

Article

Predicting the Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using an Explainable AI Approach

Gerasimos Grammenos ¹, Aristidis G. Vrahatis ¹, Panagiotis Vlamos ¹, Dean Palejev ², Themis Exarchos ^{1,*}
and for the Alzheimer's Disease Neuroimaging Initiative [†]

¹ Bioinformatics and Human Electrophysiology Laboratory, Department of Informatics, Ionian University, 7 Tsirigoti Square Corfu, 49100 Kerkira, Greece; p19gram@ionio.gr (G.G.); aris.vrahatis@ionio.gr (A.G.V.); vlamos@ionio.gr (P.V.)

² Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria; palejev@math.bas.bg

* Correspondence: exarchos@ionio.gr

[†] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf, accessed on 11 February 2024.

Abstract: Mild Cognitive Impairment (MCI) is a cognitive state frequently observed in older adults, characterized by significant alterations in memory, thinking, and reasoning abilities that extend beyond typical cognitive decline. It is worth noting that around 10–15% of individuals with MCI are projected to develop Alzheimer's disease, effectively positioning MCI as an early stage of Alzheimer's. In this study, a novel approach is presented involving the utilization of eXtreme Gradient Boosting to predict the onset of Alzheimer's disease during the MCI stage. The methodology entails utilizing data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Through the analysis of longitudinal data, spanning from the baseline visit to the 12-month follow-up, a predictive model was constructed. The proposed model calculates, over a 36-month period, the likelihood of progression from MCI to Alzheimer's disease, achieving an accuracy rate of 85%. To further enhance the precision of the model, this study implements feature selection using the Recursive Feature Elimination technique. Additionally, the Shapley method is employed to provide insights into the model's decision-making process, thereby augmenting the transparency and interpretability of the predictions.

Keywords: Alzheimer's disease; mild cognitive impairment; machine learning; gradient boosting



Citation: Grammenos, G.; Vrahatis, A.G.; Vlamos, P.; Palejev, D.; Exarchos, T.; for the Alzheimer's Disease Neuroimaging Initiative. Predicting the Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using an Explainable AI Approach. *Information* **2024**, *15*, 249. <https://doi.org/10.3390/info15050249>

Academic Editors: Arkaitz Zubiaga and Andreas Triantafyllidis

Received: 22 February 2024

Revised: 30 March 2024

Accepted: 26 April 2024

Published: 28 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mild Cognitive Impairment (MCI) is a clinical state marked by subtle yet quantifiable declines in an individual's cognitive capabilities. These declines are more pronounced than the typical age-related changes but do not reach a severity level that substantially disrupts an individual's daily life or independence. Such cognitive declines involve mild impairments in memory, cognitive skills, and cognitive functions [1]. Research has shown that MCI may serve as an indicator of Alzheimer's disease (AD), with approximately 10–15% of individuals with MCI progressing to AD annually [2]. According to recent statistics, an estimated 6.7 million Americans aged 65 and older are currently living with Alzheimer's dementia, with this number projected to reach 13.8 million by 2060 unless significant medical breakthroughs for prevention, slowing, or curing Alzheimer's disease are achieved [3]. Despite years of clinical research, there is still no cure for AD. This makes knowledge of MCI presence a crucial factor for predicting an individual's risk of developing AD and, in turn, enables the timely administration of effective treatment or intervention to slow down its progression [4]. To address this issue, we need to distinguish MCI patients

who convert to AD (MCIc) from MCI patients who remain stable (MCInc) within a specific period. To accurately distinguish those two types of MCI (MCIc vs. MCInc), multiple studies have been conducted from multiple perspectives, such as genetics, medical imaging, and pathology [5]. At present, varying perspectives exist on the efficacy of biomarkers in faithfully depicting the progression of preclinical disease over time [6,7].

Traditionally, addressing such problems, such as distinguishing MCIc (MCI due to AD) from MCInc patients, has been performed manually by clinicians using specific guidelines. Recent advances in machine learning can help experts automate this process, resulting in lower error rates and speeding up the diagnostic process. In the field of AD research, the development of such algorithms tailored for automatic diagnosis and the prediction of an individual's future clinical status based on biomarkers has become increasingly prevalent [8]. Studies have shown that the combination of machine learning techniques and the analysis of large amounts of heterogeneous data collected with non-invasive techniques may be the future for precise AD diagnosis. These datasets are sourced from a wide range of inputs, including medical imaging, biosensors for biofluid marker detection, and movement sensors, among others [9].

Furthermore, machine learning can assist in identifying cost-effective biomarkers for predicting the progression of Alzheimer's disease in an efficient way. This is especially valuable given that current diagnostic methods depend on costly techniques [10]. Several novel biomarkers have shown promise in detecting Alzheimer's disease (AD). These include neurofilament light (NFL) as a biomarker for neuronal injury, neurogranin, BACE1, synaptotagmin, SNAP-25, GAP-43, and synaptophysin as biomarkers related to synaptic dysfunction and/or loss, and sTREM2 and YKL-40 as biomarkers associated with neuroinflammation. Additionally, D-glutamate, a coagonist for NMDARs, has shown correlation with cognitive impairment and potential as a peripheral biomarker for detecting MCI and AD [10].

Through the analysis of such biomarker data, including A β -amyloid, neurofilament light, and BACE1, and the application of machine learning methodologies, we can create models that effectively learn patterns to distinguish between healthy individuals and those with Alzheimer's disease [10]. When it comes to A β -amyloid positivity, machine learning can help predict A β -amyloid future presence in the non-demented population using data obtained with non-invasive techniques to improve screening processes in clinical trials [11]. Other machine learning approaches include the analysis of T1-weighted MRI scans to identify the presence of sensitive biomarkers to critical brain regions related to potential AD development to help distinguish MCIc from MCInc patients, resulting in an accuracy of 66% [12]. Radiomic feature extraction from medical images can also help predict AD and MCI in combination with random forest algorithms that can reach accuracies up to 73% [13]. Additionally, there have been approaches that classify dementia into three stages, including MCI-related conditions, which have shown promising prediction results [14].

However, hybrid approaches, which incorporate longitudinal and heterogeneous data along with supervised and unsupervised methods, have proven to be more effective under certain circumstances and generalize better, achieving results with up to 85% accuracy, particularly in distinguishing between MCI progressive from non-progressive cases [15]. Such approaches also extend to deep learning, a subset of machine learning, which combines convolutional neural networks and long short-term memory algorithms using multimodal imaging and cognitive tests to classify healthy individuals from those with early MCI (EMCI), with an accuracy of 98.5% [16].

The explainability of these models is crucial for understanding how predictions are made and, in turn, for pinpointing the key factors driving the progression of AD. This is particularly important when dealing with blackbox AI systems. Explainable model methods, such as Shapley Values, offer promising insights into individual model predictions, showcasing how specific features influence outcomes and revealing inter-feature relationships [17]. Particularly, XGBoost has demonstrated promising interpretability by utilizing

Shapley Values to uncover specific biomarkers pivotal in the progression likelihood of Alzheimer's disease [18].

In this light, we propose an explainable machine learning solution, which aims to predict cognitive decline in individuals, with a focus on distinguishing instances of MCI stable cases and MCI to AD cases within a 48-month timeframe from the initial examination with a specialized medical professional. Given that nearly all cases of AD originate from MCI, our goal is to discern which instances of MCI will progress to AD in the medium term (in less than 48 months) or will have a longer timeline before doing so (exceeding the 48-month timeframe). The importance of this prediction is high, as it can inform the healthcare professional of a potential worsening in the patient's condition and can lead to earlier pharmacologic or non-pharmacologic intervention to delay the transition. Also, the explainability that is provided can help the treating physician identify the worsening cognition parameters and specifically target them. While diagnosing MCI can be challenging, its identification serves as an early indicator of potential progression to Alzheimer's disease. Consequently, the development of a predictive model for MCI progression holds immense promise, offering significant advantages in both research and practical applications. Although there is significant literature in the application of AI techniques in the ADNI repository, there are only a few that focus on developing progression models, while most focus on diagnostic models. In addition, most of the previous research works utilize individual ADNI cohorts (ADNI1 or ADNI2 or ADNI3 or ADNI GO), depending on the features required for analysis. In our case, we used all four ADNI cohorts in an attempt to have a more complete cohort and increase the significance of the obtained results.

2. Materials and Methods

Data used in the preparation of this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of Mild Cognitive Impairment (MCI) and early Alzheimer's disease (AD).

2.1. Data and Preprocessing

We used the ADNIMERGE dataset to train and evaluate our proposed model. ADNIMERGE is an R package that combines trivial key predictors from all four phases of the ADNI dataset project coded for applying data analysis and machine learning techniques. The ADNI (Alzheimer's Disease Neuroimaging Initiative) dataset is a comprehensive and longitudinally collected collection of clinical, imaging, genetic, and cognitive data from participants with Alzheimer's disease, Mild Cognitive Impairment, and healthy controls. The principal goal of the Alzheimer's Disease Neuroimaging Initiative (ADNI) is to explore and identify biomarkers that can be employed in clinical trials aimed at advancing the development of treatments for Alzheimer's Disease [19,20].

The initial dataset comprises a cohort of 2378 patients, encompassing participants across all ADNI phases. The dataset incorporates a comprehensive set of 64 features, encompassing a diverse range of information from various domains. These features encompass both biological data, such as APOE4 status, and imaging data, including FDG-PET scans. Additionally, screening tests such as the Mini-Mental State Examination (MMSE) and Everyday Cognition Scale are incorporated into the dataset.

A brief description of subjects' demographics used in this study is shown in Figure 1. This study predominantly involved male participants, averaging 73 years in age. Furthermore, most of the subjects identified as Caucasian and were married.

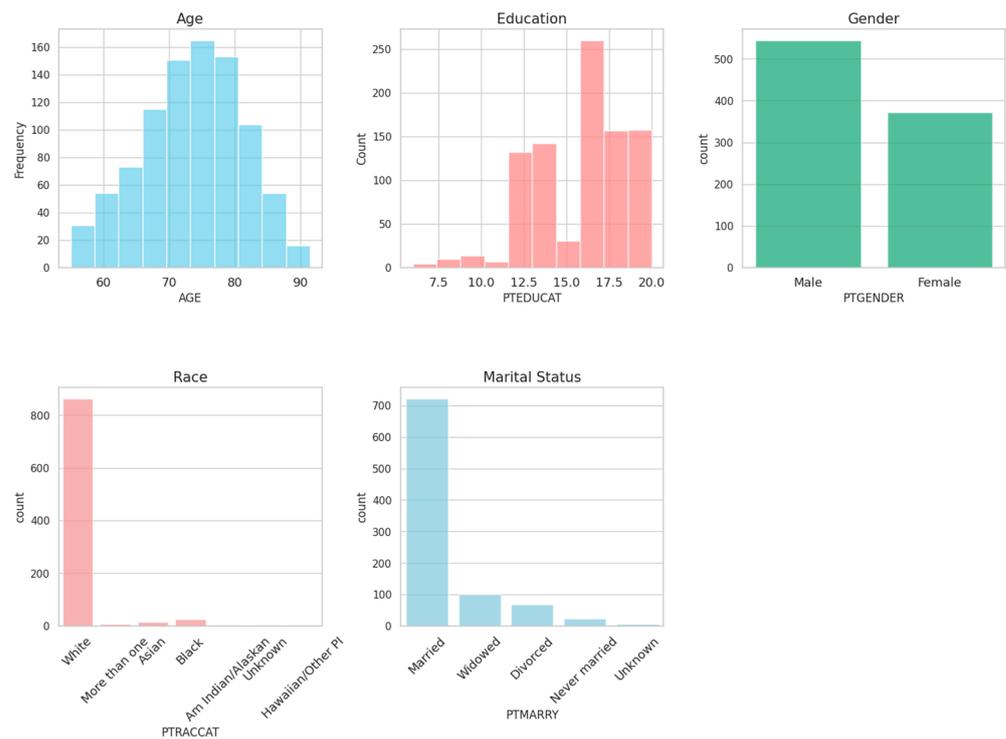


Figure 1. Demographics summary.

Given the longitudinal nature of the dataset, each patient is represented by multiple entries corresponding to distinct time points, aligning with the data collection protocol employed in the respective ADNI phase. Most diagnoses and visits are concentrated in the early months subsequent to the initial visit, with a discernible decline over time. One plausible interpretation of this pattern is that participants display a heightened level of compliance and engagement at the outset of the clinical trial, which diminishes as the study progresses. Notably, despite the diminishing frequency of visits over time, there is a consistent adherence to annual intervals, occurring in multiples of 12 months (e.g., 24 months, 36 months, etc.).

Since there is a high consistency for annual visits, we selected the baseline and follow-up visit records to predict the progression to AD from MCI within a 24- to 48-month timeframe. The refined subset then comprised 918 subjects, among whom 636 exhibited stability throughout the 48-month observation period, while 282 transitioned to Alzheimer’s disease (AD) within the predictive window. A comprehensive overview of disease progression within our curated dataset is depicted in the Sankey Diagram in Figure 2.

Sankey Diagram for Disease Progression



Figure 2. Sankey Diagram featuring patient’s clinical state from the baseline until the 48th-month visit.

Missing values were also present in the dataset. Figure 3 illustrates an overview of missing values for baseline visits.

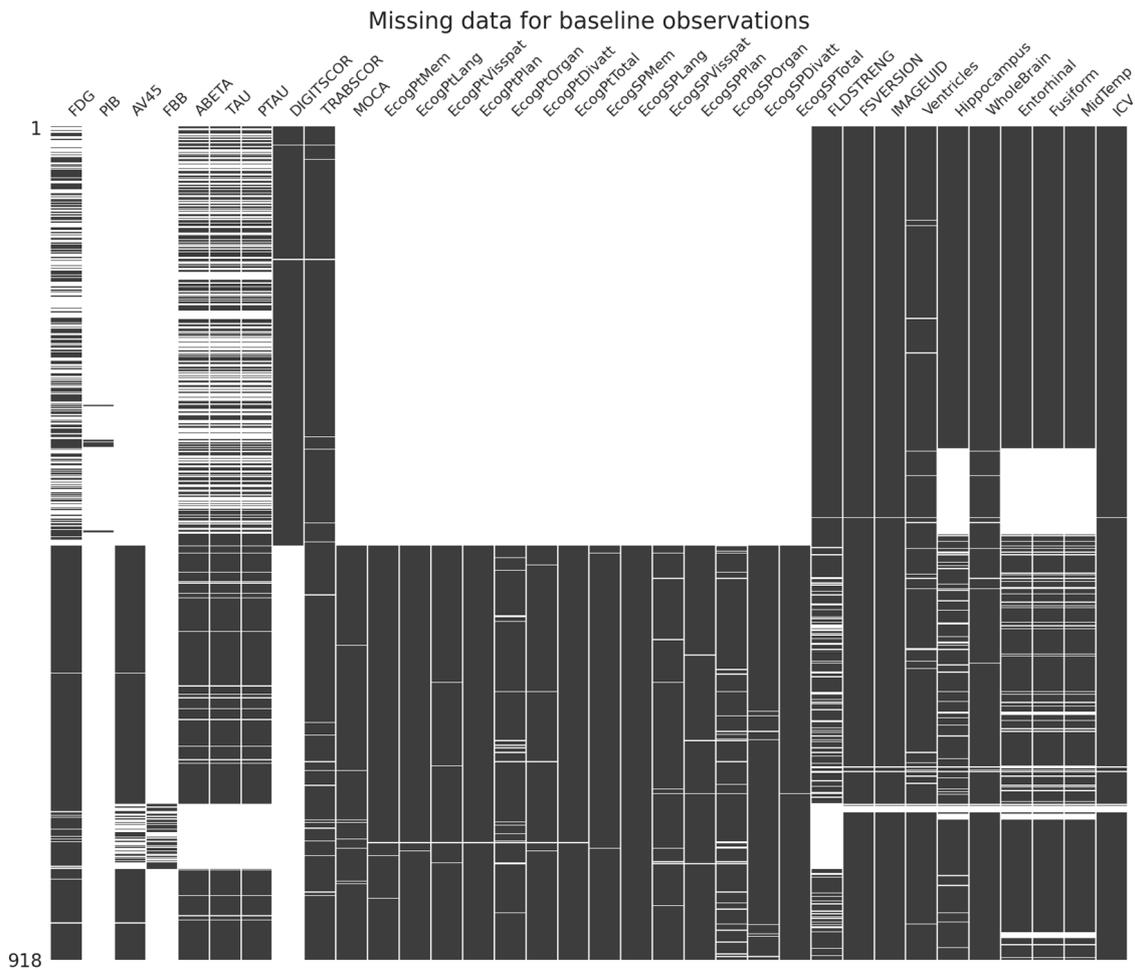


Figure 3. Missing values at baseline visit.

Considering the strict clinical guidelines defined by ADNI, we consider missing data entries as “Missing Not at Random” (MNAR). This designation signifies that data absence is not a random occurrence but rather stems from reasons related to the observed data, resulting in unanticipated lapses in data collection. Specifically, certain features were not obtained at baseline and follow-up timepoints, with the anticipation that these would be recorded at subsequent visits. One exception to this assumption is scales related to “Every Day Cognition Tests” and the “Montreal Cognitive Assessment (MOCA)”, which were not gathered during the ADNI1 phase and, consequently, were treated as missing during the training process [21]. For the remaining missing data, the k-nearest neighbors (KNN) imputation method with $k = 5$ was employed, considering the target class [22]. Notably, ADNI1 was imputed separately from the rest of the dataset to mitigate potential imputation bias, given the absence of data for the scales mentioned above during this phase. To enhance the representation of categorical variables, one-hot encoding was applied.

We modeled the progression from MCI to AD as a binary feature, assigning a value of “0” denoting the likelihood of the patient remaining stable in the MCI state over the subsequent 36 months and a value of “1” indicating the likelihood of transitioning to AD within the same timeframe. (See Figure 4 for a visual depiction of the prediction window). This determination was made by analyzing the diagnosis records of each patient within the cohort utilized in this study to represent the longitudinal progression of the disease in one feature. It is worth noting that in some instances, patient visits were recorded under different ADNI protocols (e.g., some visits occurred during the transition from ADNI 1 to ADNI 2).

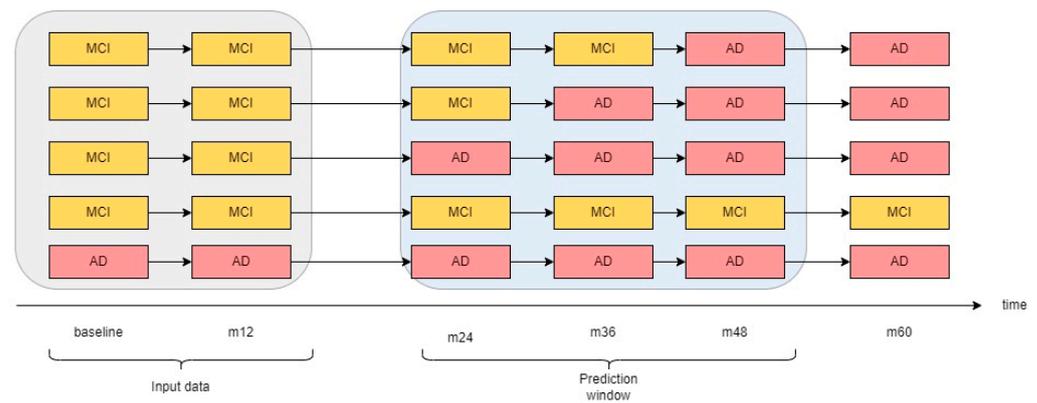


Figure 4. Prediction method using baseline and first annual visit data.

2.1.1. Model Selection

We employed a comprehensive approach to model selection by comparing the performance of various machine learning algorithms: XGBoost, CatBoost, LightGBM, Logistic Regression, Naïve Bayes, and Decision Tree. The goal was to identify the model that best captures the underlying patterns in the data and demonstrates robust generalization capabilities. To achieve this, we conducted a thorough evaluation using a 5-fold cross-validation strategy and evaluated each model's performance.

2.1.2. XGBoost

XGBoost stands for eXtreme Gradient Boosting, and it is a popular machine learning algorithm that falls under the category of ensemble classifiers. The fundamental principle behind the algorithm involves the iterative incorporation of Decision Trees by learning the negative gradient of the loss function, computed with respect to the disparity between the predicted value from the preceding tree and the actual value. It systematically performs feature splitting to grow the ensemble tree. The algorithm leverages the second-order derivative of the loss function to learn the negative gradient. It also has regularization capabilities by using a penalty term to prevent overfitting and improve generalization. Additionally, the algorithm addresses class imbalance by computing the inverse ratio of positive to negative classes, employing it as a weighting operator for each class [23].

2.1.3. CatBoost

CatBoost, introduced by Prokhorenkova et al. in 2018 [24], is a gradient boosting algorithm known for its proficiency in handling categorical features, with minimal information loss. CatBoost introduces two innovative methods, contributing to its state-of-the-art performance. Firstly, it addresses the challenge of categorical features by employing a unique encoding approach, which involves dividing and substituting categories with several numeric features, thus minimizing information loss. This process involves random sorting of input samples and calculating average values for each category. Secondly, CatBoost introduces Ordered Boosting to overcome the gradient estimation bias in traditional Gradient Boosting Trees (GBDTs). This method utilizes sorting to generate a random arrangement order and number the dataset, enabling unbiased gradient estimation. Though it involves training a model for each sample, which increases space complexity, CatBoost optimizes Ordered Boosting and retains the foundational concepts of GBDT to improve model generalization and stability [25,26].

2.1.4. Light Gradient Boosting Machine

The Light Gradient Boosting Machine (LGBM) is another Gradient Boosting algorithm that uses leaf-wise techniques to grow trees vertically. To enhance the training process, LGBM uses the Gradient-based-One-Side-Sampling algorithm (GOSS) that is designed to emphasize the significance of data instances. GOSS focuses on data samples exhibiting

larger gradients while intentionally disregarding samples with lower gradients, assuming that they have lower errors. To mitigate potential bias towards samples with larger gradients, GOSS performs random sampling on data instances with small gradients while retaining all samples with large ones. Additionally, GOSS introduces a mechanism for adjusting the weights of data samples with small gradients during the computation of information gain, aiming to rectify and account for the inherent bias in the dataset [27].

2.1.5. Standard Machine Learning Classifiers

We compared the Gradient Boosting algorithms against standard machine learning classifiers, such as Naive Bayes, Decision Trees, and Logistic Regression. By contrasting these advanced techniques with simpler, more traditional approaches, we aimed to provide a comprehensive understanding of the performance differences between them. Naive Bayes, recognized for its straightforward probabilistic approach, Decision Trees, prized for their intuitive structure and ease of interpretation, and Logistic Regression, esteemed for its linear model formulation, serve as benchmarks in this discourse. Comparing these standard classifiers against Gradient Boosting algorithms allows for a comprehensive evaluation of predictive performance, scalability, and computational efficiency.

2.1.6. Experiment Setup and Model Comparison

Instances of each model were trained on the dataset delineated in Section Shapley Additive Explanations using the default parameters. The implementation was conducted utilizing Python 3.10 and the scikit-learn library. Subsequently, 5-fold cross-validation was employed across all models, and the resultant average metrics encompassing accuracy, precision, and recall were scrutinized for comparative analysis. The selection of the most suitable model for our dataset was based on these metrics. Detailed findings of this comparative evaluation are presented in Table 1.

Table 1. Machine learning algorithm comparison.

Algorithm	Class	Precision	Recall	Accuracy
XGBoost	Stable	0.90	0.89	0.86
	Transition	0.79	0.77	
CatBoost	Stable	0.90	0.88	0.85
	Transition	0.77	0.76	
Light Gradient Boosting	Stable	0.89	0.90	0.86
	Transition	0.76	0.75	
Decision Tree	Stable	0.85	0.86	0.79
	Transition	0.67	0.66	
Logistic Regression	Stable	0.70	0.91	0.66
	Transition	0.32	0.11	
Naive Bayes	Stable	0.87	0.62	0.66
	Transition	0.47	0.78	

The comparative analysis of machine learning algorithms demonstrates that XGBoost is the most effective performer, exhibiting superior precision and recall scores of 0.79 and 0.77, respectively, alongside an accuracy of 0.86. CatBoost and Light Gradient Boosting closely follow with similar precision, recall, and accuracy metrics, achieving 0.77, 0.76, and 0.85, and 0.76, 0.75, and 0.86 respectively. Decision Tree exhibits moderate performance, with precision and recall rates of 0.67 and 0.66, resulting in an accuracy of 0.79. However, Logistic Regression shows notably lower precision and recall values at 0.32 and 0.11, respectively, yielding an accuracy of 0.66. Naive Bayes, while displaying satisfactory recall at 0.78, demonstrates lower precision at 0.47, leading to an accuracy of 0.66, showcasing its precision–recall trade-offs compared to other algorithms.

2.2. Training Pipeline

After selecting the best model, we performed the training process using the workflow presented in Figure 5. We first trained an XGBoost instance using the default parameters to perform feature selection using the Recursive Feature Elimination algorithm (RFE). This is a powerful wrapper feature selection algorithm first introduced by Guyon et al. 2002. RFE removes iteratively weak features whose removal has the lowest impact on training errors and keeps strong features to improve the model generalization. The weakness of each feature is determined by a criterion based on the model that the algorithm wraps. In most cases, the weakness criterion is the feature importance scores assigned by the model during training. During each iteration, RFE selectively removes a specified number (N) of features based on their perceived weakness, as indicated by the importance criterion. Following the removal of features, the algorithm retrains the model on the new set of features and assesses its performance using a predetermined evaluation metric. This iterative process is repeated until there is no further improvement in model performance, at which point the algorithm concludes and returns the set of selected features [28].

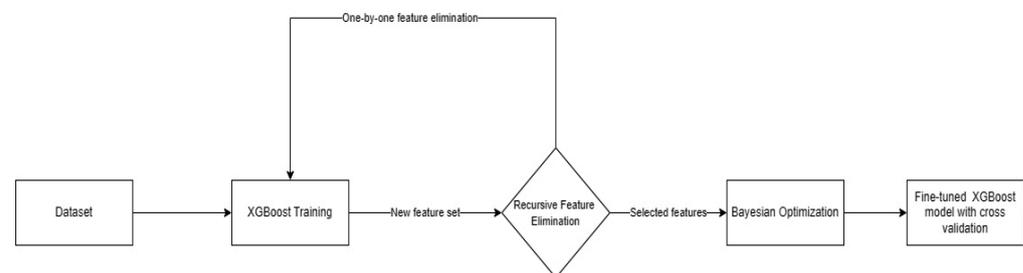


Figure 5. Training workflow used for creating the model.

In our case, we used RFE combined with 5-fold cross-validation for evaluation purposes. We adopted a stepwise approach with the step size set to 1, indicating the removal of one feature at each iteration during the Recursive Feature Elimination with Cross-Validation (RFE-CV) process. Our evaluation criteria were based on the F1 score, as defined in Equation (1). The F1 score, encompassing both precision (Equation (2)) and recall (Equation (3)), is particularly instrumental in comprehensively assessing the performance of the model. To perform feature selection, we utilized the RFE-CV technique implemented in the scikit-learn Python package. The workflow and outcomes of our RFE-CV implementation are depicted in Figure 6, illustrating the systematic elimination of features and the corresponding impact on the model's F1 score. This approach provides a robust means of feature selection, allowing us to identify the optimal subset of features that maximizes the model's predictive performance.

$$F_1 \text{ score} = 2 \times \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (1)$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (3)$$

RFE feature selection concluded with a subset of 75 features, chosen from an initial set of 105 features. The features selected after RFE can be found in Appendix A. It is important to note that the feature vector of 105 features included baseline observations that were engineered as a separate feature to maintain the baseline of each individual piece of knowledge across all samples.

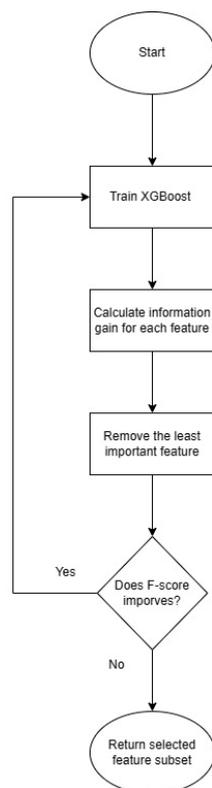


Figure 6. Recursive Feature Elimination workflow as implemented in scikit-learn.

Hyperparameter Tuning

After the feature selection process, we performed hyperparameter tuning to find the best set of hyperparameters to tune our XGBoost model and maximize the performance. Hyperparameters are a set of configuration parameters for a machine learning model that cannot be learned during the training process and require manual adjustment by the user. Selecting the optimal hyperparameters can prove challenging, often necessitating trial-and-error approaches when performed manually. To address this problem, various automatic hyperparameter search algorithms have been proposed throughout the years, treating the search problem as an optimization problem where the objective function is unknown or a blackbox function. In this paper, we chose the Bayesian Optimization (BO) algorithm to efficiently determine the optimal set of hyperparameters using a few samples of the dataset. Bayesian optimization proves to be a valuable approach in tackling functions where determining extrema is computationally expensive. Its applicability extends to functions lacking a closed-form expression and those that involve expensive calculations, challenging derivative evaluations, or exhibit non-convex characteristics. In summary of the process, BO models the relationship between a given set of parameter combinations and model's performance by fitting a Gaussian Process (GP). Then, it optimizes the GP to find the next promising combination of parameters. Each combination is evaluated by training the model and evaluating it using a given evaluation metric. In this paper, the evaluation metric is the F1-score presented in Equation (1). The process is repeated until a specific criterion is met. In most cases, the ending criterion is the maximum number of iterations provided by the researcher [29,30].

After the application of Bayesian Optimization, the parameters presented in Table 2 were attained.

Table 2. Parameters selected after Bayesian Optimization.

Parameter	Value
colsample_bytree	0.6714223800630487
gamma	0.7244817045778367
Learning_rate	0.01
min_child_weight	10
n_estimators	1000
Scale_pos_weight	5.0
subsample	0.5676926067435525
Max_depth	7

3. Results

We trained the final XGboost instance using the features selected from RFE and the hyperparameters chosen with Bayesian Optimization. To assess the robustness and generalization capability of the model, we employed stratified 5-fold cross-validation. This technique involves splitting the dataset into five subsets (folds), ensuring that each fold maintains the same class distribution as the original dataset. The model is trained on four folds and validated on the remaining one in each iteration. This process is repeated five times, and the performance metrics, including F1-score, precision, recall, accuracy, and ROC AUC score, are averaged over the folds to provide a comprehensive evaluation. ROC AUC score stands for Area Under the Receiver Operating Characteristic curve and shows how well a binary classification can distinguish the two classes. In medical diagnosis, such information is vital, as misclassifying patients can lead to greater disease severity. Summarized results from the table process are presented in Table 3.

Table 3. Average results from 5-fold cross-validation.

Target Class	Precision	Recall	F1-Score	Accuracy	ROC AUC
Stable	0.94	0.84	0.90	0.85	0.86
Transition	0.71	0.88	0.79		

As expected, the model demonstrates higher accuracy in predicting the stable class compared to the transition class, achieving precision and recall scores exceeding 0.90 in some folds. This can be ascribed to the absence of any rebalancing methodology applied during the dataset curation process, thereby facilitating a dataset composition that endeavors to mirror real-world conditions to the greatest extent possible. Also, the difference observed between recall and precision in the transition class is ascribed to the use of the scale_pos_weight parameter during the hyperparameter optimization phase. This parameter imparts a heightened sensitivity of the model to the transition class as a measure to mitigate class imbalance. This technique results in slightly higher recall scores accompanied by a marginal reduction in precision scores, which deviate slightly from anticipated levels. The rationale for accepting this trade-off lies in the heightened significance accorded to accurately classifying patients within the transition class (as indicated by the higher recall), even if it entails the inclusion of certain false-positive predictions in the resultant predictive outcomes. Additionally, it is imperative to recognize the inherent complexity of the Alzheimer’s disease (AD) diagnosis and prognosis, constituting an exceptionally challenging problem that necessitates a substantial volume of heterogeneous data for achieving accurate predictions. This effect can be visually observed in the summarized confusion matrix in Figure 7.

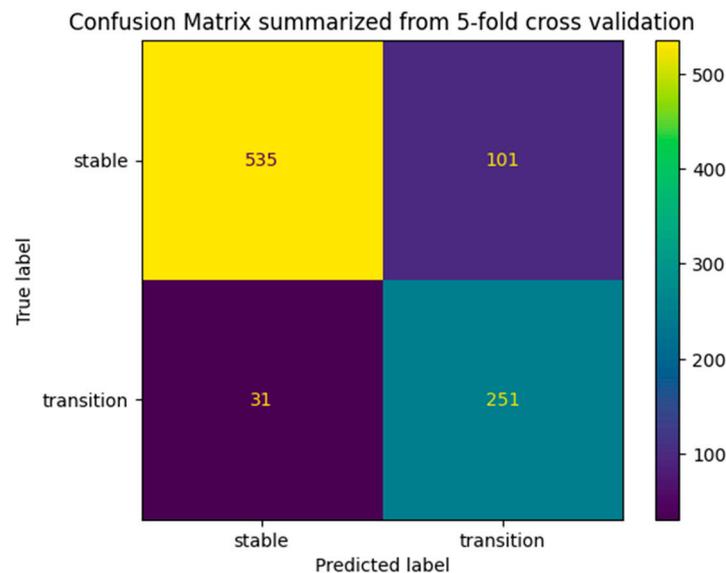


Figure 7. Summarized confusion matrix from 5-fold cross-validation.

Shapley Additive Explanations

Interpreting the model's predictions is crucial when it comes to medical diagnosis. To address this issue, there have been numerous sophisticated methodologies to interpret blackbox machine learning models. One of the most famous explainable AI methodologies is SHAP values, which find their origin in game theory. Fundamentally, SHAP transforms the feature space of XGBoost into a clinical variable space, wherein each transformed SHAP value corresponds to an original variable. This transformation facilitates a more clinically meaningful interpretation of the model's output. Graphically, SHAP often visualizes XGBoost predictions to enhance interpretability. For instance, the SHAP summary plot succinctly illustrates the magnitudes and directions of predictions, wherein the size of a SHAP value signifies the contribution of a specific feature. Larger values indicate a more substantial contribution to prediction performance. Moreover, the SHAP dependency plot offers insights into the distribution of SHAP values across individuals for a given feature. As SHAP values vary among individuals, so do the predictions of the corresponding feature mappings for those individuals. This understanding provides a comprehensive view of the impact of individual features on model predictions, offering a valuable tool for researchers and practitioners in comprehending and validating the XGBoost model's decision-making processes [18,31]. We applied SHAP Additive Explanations to interpret the model's decision-making processes.

4. Discussion

In this study, an interpretable machine learning model was developed using eXtreme Gradient Boosting and the Shapley explanation framework. The model was trained on the ADNI dataset, including all four phases of ADNI, with the latest being ADNI-3 at the time of writing. The dual purpose of this study was to develop a machine learning model with the primary goal of accurately distinguishing between individuals with Mild Cognitive Impairment (MCI) who will experience stability in their cognitive function (MCIs) and those who will progress to Alzheimer's disease soon (MCIs). Additionally, the secondary aim was to interpret the model's decision process to help clinicians understand the reasons behind each prediction to make them trust the ML system. To achieve this, we chose to train the model across all ADNI phases in an effort to develop a framework that can be adopted in heterogeneous dataset types since the data collection protocol has changed gradually from ADNI1 to the ADNI3 phase. It is essential to acknowledge that we did not employ any direct balancing methodology in the development of our model. However, to mitigate potential issues arising from the unbalanced nature of our data, we utilized

stratified k-fold cross-validation. This decision was deliberate, as it reflects a conscious choice to align our modeling approach with the prevalence of Mild Cognitive Impairment and Alzheimer's disease patients in the general population. In real-life conditions, individuals diagnosed with MCI and AD represent a minority within the broader demographic landscape. This approach is particularly crucial in the development of medical models, where accurately mirroring real-world prevalence is essential for model reliability and relevance. In addressing the target class modeling problem, we adopted a unique approach. Our strategy involved grouping MCI patients, regardless of the onset time of AD up to their 48th visit since baseline, subsequently annotating the conversion using a binary target feature created in the dataset used. This approach was used to construct a model capable of providing predictions that reflect diverse real-life scenarios. Through exposing the model to a variety of samples showcasing conversion tendencies, our objective was to direct its learning process in discerning the factors indicative of a patient's potential progression to AD. Our methodology surpassed similar techniques in performance, although it is worth acknowledging the current limitations of our comparative analysis. At the time of writing, most of the research available mainly focuses on developing diagnostic or potentially prognostic methodologies, rather than specifically tailored prognostic techniques. Additionally, while most prognostic methods predict within a fixed time window for conversion, our approach offers a more adaptable timeframe of 4 years (48 months) from the baseline visit. A method proposed by Fuliang et al. (2023) can partially be compared with our proposed solution since we both used similar approaches in terms of XGBoost and SHAP setup, but their approach is purely diagnostic. Our model showcased higher-sensitivity performance, which is crucial in medical models, despite their diagnostic orientation [18]. Other studies that use machine learning for the early diagnosis of AD analyzing MRIs and their variants also seem to slightly fall under in terms of performance when compared to our approach. Although they exhibit accuracy levels similar to ours, their sensitivity scores are slightly lower in the same time window [5,32,33].

In Figure 8, the SHAP summary plot illustrates the top-10 features generated by XGBoost. These features are arranged in descending order based on their SHAP values for all predictions. The SHAP values indicate the positive or negative associations of the respective features, and the absolute SHAP value for each feature is displayed on the left side. Every data point depicted in the plot represents an individual sample, with the horizontal axis indicating the SHAP value of a specific feature across subjects. This axis serves as a gradient, ranging from low (depicted in blue) to high (represented in red), reflecting the magnitude of the SHAP value. Apparently, cognitive test results have a moderate influence on model outcome. For example, high scores on the ADAS-Cog 13 scale indicate a possible cognitive impairment and, thus, a possible transition to AD in the near future [34]. In contrast with ADAS-Cog-13, lower scores on the PACC scale (mPACCtrailsB and mPACCdigit) indicate potential cognitive decline [35]. As anticipated, additional biomarkers derived from imaging and clinical tests exert a considerable impact on the ultimate decision. Biomarkers such as ABETA and Entorhinal, indicative of the brain's status, exhibit notably low values, contributing significantly to the overall outcome.

The features that impact the model based on their average SHAP value are presented in Figure 9. It is important to highlight the role of the "DX_bl" feature, representing the baseline diagnosis, in our prediction process. Our dataset includes Alzheimer's disease (AD) patients who did not show cognitive decline at the start, aligning with our study's focus. This feature mainly helps the model differentiate between AD and Mild Cognitive Impairment (MCI) patients, rather than distinguishing between progressing MCI (MCIc) and stable MCI (MCIs). As expected, biomarkers and assessment tests related to cognitive function have a moderate impact on the model's outcome.

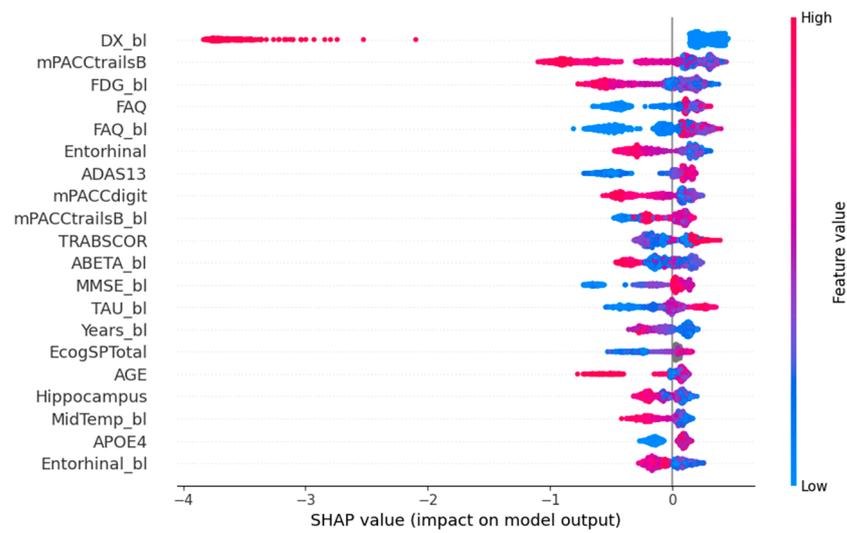


Figure 8. SHAP summary plot.

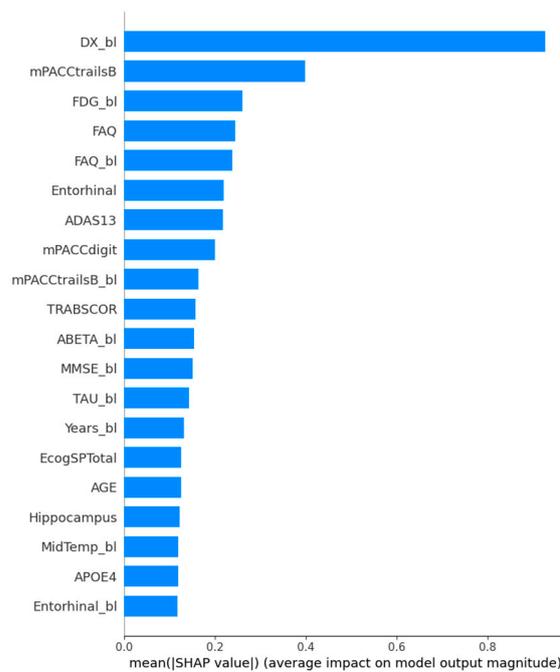


Figure 9. SHAP feature importance plot.

5. Conclusions

In real-world clinical studies, the early identification of Alzheimer’s disease is a complex and challenging task for clinicians involving various procedures, including cognitive, psychiatric, and biomarker tests. Not only are some of these assessments time-consuming for both clinicians and patients but they also involve invasive procedures, rendering them not only inefficient in terms of time but also cost-ineffective. Machine learning can improve the prognostic process for Alzheimer’s disease by identifying useful information within large amounts of clinical data and, thus, pave the way for more time-effective and non-invasive prognostic methods. At the same time, machine learning can help clinicians understand how the disease progresses in patients through biomarker analysis, enabling the timely administration of appropriate treatment. Our proposed framework can contribute to the above aims. Our results showed that the proposed XGBoost model can accurately predict the conversion from MCI to AD within 36 months from baseline with low error rates. Integrating the SHAP framework with the prediction model allows us to

interpret the model's decision process for each given sample, providing transparency to clinicians and enhancing the model's reliability. Despite these promising results, further research is essential to develop models with minimal error rates and implement them in real-world conditions. To achieve this, it is imperative to train models on extensive datasets, representing real-life scenarios that may differ across various clinics and hospitals. Furthermore, the optimization of these models requires attention to factors influencing the decision-making process in different healthcare settings, always prioritizing the well-being and safety of patients.

6. Limitations

This study has potential limitations. The proposed model was specifically trained to predict the transition from Mild Cognitive Impairment (MCI) to Alzheimer's disease (AD), while excluding the possibilities of transitions from Cognitively Normal (CN) to MCI and CN to AD. The proposed approach needs further clinical verification against an external dataset. The model needs two timepoints to make an accurate prediction, which translates to two physical visits.

Author Contributions: Conceptualization, G.G. and T.E.; methodology, G.G.; software, G.G.; validation, T.E., P.V. and A.G.V.; formal analysis, G.G.; investigation, D.P.; resources, D.P.; data curation, G.G.; writing—original draft preparation, G.G.; writing—review and editing, D.P. and T.E.; visualization, G.G.; supervision, P.V.; project administration, A.G.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: This study is secondary analysis of already available data.

Data Availability Statement: Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U19 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org, accessed on 11 February 2024). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

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

Appendix A

Below is the subset of features selected after Recursive Feature Elimination. For continuous values, the median, mean, max, and standard deviation are provided, and for discrete values, the membership status is provided.

	Feature	Median	Mean	Max	Std	Membership Status
0	DX_bl					[2, 0, 1, 0]
1	AGE	74	73.53801743	91.4	7.610487740946301	
2	PTEDUCAT	16	15.790849673202615	20	2.8848952582917042	
3	APOE4					[1, 0, 0, 0, 2, 0]
4	FDG	1.422475	1.279733683442266	1.57338	0.17810577352607126	
5	CDRSB	2	2.6919389978213513	15	2.3115707124725917	
6	ADAS11	11.33	13.06316557734205	56.33	7.848523985235578	
7	ADAS13	19	20.335211328976037	71.33	10.712926813970377	
8	ADASQ4	7	6.328322440087145	10	2.8895712942428418	
9	MMSE	27	25.861220043572985	30	3.819875621530087	
10	RAVLT_immediate	30	30.709586056644877	71	11.989765521262475	
11	RAVLT_learning	3	3.3753812636165574	12	2.626026021958254	
12	RAVLT_forgetting	5	4.550762527233116	14	2.727335365616795	
13	RAVLT_perc_forgetting	83.3333	68.94227735294118	100	66.12887476892881	
14	LDELTOTAL	4	5.311111111111112	25	5.1133143144938495	
15	TRABSCOR	107.5	138.7888888888889	300	83.34054120772674	
16	FAQ	4	7.128322440087145	30	7.761732851339966	
17	Ventricles	39,943.79	44,969.496797385626	151,426	23,686.835523710066	
18	Hippocampus	6390.5	6366.0328758169935	10,452	1201.5157640328403	
19	WholeBrain	1,007,820	1,009,664.1372549019	1,428,190	112,656.48231469237	
20	Entorhinal	3281.6	3285.5557734204795	5770	773.1931836208981	
21	Fusiform	16,739.5	16,767.563834422657	28,878	2781.09519	
22	MidTemp	18,553.4	18,716.070370370373	29,006	2960.8444166415316	
23	ICV	1,522,580	1,540,374.3877995643	2,100,210	164,601.06052299083	
24	mPACCdigit	-8.21001	-8.66117474	5.95912	6.933461669969725	
25	mPACCtrailsB	-7.9708	-8.337507947	6.13315	6.824617271140973	
26	CDRSB_bl	1.5	2.070806100217865	10	1.5453513515946802	
27	ADAS11_bl	11	11.881372549019607	36	5.700384925550372	
28	ADAS13_bl	18	18.951222222222224	50	8.290366685918455	
29	MMSE_bl	27	26.715686274509803	30	2.550000345894958	
30	RAVLT_immediate_bl	30	32.001960784313724	68	10.788177445521617	
31	RAVLT_learning_bl	3	3.621786492374728	11	2.575508392318168	
32	RAVLT_forgetting_bl	5	4.614596949891068	13	2.296667824916341	
33	RAVLT_perc_forgetting_bl	71.4286	66.49644300653596	100	32.46909675562216	
34	LDELTOTAL_BL	4	4.790849673202614	18	3.62889217	
35	TRABSCOR_bl	105.1	131.97015250544663	300	76.91764281066624	
36	FAQ_bl	3	5.153159041394336	30	6.195501396652368	
37	mPACCdigit_bl	-7.385415	-7.820029706	2.23768	4.944942117584728	
38	mPACCtrailsB_bl	-6.988085	-7.485260064	2.7732	4.926875777371472	
39	Ventricles_bl	37,837.5	42,501.538061002175	157,713	22,875.046972145934	
40	Hippocampus_bl	6528	6547.429477124183	9929	1178.2954938265814	
41	WholeBrain_bl	1,015,965	1,020,942.0840958606	1,443,990	113,929.04757611542	
42	Entorhinal_bl	3360.5	3365.5603485838783	5896	770.1566993159872	
43	Fusiform_bl	17,023.5	17,064.395642701526	26,280	2763.6379991806716	
44	MidTemp_bl	19,086.3	19,171.861437908494	29,292	2966.722264517154	
45	ICV_bl	1,527,190	1,542,547.285	2,714,340	169,932.87108013846	
46	FDG_bl	1.1870850000000002	1.2005082629629629	1.70113	0.1342208035987261	
47	Years_bl	1.00205	1.0118026601307188	1.2512	0.049772686	
48	TAU_bl	298.89	305.8513442265795	816.9	110.61959394290388	
49	ABETA_bl	754.74	867.6725054466232	1700	380.51187522862955	
50	PTAU_bl	29.145	30.04860784313726	94.86	12.535837824740023	
51	TAU	297.9	317.38917211328976	802.4	73.68806040025473	
52	DX					[2, 1]
53	MOCA	23	22.769162995594716	30	4.020161189133941	
54	EcogPtLang	1.77778	1.8973843788546254	4	0.6874612645623314	
55	EcogPtVisspat	1.28571	1.507761947136564	4	0.6135489175939899	
56	EcogPtPlan	1.4	1.5351908810572688	3.8	0.6146249328058496	
57	EcogPtDivatt	2	2.006057268722467	4	0.8197753125466491	
58	EcogPtTotal	1.74359	1.8405794669603526	3.69231	0.5772989648737149	
59	EcogSPLang	1.58611	1.8575385903083699	4	0.7996222455335457	
60	EcogSPPlan	1.5	1.8114684140969162	4	0.9016202278521113	
61	EcogSPOrgan	1.66667	1.9126724317180617	4	0.9180337285803657	
62	EcogSPDivatt	2	2.182672577092511	4	0.9494293241028594	
63	EcogSPTotal	1.74359	1.9563972290748899	3.97368	0.7966188010118953	
64	MOCA_bl	23	22.822907488986782	30	3.5954600273400152	
65	EcogPtLang_bl	1.77778	1.9069771189427316	4	0.6946476139514485	
66	EcogPtVisspat_bl	1.28571	1.4918661453744495	4	0.619250998	
67	EcogPtOrgan_bl	1.416666	1.6129223039647576	4	0.6744607742485486	
68	EcogPtDivatt_bl	1.75	1.9948237885462554	4	0.8070685154926579	
69	EcogPtTotal_bl	1.7142400000000002	1.8485160044052862	3.85294	0.5781348132153441	
70	EcogSPMem_bl	2.25	2.3429122026431717	4	0.8585914694597968	
71	EcogSPLang_bl	1.55556	1.7631507488986786	4	0.7370702823814411	
72	EcogSPVisspat_bl	1.266666	1.5279054140969162	4	0.6853481425640744	
73	EcogSPOrgan_bl	1.5	1.735022181	4	0.800743658	
74	EcogSPTotal_bl	1.71053	1.8598564933920703	3.89744	0.6867589759636067	
75	Transition					[0, 1]

References

- Portet, F.; Ousset, P.J.; Visser, P.J.; Frisoni, G.B.; Nobili, F.; Scheltens, P.; Vellas, B.; Touchon, J.; MCI Working Group of the European Consortium on Alzheimer's Disease. Mild cognitive impairment (MCI) in medical practice: A critical review of the concept and new diagnostic procedure. Report of the MCI Working Group of the European Consortium on Alzheimer's Disease. *J. Neurol. Neurosurg. Psychiatry* **2006**, *77*, 714. [[CrossRef](#)] [[PubMed](#)]
- Alzheimers Facts and Figures Report 2022. 2022. Available online: <https://www.alz.org/media/Documents/alzheimers-facts-and-figures-special-report-2022.pdf> (accessed on 6 May 2023).
- 2023 Alzheimer's disease facts and figures. *Alzheimer's Dement.* **2023**, *19*, 1598–1695. [[CrossRef](#)] [[PubMed](#)]
- Mofrad, S.A.; Lundervold, A.; Lundervold, A.S. A predictive framework based on brain volume trajectories enabling early detection of Alzheimer's disease. *Comput. Med. Imaging Graph.* **2021**, *90*, 101910. [[CrossRef](#)] [[PubMed](#)]
- Zhang, T.; Liao, Q.; Zhang, D.; Zhang, C.; Yan, J.; Ngetich, R.; Zhang, J.; Jin, Z.; Li, L. Predicting MCI to AD Conversion Using Integrated sMRI and rs-fMRI: Machine Learning and Graph Theory Approach. *Front. Aging Neurosci.* **2021**, *13*, 688926. [[CrossRef](#)] [[PubMed](#)]
- Graff-Radford, J.; Yong, K.X.; Apostolova, L.G.; Bouwman, F.H.; Carrillo, M.; Dickerson, B.C.; Rabinovici, G.D.; Schott, J.M.; Jones, D.T.; Murray, M.E. New Insights into Atypical Alzheimer's Disease in the Era of Biomarkers. *Lancet Neurol.* **2021**, *20*, 222. [[CrossRef](#)] [[PubMed](#)]
- Blennow, K.; Zetterberg, H. Biomarkers for Alzheimer's disease: Current status and prospects for the future. *J. Intern. Med.* **2018**, *284*, 643–663. [[CrossRef](#)] [[PubMed](#)]
- Bron, E.E.; Bron, E.E.; Klein, S.; Papma, J.M.; Jiskoot, L.C.; Venkatraghavan, V.; Linders, J.; Aalten, P.; De Deyn, P.P.; Biessels, G.J.; et al. Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease. *Neuroimage Clin.* **2021**, *31*, 102712. [[CrossRef](#)] [[PubMed](#)]
- Vrahatis, A.G.; Skolariki, K.; Krokidis, M.G.; Lazaros, K.; Exarchos, T.P.; Vlamos, P. Revolutionizing the Early Detection of Alzheimer's Disease through Non-Invasive Biomarkers: The Role of Artificial Intelligence and Deep Learning. *Sensors* **2023**, *23*, 4184. [[CrossRef](#)]
- Chang, C.H.; Lin, C.H.; Lane, H.Y. Machine Learning and Novel Biomarkers for the Diagnosis of Alzheimer's Disease. *Int. J. Mol. Sci.* **2021**, *22*, 2761. [[CrossRef](#)]
- Ko, H.; Ihm, J.J.; Kim, H.G. Cognitive profiling related to cerebral amyloid beta burden using machine learning approaches. *Front. Aging Neurosci.* **2019**, *11*, 439698.
- Salvatore, C.; Cerasa, A.; Battista, P.; Gilardi, M.C.; Quattrone, A.; Castiglioni, I. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach. *Front. Neurosci.* **2015**, *9*, 144798.
- Singh, A.; Kumar, R.; Tiwari, A.K. Prediction of Alzheimer's Using Random Forest with Radiomic Features. *Comput. Syst. Sci. Eng.* **2022**, *45*, 513–530. [[CrossRef](#)]
- Akter, L.; Ferdib-Al-Islam. Dementia Identification for Diagnosing Alzheimer's Disease using XGBoost Algorithm. In Proceedings of the 2021 International Conference on Information and Communication Technology for Sustainable Development, ICICT4SD 2021, Dhaka, Bangladesh, 27–28 February 2021; pp. 205–209. [[CrossRef](#)]
- Bucholc, M.; Titarenko, S.; Ding, X.; Canavan, C.; Chen, T. A hybrid machine learning approach for prediction of conversion from mild cognitive impairment to dementia. *Expert. Syst. Appl.* **2023**, *217*, 119541. [[CrossRef](#)]
- Balaji, P.; Chaurasia, M.A.; Bilfaqih, S.M.; Muniasamy, A.; Alsid, L.E.G. Hybridized Deep Learning Approach for Detecting Alzheimer's Disease. *Biomedicines* **2023**, *11*, 149. [[CrossRef](#)] [[PubMed](#)]
- Bogdanovic, B.; Eftimov, T.; Simjanoska, M. In-depth insights into Alzheimer's disease by using explainable machine learning approach. *Sci. Rep.* **2022**, *12*, 6508. [[CrossRef](#)] [[PubMed](#)]
- Yi, F.; Yang, H.; Chen, D.; Qin, Y.; Han, H.; Cui, J.; Bai, W.; Ma, Y.; Zhang, R.; Yu, H. XGBoost-SHAP-based interpretable diagnostic framework for alzheimer's disease. *BMC Med. Inform. Decis. Mak* **2023**, *23*, 137. [[CrossRef](#)] [[PubMed](#)]
- ADNI | About. Available online: <https://adni.loni.usc.edu/about/> (accessed on 7 November 2023).
- ADNIMERGE: Clinical and Biomarker Data from All ADNI Protocols • ADNIMERGE. Available online: <https://adni.bitbucket.io/index.html> (accessed on 7 November 2023).
- ADNI_General Procedures Manual. 2006. Available online: https://adni.loni.usc.edu/wp-content/uploads/2024/02/ADNI_General_Procedures_Manual_29Feb2024.pdf (accessed on 7 November 2023).
- Troyanskaya, O.; Cantor, M.; Sherlock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D.; Altman, R.B. Missing value estimation methods for DNA microarrays. *Bioinformatics* **2001**, *17*, 520–525. [[CrossRef](#)] [[PubMed](#)]
- Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016. [[CrossRef](#)]
- Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. *arXiv* **2019**, arXiv:1706.09516v5.
- Dorogush, A.V.; Ershov, V.; Yandex, A.G. CatBoost: Gradient Boosting with Categorical Features Support. October 2018. Available online: <https://arxiv.org/abs/1810.11363v1> (accessed on 3 December 2023).
- Zhenyu, Z.; Yang, R.; Wang, P. Application of explainable machine learning based on Catboost in credit scoring. *J. Phys. Conf. Ser.* **2021**, *1955*, 12039. [[CrossRef](#)]

27. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Available online: <https://github.com/Microsoft/LightGBM> (accessed on 3 December 2023).
28. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422.
29. Wu, J.; Chen, X.Y.; Zhang, H.; Xiong, L.D.; Lei, H.; Deng, S.H. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26–40. [[CrossRef](#)]
30. Bloch, L.; Friedrich, C.M. Using Bayesian Optimization to Effectively Tune Random Forest and XGBoost Hyperparameters for Early Alzheimer’s Disease Diagnosis. In *Wireless Mobile Communication and Healthcare, Proceedings of the 9th EAI International Conference, MobiHealth 2020, Virtual Event, 19 November 2020*; Springer: Cham, Switzerland, 2021. [[CrossRef](#)]
31. Lundberg, S.M.; Allen, P.G.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. Available online: <https://github.com/slundberg/shap> (accessed on 11 December 2023).
32. Syaifullah, A.H.; Shiino, A.; Kitahara, H.; Ito, R.; Ishida, M.; Tanigaki, K. Machine Learning for Diagnosis of AD and Prediction of MCI Progression from Brain MRI Using Brain Anatomical Analysis Using Diffeomorphic Deformation. *Front. Neurol.* **2021**, *11*, 576029. [[PubMed](#)]
33. Lin, W.; Gao, Q.; Yuan, J.; Chen, Z.; Feng, C.; Chen, W.; Du, M.; Tong, T. Predicting Alzheimer’s Disease Conversion from Mild Cognitive Impairment Using an Extreme Learning Machine-Based Grading Method with Multimodal Data. *Front. Aging Neurosci.* **2020**, *12*, 509232. [[CrossRef](#)] [[PubMed](#)]
34. Anderson, N.H.; Woodburn, K. Old-age psychiatry. In *Companion Psychiatric Studies*; Elsevier: Amsterdam, The Netherlands, 2010; pp. 635–692. [[CrossRef](#)]
35. Donohue, M.C.; Sperling, R.A.; Salmon, D.P.; Rentz, D.M.; Raman, R.; Thomas, R.G.; Weiner, M.; Aisen, P.S.; Australian Imaging, Biomarkers, and Lifestyle Flagship Study of Ageing; Alzheimer’s Disease Neuroimaging Initiative; et al. The Preclinical Alzheimer Cognitive Composite: Measuring Amyloid-Related Decline. *JAMA Neurol.* **2014**, *71*, 961. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.