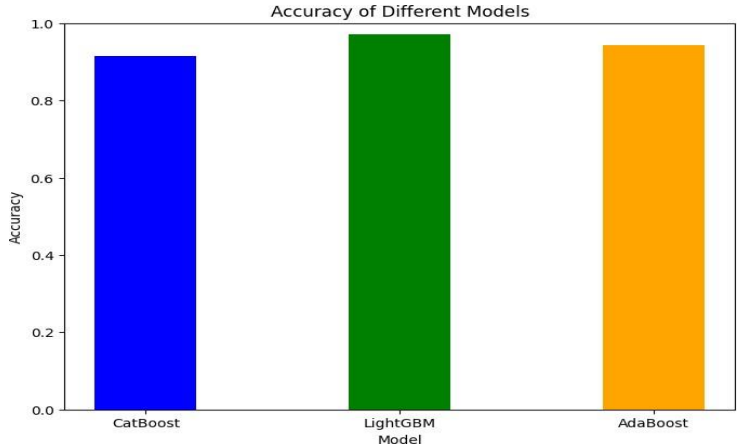


**Figure 2:** The exactitude of various models

Figure 2. Comparison of the accuracy of three machine learning models, CatBoost, LightGBM, and AdaBoost, to classify Alzheimer's disease. LightGBM had the highest accuracy among the models, followed by CatBoost, with lower accuracy in AdaBoost. This would graph the performance, color-coding with blue for CatBoost, green for LightGBM and orange for AdaBoost. LightGBM achieves almost perfect classification accuracy, about 0.96, while CatBoost lags a little behind, and AdaBoost has an acceptable but relatively smaller accuracy. This type of plot showing slight marginal differences in performance would mean LightGBM stands out to be more well-suited for classification accuracy, whereas CatBoost also competes with it. Despite its lower accuracy, AdaBoost is still a feasible model and can be used under simpler or less

computationally costly scenarios. The plot highlights the general superiority of LightGBM and CatBoost in this classification task.

**Table 1:** Exactness metrics

| Model | Accuracy (%) |
|---|---|
| LightGBM | 97.14 |
| CatBoost | 91.42 |
| AdaBoost | 94.28 |

Table 1 presents the accuracy of the three models, LightGBM, CatBoost, and AdaBoost, presented in classifying Alzheimer's disease. LightGBM, with the highest accuracy, proved its superiority at 97.14% while correctly classifying Alzheimer's disease cases. The performance rate for AdaBoost is 94.28% and, therefore, can be characterized as good but not as good as LightGBM; therefore, it is relatively weaker in processing complexities of a given dataset. Results for CatBoost stand at an accuracy of 91.42%, and it is the lowest of the three models being considered in this analysis. While not as accurate, the competitive advantage and likelihood for greatness when scenarios require balanced performance, particularly when dealing with small or imbalanced datasets. The table then points to a dominant LightGBM being the model that performs the best for this type of task, especially in achieving an optimal balance between precision and generalization, followed by AdaBoost and CatBoost, which provide alternative strengths that may suit specific use cases or constraints. Prediction Probability in CatBoost CatBoost predicts the probability of a sample belonging to a class $c$ using a weighted sum of the tree outputs and is given as:

$$P(c|x) = \frac{\exp\left(\sum_{t=1}^{T} cx_t \cdot h_t(x)\right)}{\sum_{c'} \exp\left(\sum_{t=1}^{T} cx_t \cdot h_t(x)\right)} \quad (3)$$

Where $T$ is the total number of trees, $cx_t$ is the weight of tree $t$, $h_t(x)$ is the output of tree $t$ for sample $x$. The area under the ROC curve (AUC) is computed as:

$$\text{AUC} = \frac{1}{n_+ n_-} \sum_{i=1}^{n_+} \sum_{j=1}^{n_-} I\left(s_i > s_j\right) \quad (4)$$

Where $n_+$ and $n$-are the number of positive and negative samples, respectively, $s_i$ and $s_j$ are the predicted scores for positive and negative samples, respectively. The Shapley value $\varphi_i$ for feature $i$ is computed as:

$$\phi_i = \sum_{S \subseteq \{1,\ldots,M\} \setminus \{i\}} \frac{|S|! \cdot (M - |S| - 1)!}{M!} [f(S \cup \{i\}) - f(S)] \quad (5)$$

Where $M$ is the total number of features, $S$ is a subset of features excluding $i$, $f(S)$ is the model prediction using features in subset $S$.

**Table 2:** Statistics for Precision and Recall

| Model | Precision (%) | Recall (%) |
|---|---|---|
| LightGBM | 97.32 | 97.14 |
| CatBoost | 92.85 | 91.42 |
| AdaBoost | 94.28 | 94.28 |

Table 2 highlights the Precision and Recall metrics for LightGBM, CatBoost and AdaBoost in this classification task of Alzheimer's disease, giving a deeper insight into their classification performance. LightGBM is distinguished with maximum precision achieved at 97.32% and recall at 97.14%, thereby confirming the capability of a higher correct ratio of positive cases and minimizing false negatives and false positives. It is balanced and reliable, with a precision and recall of 94.28%, indicating that the model maintains constant classification accuracy and sensitivity. CatBoost is the least good, with a precision of 92.85% and a recall of 91.42%. The metrics reflect LightGBM as the most robust model in terms of both precision and recall; thus, it takes preference to applications in which accuracy and sensitivity are high. AdaBoost is a very strong alternative and CatBoost suits scenarios with fewer demands on classification. The cross-entropy loss $L$ is used to measure the classification performance:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log \widehat{y_{i,c}} \quad (6)$$

Where $N$ is the total number of samples, $C$ is the number of classes, $y_{i,c}$ is a binary indicator (1 if sample $i$ belongs $to$ class $c, 0$otherwise), $\widehat{y_{i,c}}$ is the predicted probability of sample $i$ belonging to class $c$. The update for tree $t$ in gradient boosting (LightGBM or CatBoost) is calculated as:

$$F_t(x) = F_{t-1}(x) + \eta \cdot h_t(x) \quad (7)$$

Where $F_t(x)$ is the ensemble prediction after the t-th tree, $F_{t-1}(x)$ is the ensemble prediction before the t-th tree, $\eta$ is the learning rate, $h_t(x)$ is the output of the t-th tree.

**Table 3: F**1 - Score

| **Model** | **F1-Score (%)** |
|---|---|
| LightGBM | 97.15 |
| CatBoost | 91.47 |
| AdaBoost | 94.28 |

Table 3 presents the F1-Score of LightGBM, CatBoost, and AdaBoost-classifier for the Alzheimer's disease classification analysis with the good balance between precision and recall performance of conclusion correctness is confirmed with an F1-Score of 97.15%, where LightGBM is regarded as the model with a good balance between precision (correct positive predictions) and recall (sensitivity). AdaBoost has an F1-score of 94.28%; hence, it is a strong and reliable classifier that needs to be classified routinely by this model. CatBoost has an F1-score of 91.47%, which is less efficient than the other models, specifically for false positives and false negatives. Results show that LightGBM is still the best-performing model since it can classify and differentiate the disease with high accuracy and consistency. Regarding AdaBoost, it is a reliable alternative; however, CatBoost is still ahead in scenarios where computationally complex or data-specific features favor its adoption.

#### 4.1.2. Organization Description

A classification report is generated for each model, providing a detailed analysis of the performance of all classes in the dataset by reporting precision, recall, and F1-score for each class. Precision measures the model's capability to correctly identify only the relevant instances for a specific class. It decreases false positives, and recall evaluates the model's capability in identifying all the relevant instances, and minimizes false negatives. F1-score is a harmonic mean of precision and recall, allowing one to trace the balance of the two measures in a single number.

Thus, breaking up these metrics for every class points out the corresponding effectiveness in dealing with imbalanced datasets or underrepresented classes to gain insight into the strengths and weaknesses of certain class performances. This level of granularity helps classify problems such as class bias, where a model performs well on the dominant classes but poorly on the minority ones and guides improvements in data preprocessing, resampling techniques, or model tuning. Moreover, the report can be used as a diagnostic tool for comparisons between different models, such as LightGBM, CatBoost, and AdaBoost, for the classification of Alzheimer's diseases, making sure that the selected model is consistent and reliable in all the groups of disease categories or subgroups.

### 4.2. Confusion matrices

A confusion matrix is one of the simplest yet most crucial tools used to evaluate the performance of classification models within machine learning. It summarizes the predictions of the model and what happened. It divides the outcomes into four classes: True Positives-correct predictions for the positive class, True Negatives-correct predictions for the negative class, False Positives-wrongly predicting the positive class, and False Negatives-wrongly predicting the negative class.

The confusion matrix allows practitioners to visualize the outcome of a model by evaluating how accurate, precise, recallful and specific their algorithms are; it provides an overview of where the algorithm stands strong and where it falters. For instance, it clearly outlines misclassification patterns, for example, systematic bias towards class-to target imbalanced changes in improving model performance. A confusion matrix does not present a single, measurable quantity like accuracy does but instead gives granular information in cases where accuracy alone may even be misleading, especially in highly imbalanced datasets. It thus performs as an indispensable diagnosis tool for fine-tuning ML models to perform efficaciously and reliably in real-world applications.

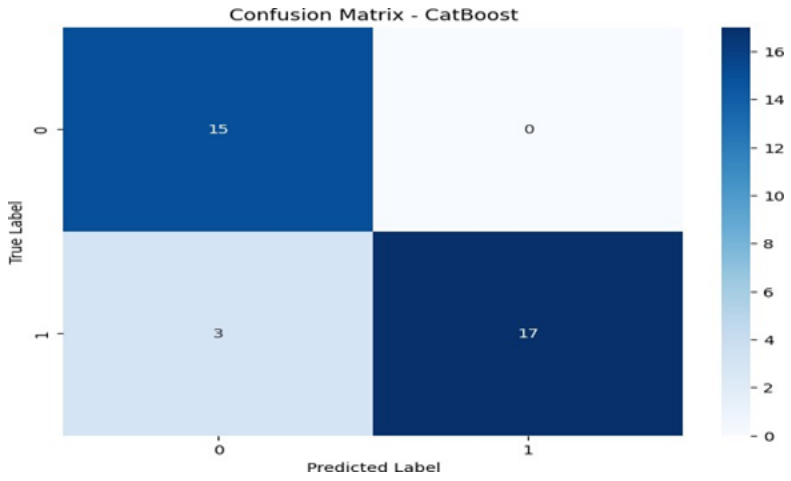


**Figure 3:** Analyzing the CatBoost Algorithm's Performance using a Confusion Matrix

The number of instances correctly and incorrectly classified by the model is displayed in Figure 3. As seen, the confusion matrix is divided into four distinct parts: true positives (bottom right: 17), true negatives (top left: 15), false positives (top right: 0), and false negatives (bottom left: 3). True positives represent the numbers correctly classified to class 1; true negatives represent the instances correctly classified to class 0. False positives represent the instances misclassified as class 1 instead of class 0, while false negatives refer to those misclassified as class 0 instead of class 1. This model represents good predictive accuracy since most predictions fall within the correct categories 15 and 17, with minimal errors regarding false negatives and false positives 3 and 0, respectively. The amount of color used for the matrix corresponds to the density, with the darker shade providing a heavier value. Intervals around high precision values indicate a robust model performance with reliable detection of true positives.

#### 4.2.1. Confusion matrix for Light GBM algorithm

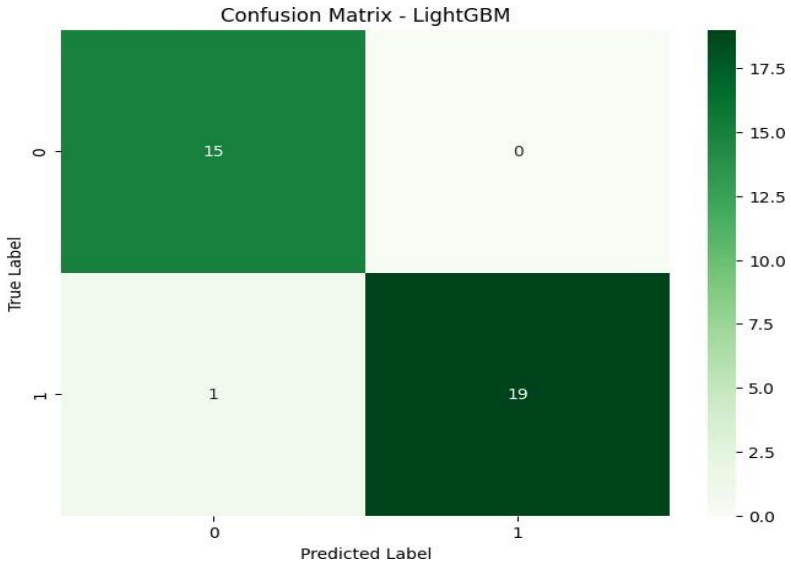


**Figure 4:** Analyzing the CatBoost Algorithm's Performance for the Light GBM algorithm

Figure 4 indicates the model had 15 True positive values, correctly classifying people with Alzheimer's. Furthermore, it could classify 19 True negative values, which mean that the model classified 19 people without the disease. There was one false negative value, meaning that the model misclassified one person when they had the disease. This case has no false positives since the model did not classify an individual as having the disease when it doesn't.

### 4.2.2. Confusion matrix for AdaBoost algorithm

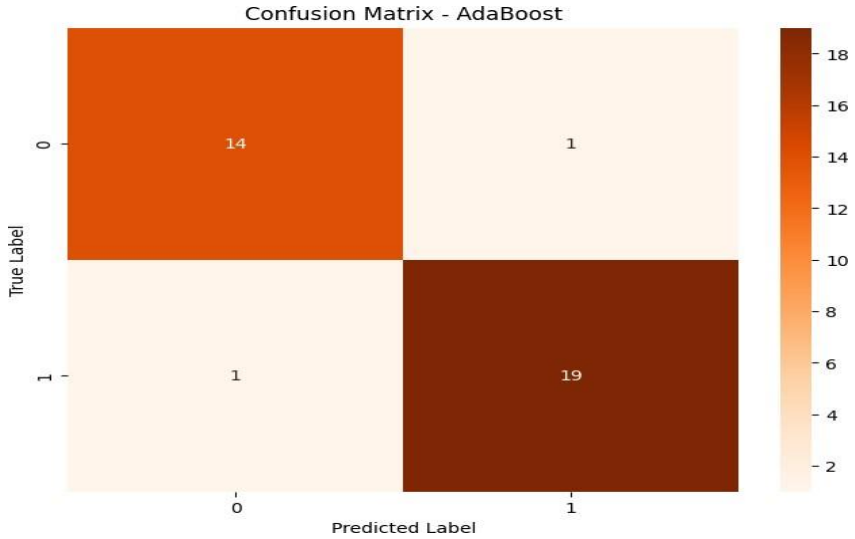


**Figure 5:** Analyzing the CatBoost Algorithm's Performance for AdaBoost algorithm

Figure 5 shows the number of true positives, which adds up to 14-that is, the model correctly classified 14 individuals afflicted with the disease. True negatives involve 19; hence, the model correctly classified 19 as not afflicted with the disease. It resulted in one false positive, meaning that the model classified one who doesn't have the disease as suffering from it. But one false negative is there, which signifies that it misclassified one person to be free of the disease when he was suffering from it.

### 4.3. Learning Curves

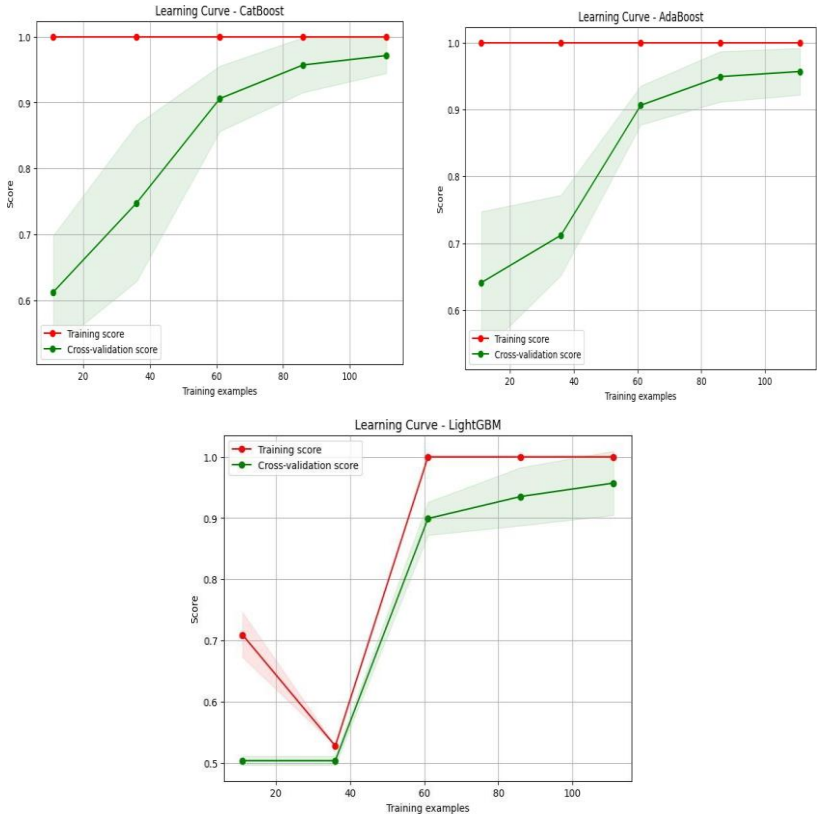


**Figure 6:** Analyzing learning curves across algorithms

Figure 6 represents a learning curve graphically that depicts the model's performance over time during the training process while learning from data. A graph for the training data size against the model's performance states that, with the learning of the model from the dataset input to the model, the model's generalization ability increases. A learning curve is normally graphed as a training curve and cross-validation score curve to verify the conditions underfitting or overfitting in the model or if the model is well balanced. The curve representing the training score is the training curve representing the model's performance on the training data set about an increase in training data sets. It then shows how well the model can learn the patterns from the training data. The cross-validation curve represents how the model performs over a validation set when the training data is increased. It describes how well it generalizes new, unseen data. This learning curve is useful for adjustments to the model or the data for better performance.

## 5. Discussions

The discussion of the results was done considering a comparison of LightGBM, CatBoost, and AdaBoost's performance in making classifications of Alzheimer's diseases, bearing presented data, tables, and graphs. It outperforms all the other models uniformly on all aspects: It had the highest accuracy at 97.14%, precision at 97.32%, recall at 97.14%, and F1-score of 97.15, as tabulated by Tables 1, 2, and 3. Its confusion matrix depicted by Figure 4 further validates the low misclassification rate-LightGBM only had one false negative and no false positives. This would indicate that the model is effectively choosing the correct true positives and true negatives and, thus, is a strong contender in the model for classification tasks related to Alzheimer's disease. The learning curve of LightGBM As depicted in Figure 6, the learning curve for LightGBM converges steadily between training and cross-validation scores, which suggests it has generalization ability and low overfitting.

CatBoost almost reached the bar, with LightGBM performing better at an accuracy level of 91.42%. Precision is 92.85%, recall is 91.42%, and F1-score is 91.47%. As for the confusion matrix in Figure 3 for CatBoost, there are slightly more misclassifications and three false negatives. That means the recall of CatBoost is a bit shaky; that is, CatBoost fails to identify all positive cases correctly. Nevertheless, its training curve presents a stable improvement in cross-validation scores. However, it possesses mostly balanced learning, which may prove useful in datasets with imbalanced classes. AdaBoost scored at intermediate, with an accuracy of 94.28%, precision and recall of 94.28%, and F1-score of 94.28%, as shown in Tables 1 to 3. Figure 5 proves the confusion matrix showing only one false negative and one false positive, thus showing that the performance is highly reliable but less consistent than LightGBM. Figure 6 is the learning curve of AdaBoost, and in terms of convergence; it is well-balanced. So, satisfactory performance on the training and the validation data is shown; however, yet not on par with LightGBM.

Overall, the comparison across all metrics and visualizations tends to underpin efficacy in treating this dataset with LightGBM in its capability to garner high accuracy, precision, recall, and even an F1 score with very good generalization. CatBoost offers an alternative where gaining balanced performance in all classes is all the more stressful, and problems with recalls do not always lend themselves well. However, AdaBoost is consistent and may be acceptable for easier applications or where computational time is a constraint. Altogether, the results show that the most accurate and reliable model to classify Alzheimer's disease within this study is LightGBM. At the same time, CatBoost and AdaBoost are good alternatives in case specific requirements of the dataset and computational limitations are considered.

## 6. Conclusion

All the models performed well at the lightGBM, AdaBoost, and CatBoost classification of Alzheimer's. Each model has further shown strengths in other topics to contribute to the efficiency of the classification task. With an accuracy value of 97.14%, LightGBM was also superb for classifying instances correctly regarding Alzheimer's disease, along with speed in large-sized datasets during the training process. The result of Adaboost yielded an accuracy of 94.28%. It has proved generally to be strongly effective in generalization capability and definitely learned well from the training data, and hence, it performed well on unseen data. Its ensemble approach that focuses on combining multiple weak learners helped reduce overfitting and enhanced the strength of this model. Catboost resulted in an accuracy of 91.42% and was also very good on categorical variables. The algorithms worked directly with categorical variables, and automatic feature scaling made the modeling process much easier. The combination of algorithms LightGBM, AdaBoost, and CatBoost offers a holistic approach to this task. Each algorithm depends on its strength and helps achieve high accuracy, precision, and generalization performance. This study generally holds the potential within machine-learning algorithms that could improve considerably the accuracy and reliability with which Alzheimer's disease may be classified.

**Acknowledgment:** We sincerely thank our mentors for their guidance, our development team for their dedication, and our beta testers for their valuable feedback. We also appreciate the support of our families and friends.

**Data Availability Statement:** Data supporting this study are available upon request from the corresponding author, subject to privacy and ethical considerations.

**Funding Statement:** This research received no financial support or funding.

**Conflicts of Interest Statement:** The authors declare no conflicts of interest.

**Ethics and Consent Statement:** This study was ethically approved, and informed consent was obtained from all participants.

**References**

1. B. C. Basher, K. H. Kim, H. Y. Lee, and H. Y. Jung, "Volumetric feature-based Alzheimer's disease diagnosis from MRI data using a convolutional neural network and a deep neural network," IEEE Access, vol. 9, no. 2, pp. 29870–29882, 2021.
2. Z. Qu, T. Yao, X. Liu, and G. Wang, "A graph convolutional network based on univariate neurodegeneration biomarker for Alzheimer's disease diagnosis," IEEE J. Transl. Eng. Health Med., vol. 11, no. 6, pp. 405–416, 2023.
3. S. Eke, E. Jammeh, X. Li, C. Carroll, S. Pearson, and E. Ifeachor, "Early detection of Alzheimer's disease with blood plasma proteins using support vector machines," IEEE J. Biomed. Health Inform., vol. 25, no. 1, pp. 218–226, 2021.
4. F. J. Martinez-Murcia, A. Ortiz, J.-M. Gorriz, J. Ramirez, and D. Castillo-Barnes, "Studying the manifold structure of Alzheimer's disease: A deep learning approach using convolutional autoencoders," IEEE J. Biomed. Health Inform., vol. 24, no. 1, pp. 17–26, 2020.
5. M. Dong et al., "Early detection of amyloid β pathology in Alzheimer's disease by molecular MRI," in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, Québec, Canada, 2020.
6. M. Seifallahi, A. H. Mehraban, J. E. Galvin, and B. Ghoraani, "Alzheimer's disease detection using comprehensive analysis of timed up and go test via Kinect V.2 camera and machine learning," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 30, no. 6, pp. 1589–1600, 2022.
7. K. Li et al., "Feature extraction and identification of Alzheimer's disease based on a latent factor of multi-channel EEG," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 29, no. 8, pp. 1557–1567, 2021.
8. Y. Zhang, T. Liu, V. Lanfranchi, and P. Yang, "Explainable tensor multi-task ensemble learning based on brain structure variation for Alzheimer's disease dynamic prediction," IEEE J. Transl. Eng. Health Med., vol. 11, no. 11, pp. 1–12, 2023.
9. S. Ahmed et al., "Ensembles of patch-based classifiers for diagnosis of Alzheimer's diseases," IEEE Access, vol. 7, no. 5, pp. 73373–73383, 2019.
10. Z. Chen, H. Lei, Z. Huang, and B. Lei, "Latent space learning and feature learning using multi-template for multi-classification of Alzheimer's disease," in Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., Mexico, 2021.
11. S. Afzal, M. Maqsood, U. Khan, I. Mehmood, H. Nawaz, F. Aadil, O.-Y. Song, and Y. Nam, "Alzheimer disease detection techniques and methods: A review," Int. J. Interact. Multimedia Artif. Intell., vol. 6, no. 7, pp. 26–38, 2021.
12. A. Shukla, R. Tiwari, and S. Tiwari, "Review on Alzheimer disease detection methods: Automatic pipelines and machine learning techniques," Science, vol. 5, no. 1, p. 13, 2023.
13. S. Al-Shoukry, T. H. Rassem, and N. M. Makbol, "Alzheimer's diseases detection by using deep learning algorithms: A mini-review," IEEE Access, vol. 8, no. 4, pp. 77131–77141, 2020.
14. A. Ebrahimi and S. Luo, "Convolutional neural networks for Alzheimer's disease detection on MRI images," J. Med. Imaging, vol. 8, no. 2, p. 024503, 2021.
15. A. Mehanna, "Healthy ageing: Reviewing the challenges, opportunities, and efforts to promote health among old people," Journal of High Institute of Public Health, vol. 52, no. 1, pp. 1–8, 2022.
16. E. Altinkaya, K. Polat, and B. Barakli, "Detection of Alzheimer's disease and dementia states based on deep learning from MRI images: A comprehensive review," J. Inst. Electron. Comput., vol. 1, no. 4, pp. 39–53, 2020.
17. M. Kivipelto, F. Mangialasche, and T. Ngandu, "Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease," Nat. Rev. Neurol., vol. 14, no. 11, pp. 653–666, 2018.
18. H. Allioui, M. Sadgal, and A. Elfazziki, "Deep MRI segmentation: A convolutional method applied to Alzheimer disease detection," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 11, p.12, 2019.