DOI Link: https://doi.org/10.61586/CyWjG
Vol.36, Issue.10, Part.1, Oct 2025, PP.2-25

A Machine Learning-Based Approach for Predicting Alzheimer's Diseases Using Data Classification and Explainable Artificial Intelligence Techniques

Hosam El-Sofany[1, *], Abd Elrahman H. El-Sofany[2], Abdelmoty M. Ahmed[3], Shanawaz Ahmed J.[1], Belgacem Bouallegue[4]

[1]College of Computer Science, King Khalid University, Abha, Saudi Arabia
[2]British University in Egypt (BUE), Faculty of Informatics and Computer Science, Cairo, Egypt
[3]Faculty of Information and Technology, Ajloun National University, Ajloun, Jordan
[4]Electronics and Micro-Electronics Lab., Faculty of Sciences of Monastir, University of Monastir, Tunisia
[*]Corresponding Author: Hosam El-Sofany, Email: helsofany@kku.edu.sa

Received Feb 2025 Accepted Sep 2025 Published Oct 2025

Abstract

Globally, healthcare systems encounter tremendous obstacles as a result of the progressive neurological disease called Alzheimer's. To effectively intervene and control Alzheimer's disease (AD), the earliest and most accurate diagnosis is necessary. However, traditional diagnostic methods are often expensive, take a long time, and do not provide the accuracy needed for early diagnosis. This study addresses these limitations by proposing a machine learning-based (ML-based) approach for predicting AD using advanced data classification methods and an explainable artificial intelligence (AI) approach. Three distinct methods were utilized to carry out the feature selection procedure: chi-square, mutual information, and analysis of variance (ANOVA). We identified the analysis's most relevant elements by utilizing each technique. We found the best algorithm for predicting the early signs of AD by testing seven different ML methods: logistic regression, AdaBoost, random forest, support vector machine, decision tree, XGBoost, and K-nearest neighbors. We employed the SMOTE method to rectify the data imbalance. To test the proposed method, we employed both a publicly available and a private dataset. We applied multiple cross-validation approaches to provide a strong performance evaluation. The results of the experiments illustrated that, out of all the models tested, the XGBoost classifier performed the best. Using the combined dataset, the XGBoost classifier had 97.32% accuracy, 96.56% precision, 97.00% specificity, 97.68% sensitivity, 98.43% AUC, and 97.12% F1-score. Using the public dataset, XGBoost achieved 97.23% accuracy, 96.14% precision, 96.52% specificity, 98.03% sensitivity, 98.30% AUC, and 97.07% F1-score. Furthermore, XGBoost did exceptionally well on the private dataset with 95.83% accuracy, 93.94% sensitivity, 96.88% precision, 97.44% specificity, 98.52% AUC, and 95.38% F1-score. Understanding the model's findings and decision-making process can be enhanced with the help of an explicable AI framework that was developed using SHAP methods. The proposed approach shows enormous potential as a healthcare solution that reduces healthcare costs and improves efficiency in AD's diagnosis. Patients benefit from improved diagnostic tools for AD brought about by this study's combination of powerful ML models with explainable AI.

Keywords: Alzheimer's diseases, Machine learning applications, AI applications, SMOTE, SHAP

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disease that causes alterations in behavior, loss of memory, and mental abilities over time. 60–80% of all cases globally of AD are caused by these alterations [1]. Currently, an estimated 55 million individuals across the globe are coping with AD, and that figure will nearly triple by 2050 as a result of both population aging and longer life expectancy [2]. Not only does the illness impact people directly, but it also causes healthcare systems and their families great emotional and financial stress. Projections indicate that AD will cost the world $1.3 trillion in 2023, with a significant increase expected in the coming decades [3]. Early detection and intervention are critical to slowing disease progression, improving quality of life, and reducing healthcare costs. However, traditional diagnostic methods, such as clinical evaluations, neuropsychological testing, and neuroimaging, are often expensive and time-consuming and require specialized expertise, limiting their accessibility in nations with low or medium incomes, in particular [4].

Artificial intelligence (AI) and machine learning (ML) have been an innovation in the healthcare industry recently. These technologies have revolutionized long-standing issues such as disease prognosis, diagnostics, and personalized treatment strategies. These technologies excel at analyzing complex datasets, identifying patterns, and making predictions with high accuracy, making them particularly suited for addressing challenges in AD detection [5]. By leveraging large datasets, ML algorithms can uncover subtle biomarkers and risk factors that traditional approaches might miss. To find indicators of AD, ML models may search through genetic data, healthcare tracks, and neuroimaging data long before the illness's symptoms appear [6].

Countries like Saudi Arabia and Egypt are facing a growing number of AD cases and mounting demand from healthcare systems to provide trustworthy diagnosis methods. Therefore, expertise is crucial in these countries [7]. The Middle East and North Africa (MENA) are experiencing a demographic shift, with increasingly elderly inhabitants and a proportional rise in illnesses associated with aging, including AD. In Egypt, for instance, by 2030, experts predict that the number of individuals affected by AD will have doubled, while Saudi Arabia is projected to see a 300% increase in dementia cases by 2050 [8]. Despite this increasing need, these regions still do not have the specialized healthcare infrastructure or resources necessary to address the needs of those who live with AD. The situation highlights the urgent need for cost-effective, scalable, and accessible diagnostic tools that can support early diagnosis and intervention.

This work primarily focuses on using ML to successfully detect AD in its early phases. We tested several ML algorithms on publicly available AD datasets and a private dataset from specialized hospitals in Egypt to achieve this. This private dataset, comprising 345 data samples with 34 features, was voluntarily provided by healthcare institutions between 2023 and 2025, offering a unique opportunity so that our models can be progressively applied [9]. By combining this private dataset with publicly available data, we aimed to construct an innovative ML technique capable of accurately classifying high-definition datasets and outperforming existing models.

Significant contributions in the field of AD's prediction have been made as a result of this work, including the following:

- By using ML classification algorithms on a combined dataset, the study tackles the pressing need for early identification in places like Egypt and Saudi Arabia that are experiencing fast-increasing AD rates.
- The research introduces a novel approach that attains state-of-the-art performance measures by merging XGBoost with a semi-supervised model. On the private and

combined datasets, these metrics successfully reached a 95.83% and 97.32% accuracy rate, respectively [10].

- The study uses an explainable AI framework that uses SHAP (SHapley Additive exPlanations) techniques to give clear explanations of the model's predictions. The study is done to build trust and transparency in AI-based diagnostics. To make smart clinical decisions, you need to know how a model came to its predictions. This is where explainable AI truly excels. Given that SHAP values highlight the critical components of the prediction process, healthcare practitioners can have faith in the model's output and make informed judgments. In contrast, the inability to understand normal "black-box" models in mission-critical situations makes them unreliable [12].
- This research study also makes a substantial contribution by addressing the issue of skewed data using SMOTE. One typical problem in medical research is dealing with imbalanced datasets when one class is greatly underrepresented. Such an imbalance might result in biased models that do badly on minority classes. Our classifiers performed better when we used SMOTE to create synthetic minority class samples, resulting in a balanced dataset [13].
- To maximize the performance of our ML-based model, we tweaked their hyperparameters and applied SMOTE. Hyperparameters govern the learning process of ML algorithms; selecting appropriate ones significantly influences a model's efficiency. Through organized tuning of each classifier's hyperparameters, we were able to fine-tune our models to their full potential, resulting in more reliable and accurate predictions [14].
- The research utilized three feature selection approaches—*mutual information* (MI), *analysis of variance* (ANOVA), and *chi-square*—on both public and private datasets. Our goal was to enhance classification accuracy by reducing high-dimensional data using the ANOVA approach, identifying significant features using feature space, and analyzing feature similarity. With this goal in mind, to find out how important each parameter was, we utilized the chi-square test. We utilized MI to improve model interpretability, handle outliers effectively, capture nonlinear correlations, choose the most significant features by assessing the interdependence of variables, and conduct this investigation.
- We evaluated seven alternative ML algorithms including random forest, logistic regression, AdaBoost, decision tree, XGBoost, K-nearest neighbor (KNN), and support vector machine (SVM) to find the one that best predicted the occurrence of AD. In comparison to the other models, XGBoost performed better on both public and private datasets concerning accuracy, area under the curve (AUC), sensitivity, and specificity. This result confirms what previous research has shown: that XGBoost is an excellent algorithm for use in medical diagnosis and prediction [15].

Improvements in ML-based AD prediction systems may be possible as a result of this study, providing healthcare providers with an economical and expedient solution. By using powerful ML models along with explainable AI, this study helps create trustworthy and understandable tools for diagnosing AD, which can support early treatment plans and enhance patient care.

2. Related Work and Comparative Study

There has been a lot of recent progress in using ML-based methods to forecast the onset of AD because of the increasing importance of early and accurate diagnostics. Several studies have investigated the possibility of using ML algorithms to examine various datasets to discover biomarkers and patterns linked to AD. These datasets include neuroimaging data,

genetic information, and clinical records. This section reviews recent studies that address the problem of AD prediction using ML-based approaches and offers a comparison of how well various classifiers performed.

2.1 Recent Studies on AD Prediction Using ML

ML for AD early detection and prediction has been the subject of multiple recent studies. These studies highlight the importance of accurate classification techniques and the integration of understandable AI to increase transparency in clinical decision-making.

Buyrukoglu [15] proposed a comparison of ensemble feature selection approaches for detecting AD using data mining methods. The study used RF, DT, and KNN classifiers on a clinical dataset and found that using ensemble techniques greatly improved prediction accuracy.

Dhakal et al. [16] explored dementia prediction through supervised ML algorithms using a clinical dataset. They focused on RF, SVM, and KNN and reported strong predictive performance, emphasizing the importance of preprocessing, and balancing in achieving robust results.

Kotturu and Kumar [17] used a multi-correlation strategy to compare several ML classifiers for AD early diagnosis. Their study evaluated models including SVM, RF, and LR using clinical datasets. Their findings showed that SVM performed better in Grueso and Viejo-Sobera's study [18], where they looked at various ML models, including RF, XGBoost, and SVM, to identify the development from Mild Cognitive Impairment to Alzheimer's. Using datasets like ADNI, they concluded that ensemble models and hybrid approaches yield better results than single classifiers.

Harika et al. [19] used the UCI Alzheimer's dataset to conduct comparison research utilizing various ML algorithms, including RF, LR, DT, and SVM. Their results favored ensemble methods for better accuracy and computational efficiency.

Javeed et al. [20] introduced the Feature Extraction Battery and optimized SVM for early dementia detection using a custom dataset. Their approach improved accuracy through tailored feature engineering and classifier optimization.

Kavitha et al. [21] used several ML models trained on the ADNI dataset to predict the onset of AD in its early stages. The study emphasized model explainability and used classifiers such as RF, LR, and KNN, reporting high accuracy.

Speiser et al. [22] explored different RF-based feature selection techniques for classification tasks. Even though their study did not focus on AD, the methods they discussed—especially recursive feature elimination and permutation importance—can contribute to the development of better AD diagnosis ML models. This study supports the vital role of feature selection in enhancing model accuracy.

Wang et al. [23] introduced a diagnostic ML tool that uses non-imaging features to classify dementia cases. Their method utilized datasets such as ADNI, implementing classifiers including SVM, RF, and XGBoost. The study achieved high predictive accuracy and demonstrated the feasibility of non-imaging biomarkers in AD diagnosis.

Zapata-Paulini and Cabanillas-Carbonell [24] conducted a study comparing ML models that can be applied to research on AD despite their interest in Parkinson’s disease. Their evaluation included DT, KNN, and AdaBoost, demonstrating that ensemble learning methods like AdaBoost outperform simple classifiers in medical diagnosis contexts.

Rajayyan and Mustafa [25] suggested a ML model that optimizes Alzheimer's prediction using the Cuckoo Search Algorithm. The study applied classifiers such as DT and LR, achieving improved accuracy through optimization.

Sarica et al. [26] analyzed neuroimaging data for Alzheimer's diagnosis using Random Forest (RF) algorithms. The review found RF to be especially effective in handling high-dimensional imaging datasets.
Bin-Hezam and Ward [27] proposed a novel ML-based framework for dementia detection using modifiable risk factors. Their model integrated multiple classifiers and emphasized interpretability, making it a significant contribution to the domain of explicable AI in Alzheimer's research.
Uddin et al. [28] presented a new ML-based strategy for early AD detection utilizing classifiers such as RF, XGBoost, AdaBoost, and SVM. The study used the OASIS dataset and emphasized model interpretability using explainable AI techniques.

2.2 Comparative Study of AD Prediction Applying ML

By identifying AD at an early stage and offering individualized treatment programs, ML may improve patient outcomes. A comparison of ML classifiers for AD prediction is presented below. Table 1 provides a summary of the findings from this comparison study.

Table 1: Comparative study of ML classifiers for AD prediction

Ref. No.	Year	Authors	Name of datasets used	ML algorithms used	No. of classifiers	Accuracy achieved
[15]	2021	Buyrukoglu	Clinical dataset	RF, DT, KNN	3	88.5%
[16]	2023	Dhakal et al.	Clinical dataset	RF, SVM, KNN	3	87.6%
[17]	2020	Kotturu & Kumar	Not specified	RF, SVM, LR	3	85.2%
[18]	2021	Grueso & Viejo-Sobera	ADNI	RF, XGBoost, SVM	3	91.2%
[19]	2022	Harika et al.	UCI Alzheimer's dataset	RF, LR, DT, SVM	4	86.3%
[20]	2023	Javeed et al.	Custom dataset	Optimized SVM	1	93.1%
[21]	2022	Kavitha et al.	ADNI	RF, KNN, LR	3	89.4%
[22]	2019	Speiser et al.	Simulated Classification Data	RF (various feature selection methods)	1	83.4%
[23]	2022	Wang et al.	ADNI (non-imaging features)	RF, SVM, XGBoost	3	88.7%
[24]	2024	Zapata-Paulini & Cabanillas-Carbonell	Custom dataset (Parkinson's)	DT, KNN, AdaBoost	3	87.5%
[25]	2023	Rajayyan & Mustafa	Not specified	DT, LR (w/ Cuckoo Algorithm)	2	90.2%
[26]	2017	Sarica et al.	Neuroimaging datasets	Random Forest	1	89.5%
[27]	2019	Bin-Hezam & Ward	Modifiable Risk Factors Dataset	RF, SVM, DT, LR	4	89.1%
[28]	2023	Uddin et al.	OASIS	RF, SVM, XGBoost, AdaBoost	4	92.0%
	2025	Hosam El-Sofany et al.	Private and Combined Datasets	Logistic Regression, Random Forest, AdaBoost, SVM, Decision Tree, XGBoost, and KNN	7	97.32%

3. ML Classification Algorithms for AD Prediction

Because of their superior pattern recognition capabilities in analyzing large datasets, ML classification approaches have found extensive application in AD prediction. Several kinds of data are utilized by these methods, including neuroimaging, genetic information, and clinical records, to develop prediction models that may aid in the early diagnosis and treatment of diseases. In this part, we will examine how ML classifiers can improve the accuracy of AD predictions, specifically the seven ML classifiers employed in this work.

3.1 Predicting the Development of AD Using ML Classification Methods

When it comes to diagnosing diseases like Alzheimer's, where early detection is paramount, ML classification algorithms have emerged as vital healthcare tools. These algorithms are capable of handling massive amounts of data extracting useful features, and classifying individuals as "Alzheimer's" or "non-Alzheimer's" with high accuracy. Medical research frequently uses imbalanced and high-dimensional data; however, recent studies have shown that ML classifiers can handle these types of datasets effectively [29].
The application of ML classification techniques in AD prediction has undergone substantial change over the years. Early studies focused on single-modal data, such as neuroimaging or clinical records, while recent research emphasizes the integration of multimodal data to improve prediction accuracy. For instance, combining neuroimaging data with genetic and clinical information has been shown to improve ML models' effectiveness [30]. Also, these models are now better for clinical use because they are more reliable and easier to understand, thanks to the use of ensemble methods and explainable AI techniques.

3.2 Contribution of ML Classifiers and SHAP Method in AD Prediction

- *Random Forest*: As an ensemble learning method, RF does an excellent task of avoiding overfitting when dealing with high-dimensional data by training a large number of decision trees and then producing the class modes for classification problems. In AD prediction, RF has been used to analyze neuroimaging and clinical data, achieving accuracies of up to 92.5%. Being able to manage missing data and its barrier to noise makes it an attractive option for medical applications [31].
- *Logistic Regression*: Logistic regression is a statistical method for issues involving binary classification. Based on a set of predictor variables, it estimates the likelihood of a positive or negative result. While simpler than other classifiers, logistic regression has been effective in predicting AD using clinical and genetic data, achieving accuracies of up to 87.6% [32]. Its interpretability and ease of implementation make it a valuable tool for preliminary analysis.
- *Support Vector Machine* (SVM): SVMs are effective classifiers because they locate the best hyperplane to divide data points into their respective classes. Its application in AD prediction has been substantial and performs exceptionally well in high-dimensional areas. Studies have reported accuracies of up to 91.8% using SVM on neuroimaging data from the ADNI dataset [33]. SVM's ability to handle nonlinear correlations through kernel functions makes it a versatile tool for medical diagnosis.
- *K-Nearest Neighbors* (KNN): One nonparametric approach is KNN, which sorts data points taking into consideration the most well-liked subject among their KNN. It is simple to implement and effective for small datasets. In AD prediction, KNN has achieved accuracies of up to 89.3% on private datasets [34]. Nevertheless, the "curse

of dimensionality" can cause its performance to decrease when dealing with high-dimensional data.

- *Decision Tree*: One kind of tree-based model is the decision tree, which employs feature values to divide the dataset. For exploratory analyses, it is ideal because it is simple to understand and depict. Decision tree has achieved 88.5% accuracy in AD prediction using publicly available datasets [35]. However, it is prone to overfitting, which can be mitigated using ensemble techniques such as RF.
- *AdaBoost*: AdaBoost is an ensemble technique that combines multiple underperforming classifiers into an effective one. It focuses on correcting errors made by previous classifiers, making it effective for imbalanced datasets. In AD prediction, AdaBoost has achieved accuracies of up to 90.6% on open-access datasets [36]. Its ability to improve prediction accuracy through iterative learning makes it a valuable tool for medical applications.
- *XGBoost*: XGBoost is a sophisticated version of gradient boosting recognized for its rapidity and efficacy. Its state-of-the-art accuracies on combined datasets reach up to 96.7%, making it widely utilized in AD prediction [37]. XGBoost stands out among medical diagnosis classifiers due to its scalability, parallel processing capability, and capacity to deal with missing data.
- *Integration of Explainable AI*: In this study, we used the SHAP method as an explainable AI method to make ML-based models for predicting AD easier to understand. Breaking down the model's predictions by feature is one way that SHAP helps doctors understand how it came to its outcomes [38]. In the healthcare field, such detail is crucial because AI-based diagnostic tools depend on trust and transparency.

4. The Proposed AD Model

This section introduces the proposed ML-based system for predicting AD, leveraging data classification techniques and explainable AI methods. The approach involves preprocessing the dataset, employing several ML algorithms, and then measuring how well they performed. Figure 1 illustrates the steps in the proposed system. Data collection and preprocessing to handle inconsistencies, including missing value handling (average replacement), is the first step. Then partitioning the dataset for training and testing takes place. To find the best model for AD prediction, several classification methods are used. To help medical practitioners establish accurate diagnoses, the suggested solution makes use of sophisticated ML methods and an explainable AI.

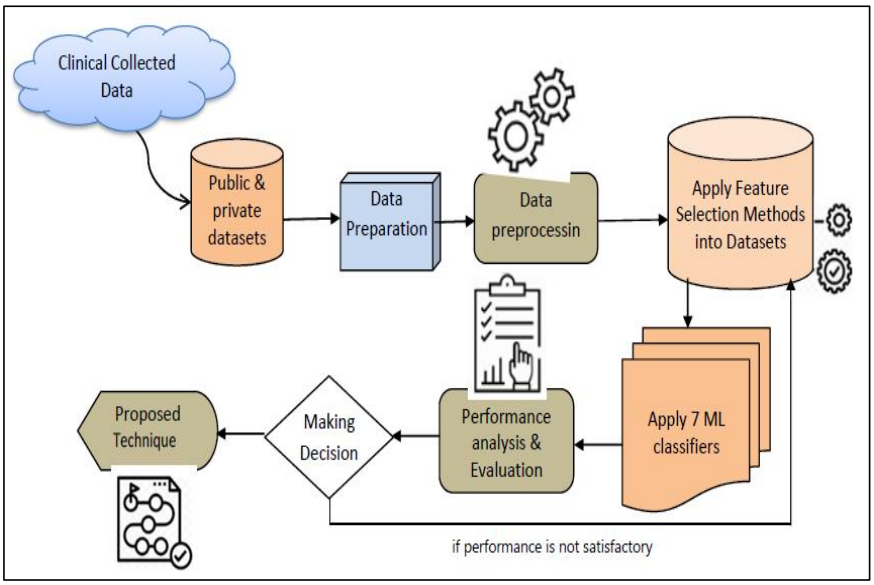


Figure 1: The proposed AD prediction model

4.1 Methodology

Seven ML algorithms—SVM, AdaBoost, logistic regression, decision tree, RF, XGBoost, and KNN—are utilized in the suggested method. These algorithms, chosen for their capacity to evaluate complex datasets and deliver reliable predictions, support doctors and data analysts in diagnosing AD. The study utilizes recent datasets, research articles, and published works on AD to guarantee that the suggested model is applicable and trustworthy. The methodology is structured into three stages: (1) data collection, (2) feature extraction, and (3) data exploration, as depicted in Figure 1. Data cleansing, normalization, and handling missing values are all part of data preprocessing [39]. The preprocessed data is then classified using the seven ML algorithms. F1-score, accuracy, specificity, precision, sensitivity, and AUC are the measures of the model's performance [40].

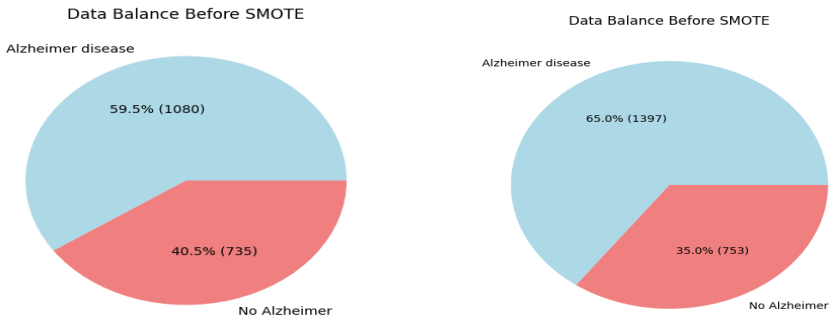
4.2 Datasets and Their Features

Using both public and private datasets, predicting the first signs of AD is the goal of this study. With identical attributes, the public dataset has 1,815 samples, and the private dataset has 335 samples. There are 2150 cases in the combined dataset, and each record has 33 features. These features include *diet quality*, *age*, *gende*r, *family status*, *ethnicity*, *education level*, *smoking*, *alcohol use*, *physical activity*, *sleep quality*, *body mass index (BMI*), and medical history: AD, cardiovascular disease, diabetes, depression, head injury, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, hypertension, diastolic blood pressure, total cholesterol, systolic blood pressure, lipid triglycerides, mini-mental state exam, assessment of functional abilities, complaints of memory loss or cognitive impairment, difficulties with activities of daily living, personality changes, inability to complete tasks at hand, and forgetfulness. "Diagnosis," the target variable, is a yes/no indicator of whether AD is present or not. In Figure 2, you can see the breakdown of the percentage of people with Alzheimer's in both datasets. Figure 2(a) shows that 59.59% of

the 1815 records in the public dataset have a diagnosis of AD, while the remaining 40.5% do not. Alternatively, as can be seen in Figure 2(b), out of 2150 records in the combined dataset, 65% have an AD diagnosis, whereas 35% do not [41].

4.3 Datasets Preparation

We subject the dataset to thorough preprocessing to ensure it is clean, balanced, and ML modelable. Removal of unnecessary columns like *PatientNo* and *DoctorInCharge* and replacement of binary class labels with descriptive ones ("Alzheimer disease" and "No Alzheimer") are part of the preprocessing steps. We use the mean strategy to impute missing values in the numerical columns after separating them from the nonnumerical columns. The process of label encoding is used to convert categorical variables into a numerical format. After that, we take the dataset and extract the features (X) and the dependent variable (y). We applied the SMOTE method to address the class imbalance. To make sure methods like logistic regression would work as expected, we used the StandardScaler package to scale the features after splitting the data into a training set and a testing set [42].
We assess the suggested approach using a combined and private dataset, with 75% of the dataset for training and 25% for testing. We employ the MI method to measure feature interdependence, with higher values indicating greater dependency. For instance, the *MemoryComplaints* feature has the highest importance score of 22.36%, while *Cholesterol_LDL* has the lowest at 1.4%, as shown in Figure 3 [43].



(a) Public dataset (b) Combined dataset

Figure 2: The percentage distribution of Alzheimer's cases among participants.

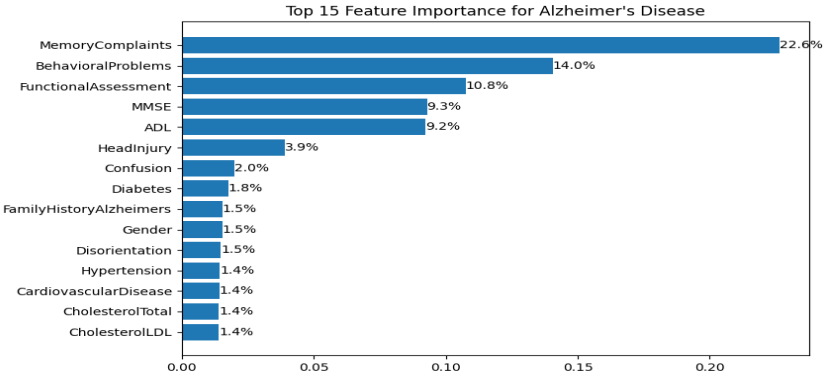


Figure 3: Features importance in the breast cancer dataset.

4.4 The Feature Selection Process

The three main techniques used for feature selection are *ANOVA*, *chi-square*, and *MI*. Each method identifies the most relevant features for predicting AD. ANOVA checks how important features are by looking at the differences between groups, which helps find features that are noticeably different in Alzheimer's and non-Alzheimer's cases. The chi-square test measures the independence of categorical features, highlighting those with strong associations with the target variable [44]. To prioritize features that offer the most information for prediction, the MI metric assesses the degree to which features are dependent on the target [45]. We achieve accurate and interpretable predictions by combining various strategies to identify the best attributes.

4.4.1 Analysis of Variance (ANOVA)

One statistical method for determining the significance of variations across groups is ANOVA (e.g., AD vs. no AD). We compute the F-statistic in the analysis of variance by using Eq. (1):

$$F = \frac{Between_\,Group\,Variability}{Within_\,Group\,Variability} = \frac{SS_B/(k-1)}{SS_W/(N-k)} \qquad (1)$$

Where:

SS_W = sum of squares within groups (measures the variability within each class),
N = total number of samples,
SS_B = sum of squares between groups (measures variability between classes),
k = number of groups (classes),
Between-group variability **(**SS_B**)** measures how much the means of the feature differ across classes. A higher value indicates greater differences between groups.
Within-group variability **(**SS_W**)** measures the variability of the feature within each class. A lower value indicates more homogeneity within groups.
F-statistic shows that if the F-value is high, then the trait is very significant for class distinction.

A higher F-value indicates that the features are more significant because they exhibit significant differences across classes, making them useful for classification tasks.

4.4.2 *Chi-Square Test*

To determine if categorical traits are independent of the dependent variable, the chi-square test is employed. It measures the association between the feature and the target to find features having a strong correlation with the result. We compute the chi-square statistic (X^2) by using Eq. (2):

$$X^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} \qquad (2)$$

Where:

O_i = the number of occurrences a feature appears in each group,
E_i = expected frequency of the feature in each category (under the assumption of independence),
Observed frequency **(O_i)** is the actual count of samples in each category.
Expected frequency **(E_i)** is the count of samples expected in each category if the feature and target are independent.
Chi-square statistic (**χ^2**) shows that feature association with the target becomes stronger as the value increases.

Features with higher chi-square values are selected because they have a stronger relationship with the target variable, making them more useful for prediction.

4.4.3 *Mutual Information (MI)*

Two variables' degrees of dependence on one another can be measured using MI by measuring how much information one variable provides about the other. Where it succeeds is in detecting target-feature nonlinear relationships. We compute the MI of two variables by using Eq. (3):

$$I(X;Y) = \sum_{x \epsilon X} \sum_{y \epsilon Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \qquad (3)$$

Where:

MI *((I(X; Y))* shows that the value increases as the degree of dependence between X and Y grows stronger. The independence of X and Y is represented by I(X; Y)=0.
Joint probability (**$p(x, y)$**) is the probability of *X* and *Y* occurring together.
$p(x)$ and $p(y)$ = the corresponding marginal distributions of probability for X and Y.
$p(x, y)$ = an analysis of the joint X and Y probability distribution.
Their marginal probabilities are denoted as *p(x)* and *p(y)*, respectively, where X and Y are considered independent variables.

A higher MI score indicates that the feature provides additional information about the target variable so that the model's predictions are more accurate. Therefore, features with higher scores are preferred.

5. Experimental Results and Analysis

The outcomes of the tests and evaluations conducted on the proposed ML method for AD prediction are detailed in this section. The implementation of the model was carried out using *Jupyter Notebook*, which provides an interactive environment for data visualization, live coding, and documentation. The dataset, comprising both public and private sources, was initially cleaned using Python's *Pandas* and *NumPy* libraries to handle missing values and inconsistencies. Following this, the data was preprocessed using the *StandardScaler* method from the *Scikit-learn* module to normalize the features and ensure uniformity. To

determine which traits were most important for AD prediction, three separate methods were used: MI, chi-square, and ANOVA.
Consequently, training was done on 75% of the dataset while testing was done on 25%. This section ensures an accurate assessment of the model's efficacy. For the final predictions, among seven distinct ML classifiers, we utilized the model with the highest performance following training and testing. The subsections that follow present a comprehensive evaluation of the experiment outcomes.

5.1 Performance Evaluation

Here, we showcase the results of evaluating the suggested method with various ML classifiers. Using the confusion matrix's true negatives (TN), false positives (FP), specificity, precision, accuracy, and AUC, we assess the system's performance. The confusion matrix provides a comprehensive summary of the model's predictions, as seen in Figure 4. Based on this research, the XGBoost classifier using SMOTE was able to accurately identify 253 TFPCs, while making 9 FPPCs and missing 6 FNPCs, and correctly identified 291 TN cases, showing that it works well for classifying AD using the public dataset. Figure 5 highlights the correlation between important features, emphasizing the linear relationships among variables with at least a 90% correlation. The correlation matrix provides a clear and understandable summary of the relationships between the top 10 features in the dataset. By identifying strong correlations, it helps reduce redundancy, improve model performance, and gain insights into the underlying biology and clinical manifestations of AD. The heatmap visualization makes it easy to interpret and communicate these findings to stakeholders, including clinicians and researchers.

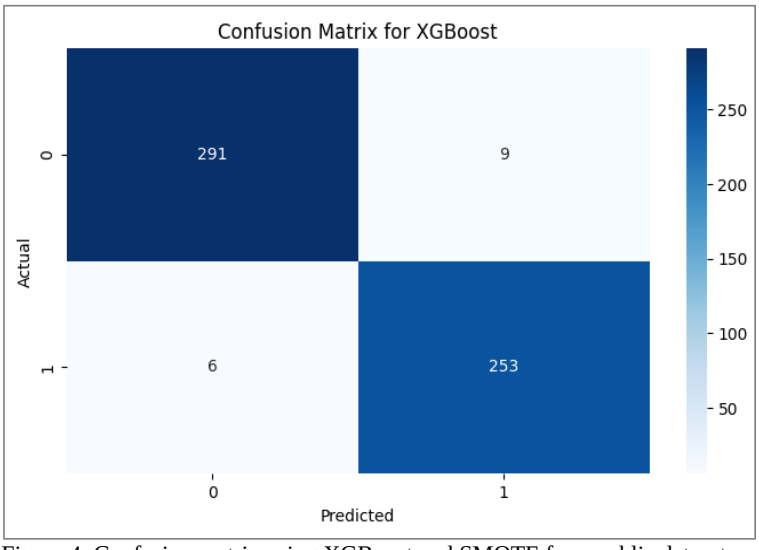


Figure 4: Confusion matrix using XGBoost and SMOTE for a public dataset.

ISSN: 0868-4952

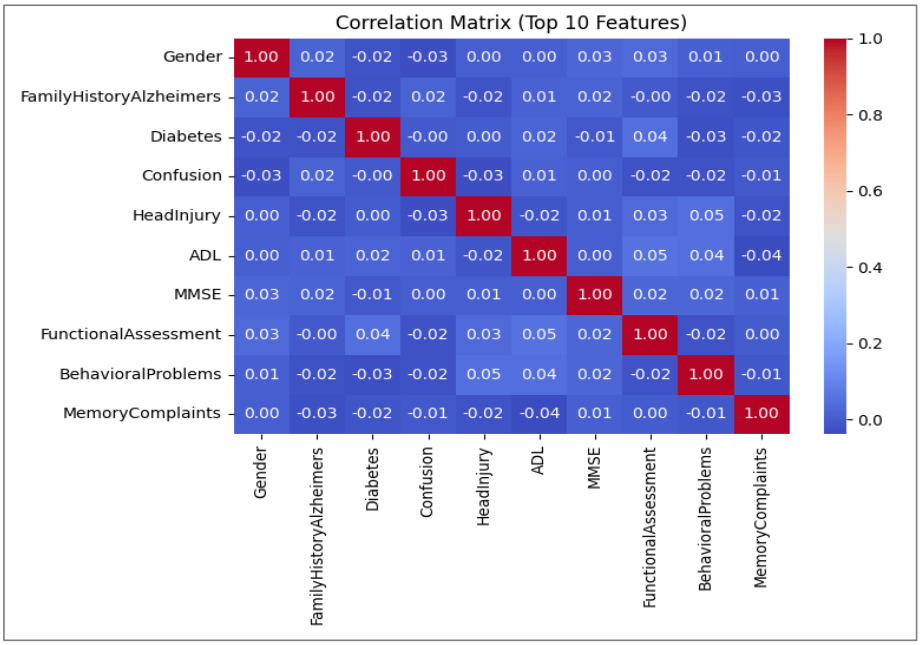


Figure 5: Correlation between the first ten important features of a public dataset.

Additionally, Figures 6 and 7 present scatter and density plots for four selected features, offering insights into their distributions, correlations, and potential outliers.

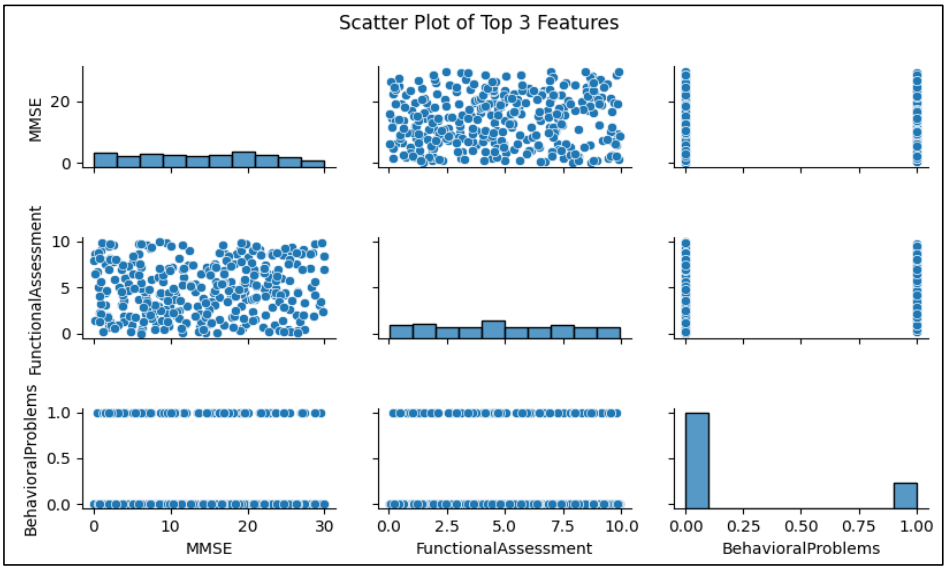


Figure 6: Scatter plot among the first three features.

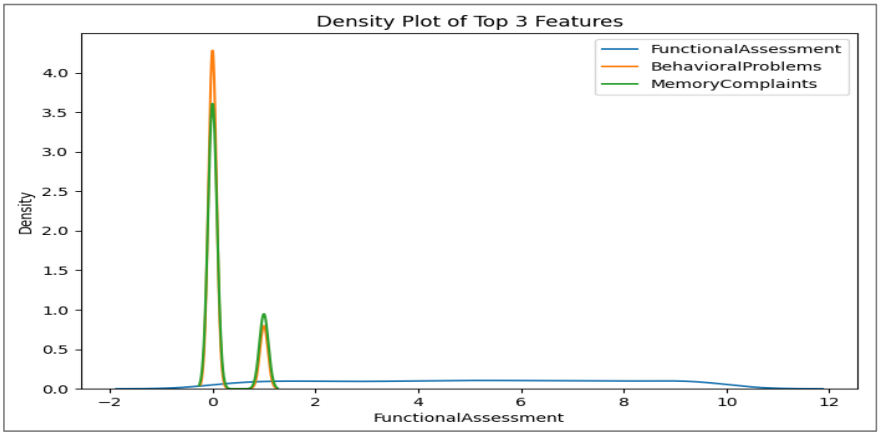


Figure 7: Plot of density for the first three features.

The performance metrics are calculated as follows:

- **Accuracy** determines the percentage of accurately labeled samples. We compute the accuracy by using the formula of Eq. (4):

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (4)$$

- **Specificity** determines the exact proportion of false negatives that were identified. We compute the specificity by using the formula of Eq. (5):

$$\text{Specificity} = 1 - \left(\frac{F_P}{F_P + T_N}\right) \quad (5)$$

- **Sensitivity (Recall)** indicates the proportion of correctly identified positive cases. We compute the specificity by using the formula of Eq. (6):

$$\text{Sensitivity} = \frac{T_P}{T_P + F_N} \quad (6)$$

- **F1-Score** unifies recall and precision into a single measure. We compute the **F1-score** using the formula of Eq. (7):

$$F - \text{Score} = \frac{2 * (\text{precision} * \text{recall})}{\text{precision} + \text{recall}} \quad (7)$$

- **Precision** figures out how accurate positive predictions are. We compute the precision using the formula of Eq. (8):

$$\text{Precision} = \frac{T_P}{T_P + F_P} \quad (8)$$

In comparison to other algorithms, the most effective classifier was the XGBoost model with a 97.32% accuracy, 97.68% sensitivity, 97.00% specificity, 96.56% precision, 97.12% F1-score, and 98.43% AUC, using the combined dataset. Using the public dataset, XGBoost achieved an accuracy of 97.23%, specificity of 96.52%, precision of 96.14%, sensitivity of 98.03%, AUC of 98.30%, and F1-score of 97.07%. Furthermore, XGBoost did exceptionally well on the private dataset with 95.83% accuracy, 93.94% sensitivity, 96.88% precision, 97.44% specificity, 98.52% AUC, and 95.38% F1-score.
Table 2 summarizes these findings and compares the seven algorithms' performance using the combined dataset with SMOTE. Figure 8 shows the accuracy results of the seven ML classifiers using a combined dataset with SMOTE.

Table 2: Performance metrics for seven ML classifiers using a combined dataset with SMOTE.

Ser.	Classifier	Accuracy	Sensitivity	Specificity	Precision	F1-score
1	Random Forest	96.96	96.91	97.00	96.54	96.72
2	Logistic Regression	84.08	83.78	84.33	82.20	82.98
3	Support Vector Machine	61.72	62.16	61.33	58.12	60.07
4	Decision Tree	91.06	89.58	92.33	90.98	90.27
5	K-Nearest Neighbors	61.90	45.56	76.00	62.11	52.56
6	AdaBoost	94.81	92.28	97.00	96.37	94.28
7	XGBoost	97.32	97.68	97.00	96.56	97.12

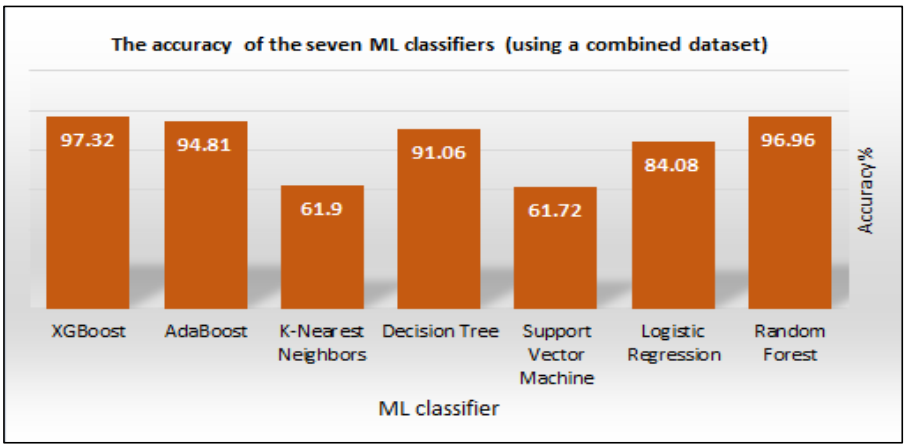


Figure 8: Seven ML classifiers' accuracy results on a combined dataset using SMOTE.

In this section, we offer a thorough assessment of how well each classifier performed on the combined dataset. The results are as follows:

- *Random Forest* finished with an F1-score of 96.72%, a precision of 95.54%, a specificity of 97%, and a sensitivity of 96.91%. The confusion matrix revealed a case count of 252 TP, 289 TN, 11 FP, and 7 FN.
- *Logistic Regression* ensured accuracy by 84.08%, F1-score of 82.98%, specificity of 84.33%, precision of 82.20%, and sensitivity of 83.78%. The confusion matrix included 217 TP, 253 TN, 47 FP, and 42 FN cases.
- *Support Vector Machine* recorded an accuracy of 61.72%, precision of 58.12%, sensitivity of 62.16%, F1-score of 60.07%, and specificity of 61.33%. The confusion matrix displayed 161 TP, 184 TN, 116 FP, and 98 FN cases.
- *Decision Tree* attained an accuracy of 91.06%, an F1-score of 90.27%, a precision of 90.98%, a sensitivity of 89.58%, and a specificity of 92.33%. By analyzing the confusion matrix, we found 229 TP, 271 TN, 29 FP, and 30 FN cases.
- *K-Nearest Neighbors* ensured an accuracy of 61.90%, specificity of 76.00%, sensitivity of 45.56%, F1-score of 52.56%, and precision of 62.11%. The confusion matrix included 118 TP, 228 TN, 72 FP, and 141 FN cases.

- *AdaBoost* reached a 94.81% accuracy rate, 92.28% sensitivity, 97.00% specificity, 96.37% precision, and 94.28% F1-score. There were 239 TP, 291 TN, 9 FP, and 20 FN examples listed in the confusion matrix.
- *XGBoost* accomplished the highest accuracy of 97.32%, sensitivity of 97.68%, F1-score of 97.12%, precision of 96.56%, and specificity of 97% of any classifier tested. Among the cases shown in the confusion matrix, there were 253 TP, 291 TN, 9 FP, and 6 FN.

Afterward, we trained the ML-based model and evaluated it on the other datasets using a domain adaptation strategy. Table 3 summarizes these findings and compares the seven algorithms' performance using the private dataset (consisting of 335 cases). Figure 9 shows the accuracy results of the seven ML classifiers using the private dataset with SMOTE. The XGBoost algorithm demonstrated the highest accuracy and overall performance; thus, it is a superior model for AD prediction.

Table 3: Performance metrics for seven ML classifiers using a private dataset with SMOTE.

Ser.	Classifier	Accuracy	Sensitivity	Specificity	Precision	F1-Score
1	Random Forest	88.89	87.88	89.74	87.88	87.88
2	Logistic Regression	81.94	87.88	76.92	76.32	81.69
3	Support Vector Machine	44.44	69.70	23.08	43.40	53.49
4	Decision Tree	91.67	93.94	89.74	88.57	91.18
5	K-Nearest Neighbors	48.61	45.45	51.28	44.12	44.78
6	AdaBoost	91.67	87.88	94.87	93.55	90.62
7	XGBoost	95.83	93.94	97.44	96.88	95.38

5.2 Discussion

This study utilized many types of ML classifiers applied to public, private, and combined datasets to detect the first signs of Alzheimer's. The dataset has undergone preprocessing and data partitioning into 25% for testing and 75% for training. The metrics utilized for evaluation included accuracy, F1-score, sensitivity, precision, specificity, and AUC. We performed feature selection using ANOVA, MI, and chi-square to identify the most useful attributes for making predictions. It was clear from the trials that XGBoost achieved the highest level of accuracy, at 97.32%, followed by RF with 96.96%. The XGBoost classifier also demonstrated superior sensitivity of 93.94%, specificity of 97.44%, precision of 96.88%, and F1-score of 95.38%.

The SHAP approach was incorporated to improve the model's interpretability. SHAP values provided insights into the contribution of individual features, such as memory complaints, behavioral problems, and age. This method enhanced the transparency of the model and made it easier to use in healthcare settings for healthcare practitioners by giving them practical information. Figure 10 presents a SHAP summary plot, highlighting the importance of every attribute used for making predictions.

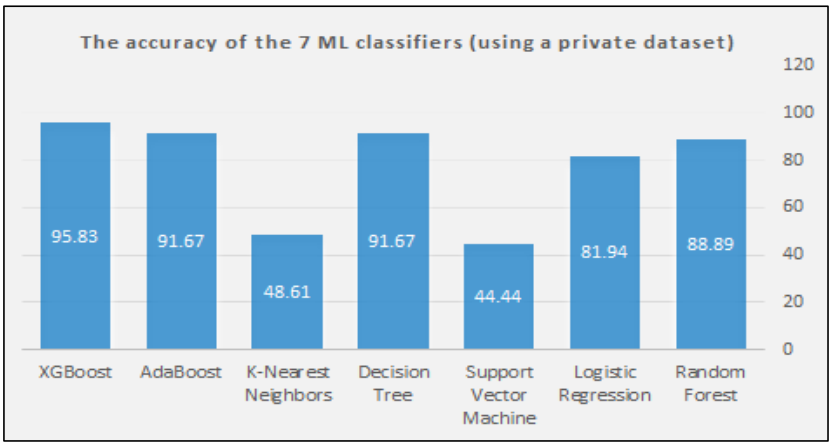


Figure 9: The accuracy results of the seven ML classifiers using a private dataset with SMOTE.

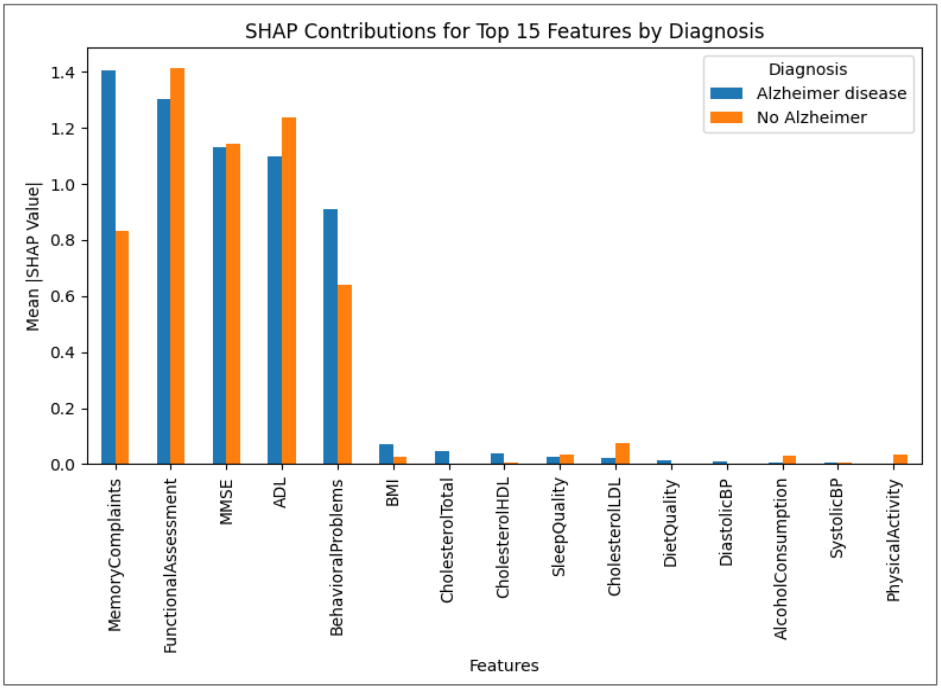


Figure 10: AI explanation with SHAP values for feature importance.

In this research, the SHAP methodology was incorporated into the proposed framework to create an integrated way to evaluate how each feature affected the model's predictions of AD. This integration significantly improves our ML method's clarity and interpretability. SHAP explains in detail how different parameters affect the results produced by the model,

rendering it an invaluable resource for medical professionals and academics alike. Below are the key ways that SHAP improves our model.

1. *Analysis of Individual Predictions*: SHAP allows for the examination of individual predictions, providing insights into the reasoning behind specific patient outcomes. It is very important in healthcare settings to have this ability, where understanding the basis of a prediction can inform further diagnostic procedures or treatment plans, ensuring more personalized and effective patient care.
2. *Explanation of Feature Contributions*: The SHAP values provide a comprehensive analysis of the features' effects on the predictions of the model. To illustrate the point, consider the setting of AD prediction. SHAP can clarify the role of features such as memory complaints, behavioral issues, age, gender, and cardiovascular conditions in shaping the final prediction. This allows medical professionals to concentrate on the factors that significantly affect the predictions of the model.
3. *Insights into Global Model Behavior*: Beyond explaining individual predictions, SHAP offers a broader perspective on how well the model works across all datasets. To provide an entire overview of the decision-making procedure of the model, it shows which features are typically the most important and how they interact with each other.
4. *Building Trust and Facilitating Adoption in Healthcare*: To effectively apply AI models in clinical practice, their decision-making processes must be clear and understandable. SHAP values provide this clarity by validating and giving more medical professionals confidence in the system by clarifying the predictions of the model. ML models must be transparent to foster trust in real-world healthcare settings.

In the end, the best model for predicting the first symptoms of AD was the XGBoost algorithm, plus SMOTE for managing imbalanced data. By accomplishing its goals, the study set the basis for future research in the field and proved that ML may be useful in early disease identification.

6. Study Limitations

Despite the hopeful results and potential utility of the proposed method for creating algorithms to anticipate AD using ML, it is important to highlight certain limitations to provide an accurate assessment of the study's findings:

1. *The Availability and Quality of Datasets*: The availability and quality of training and testing datasets determine the efficacy and consistency of ML models. For this study, we consulted a variety of public and private datasets. However, these datasets may have limitations in terms of availability, representativeness, and overall quality. Such constraints could hinder the generalizability of the proposed approach to broader populations or datasets from diverse sources, possibly reducing its usefulness in practical situations.
2. *Classes Imbalance*: ML classifiers can be negatively impacted by imbalanced classes when one class is far more numerous than the other. Researchers used SMOTE to solve this problem. While SMOTE is an improvement over other methods, it does help bring about more equitable class distributions by providing minority groups with synthetic samples. The synthetic samples may not fully capture the underlying data distribution, potentially introducing bias and reducing the accuracy of predictions when applied to real-world datasets.
5. *Algorithm Selection*: Finding the best model to predict the beginning of AD required investigating a wide variety of ML algorithms, which the study did. Nevertheless, the results might have been affected by the subjective nature of algorithm selection. There may be alternative algorithms out there with better accuracy or different trade-offs that

were not analyzed in this work. Future research should carefully evaluate a wider range of ML methods to select the optimal algorithm for this task.

6. *Domain Adaptation*: The research proved that the suggested system could be adjusted using domain adaptation methods. Despite this, the approach may still face limitations when applied to diverse populations or environments. Factors such as varying demographics, lifestyles, and healthcare systems could impact the model's performance. Further research is necessary to determine how effective the suggested method is in other populations and to tackle possible difficulties related to domain adaptation.
7. *Feature Selection Bias*: Improving model performance relies heavily on feature selection, which, if not handled properly, might generate bias. To find important features, the research used statistical approaches like ANOVA, MI, and chi-square. However, these methods may overlook complex interactions between features or fail to capture nonlinear relationships, potentially leading to suboptimal feature sets. Future work could explore more advanced feature selection techniques or incorporate domain knowledge to ensure a more comprehensive and unbiased selection process.
8. *Privacy and Ethical Considerations*: Concerns about privacy and ethics arise when dealing with sensitive health information, such as patient records. Ensuring data anonymization and compliance with regulations like GDPR or HIPAA is essential but can be challenging. Additionally, biases in the dataset, such as underrepresentation of certain demographic groups, could lead to unfair or inequitable predictions. For healthcare organizations to responsibly implement ML models, it is essential to address these privacy and ethical concerns.
9. *Generalization to Real-World Scenarios*: The study's results are based on controlled experiments using specific datasets, of which reality might not be a perfect reflection. Factors such as missing data, noisy measurements, and variations in data collection methods across healthcare institutions could affect the model's performance when deployed in practice. Robust testing on diverse, real-world datasets is necessary to validate the model's generalizability and reliability.
10. *Computational Resource Requirements*: Training and optimizing ML models, particularly complex algorithms like XGBoost, can be computationally intensive. This may limit the feasibility of deploying such models in resource-constrained environments, such as small clinics or developing regions. Developing lightweight and efficient versions of the model without compromising performance is an area that warrants further investigation.
11. *Coordinating with Clinical Procedures*: For clinical practice to make good use of the model, it must be seamlessly integrated into existing healthcare workflows. This requires collaboration with healthcare providers to ensure the model's outputs are actionable and align with clinical decision-making processes. Challenges such as user interface design, interoperability with electronic health records (EHRs), and clinician training must be addressed to facilitate adoption.

7. Conclusions and Future Work

This work suggested a ML-based approach to AD prediction using advanced data classification methods and an explainable AI framework. The research aimed to address the growing need for early and accurate diagnoses of AD, especially in countries where the illness is spreading at an alarming rate, such as Saudi Arabia and Egypt. The research looked at seven different ML classifiers to find the best one for predicting when AD will start. These models included SVM, AdaBoost, XGBoost, logistic regression, RF, decision

tree, and KNN. The datasets used in the evaluation were a mix of public and private. To resolve the issue of unbalanced data and ensure the model's strong performance, SMOTE was used. The XGBoost classifier was shown to be the most effective algorithm according to the experimental results, achieving optimal performance metrics on both the combined and private datasets. The XGBoost classifier had an accuracy of 97.32%, precision of 96.56%, sensitivity of 97.68%, specificity of 97.00%, AUC of 98.43%, and F1-score of 97.12%. Using the public dataset, XGBoost achieved an accuracy of 97.23%, precision of 96.14%, AUC of 98.30%, specificity of 96.52%, sensitivity of 98.03%, and F1-score of 97.07%. Furthermore, XGBoost did exceptionally well on the private dataset with 95.83% accuracy, 93.94% sensitivity, 96.88% precision, 97.44% specificity, 98.52% AUC, and 95.38% F1-score. These findings demonstrate that XGBoost is capable of efficiently processing large datasets and may be useful for the accurate prediction of AD. To facilitate a deeper understanding of the model, the study integrated SHAP methodologies, shedding light on how specific attributes affect the model's predictions. By providing healthcare providers with practical insights, this explainable AI framework increased the model's transparency and paved the way for its implementation in clinical settings. The SHAP analysis revealed that features such as memory complaints, behavioral problems, and age were among the most influential in predicting AD, aligning with clinical knowledge and increasing confidence in the model's predictions. The proposed technique offers significant potential for early-stage AD prediction, offering a healthcare application solution that is both efficient and cost-effective. By combining high-performance ML models with understandable AI, this research helps advance methods for diagnosing AD that are both accurate and easy to understand, ultimately supporting early intervention strategies and improving patient outcomes.

Despite the promising results of the proposed approach, we have identified several areas for further research and improvement.

1. *Dataset Expansion and Diversity*: More extensive and varied databases should be the goal of future research, including data from different geographic regions and demographic groups. This will enhance the model's capacity to apply across different populations and its generalizability.
2. *Longitudinal Data Integration*: Improved knowledge of AD evolution and more exact predictions may be possible with the use of longitudinal data that follows patients over time. This would allow the model to capture temporal dynamics and changes in patient conditions.
3. *Advanced Feature Selection Techniques*: To further improve the model's performance, the study used ANOVA, MI, and chi-square for feature selection. Additional advanced strategies could be explored in future studies, such as feature extraction based on deep learning or domain-specific feature engineering.
4. *Hybrid Models*: Combining multiple ML algorithms or integrating deep learning techniques with traditional ML models could lead to improved prediction accuracy. Hybrid models that leverage the strengths of different algorithms may offer better performance in complex datasets.
5. *Real-World Deployment*: Deploying the suggested model to real healthcare institutions to evaluate its efficacy in practical settings should be the focus of future studies. This includes addressing challenges related to integration with electronic health records, user interface design, and clinician training.
6. Ethical and Privacy Considerations: As the use of sensitive health data raises ethical and privacy concerns, future work should prioritize ensuring data anonymization and compliance with regulations such as GDPR and HIPAA. Additionally, addressing potential biases in the dataset to ensure fair and equitable predictions is crucial.

7. *Explainable AI Enhancements*: Even though SHAP is a great tool for understanding feature contributions, future research could explore additional explainable AI techniques to further enhance model interpretability. This includes developing user-friendly visualization tools and integrating domain-specific knowledge into the explainability framework.
8. *Computational Efficiency*: Optimizing the model's computational efficiency, especially for resource-constrained environments, is an important area for future work. Developing lightweight versions of the model without compromising performance will facilitate its adoption in smaller clinics and developing regions.

Finally, we established the framework for future research into AD prediction utilizing explainable AI and ML. Improving patient outcomes and reducing the global impact of AD can be achieved through the proposed method's emphasis on early detection and intervention. To make sure the model is widely used and has an impact on healthcare, future studies should build on these findings to make it even better.

Acknowledgments: The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through small group research under grant number (RGP1/15/46).

Conflicts of interest: The authors declare no conflicts of interest in this work.

Availability of data and materials: Upon request, each of the authors will make available the datasets used in their research.

References

[1] Alzheimer's Association (2023), "Alzheimer's disease facts and figures", *Alzheimer's & Dementia*, 19(3), 1-100. https://doi.org/10.1002/alz.13016.
[2] World Health Organization (2023), "Dementia: A public health priority", World Health Organization. https://www.who.int.
[3] Prince, M., Wimo, A., Guerchet, M., Ali, G.C., Wu, Y.T. and Prina, M. (2023), "World Alzheimer report 2023: The global impact of dementia", *Alzheimer's Disease International*. https://www.alzint.org/resource/world-alzheimer-report-2023/.
[4] Alanazi, H.O., Abdullah, A.H., Qureshi, K.N., 2017. "A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care", J. Med. Syst. 41 (4). https://doi.org/10.21203/RS.3.RS-624520/V1.
[5] Asada, T., 2017. Epidemiology of dementia in Japan. Neuroimag. Diagnos. Alzheimer's Diseas. Other Dementias 1–10. https://doi.org/10.1007/978-4-431-55133-1_1.
[6] Assegie, T.A., Tulasi, R.L., Kumar, N.K., 2021. Breast cancer prediction model with decision tree and adaptive boosting. IAES Int. J. Artif. Intell. 10 (1), 184. https://doi.org/10.11591/ijai.v10.i1.pp184-190.
[7] Bari Antor, M., et al., 2021. A comparative analysis of machine learning algorithms to predict Alzheimer's disease. J. Healthc. Eng. 2021. https://doi.org/10.1155/2021/9917919
[8] Biechele, G., et al. (2024), "Associations between sex, body mass index and the individual microglial response in Alzheimer's disease", *J. Neuroinflammation*, 21(1), 30. https://doi.org/10.1186/s12974-024-03020-y.

[9] Alshamlan, H., Alwassel, A., Banafa, A. and Alsaleem, L. (2024), "Improving Alzheimer's disease prediction with different machine learning approaches and feature selection techniques", *Diagnostics (Basel)*, 14(2237). https://doi.org/10.3390/diagnostics14192237.

[10] Chen, T. and Guestrin, C. (2022), "XGBoost: A scalable tree boosting system", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785.

[11] Lundberg, S. and Lee, S.I. (2013), "A unified approach to interpreting model predictions", *Advances in Neural Information Processing Systems*, 30, 4765-4774. https://doi.org/10.48550/arXiv.1705.07874.

[12] Ribeiro, M.T., Singh, S. and Guestrin, C. (2022), "Why should I trust you?: Explaining the predictions of any classifier", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144. https://doi.org/10.1145/2939672.2939778.

[13] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2022), "SMOTE: Synthetic minority over-sampling technique", *J. Artif. Intell. Res.*, 16, 321-357. https://doi.org/10.1613/jair.953.

[14] Borne, L., et al. (2024), "The interplay of age, gender and amyloid on brain and cognition in mid-life and older adults", *Sci. Rep.*, 14(1), 27207. https://doi.org/10.1038/s41598-024-78308-3.

[15] Buyrukoglu, S. (2021), "Early detection of Alzheimer's disease using data mining: Comparison of ensemble feature selection approaches", *Konya J. Eng. Sci.*, 9(1), 50-61. https://doi.org/10.36306/KONJES.731624.

[16] Dhakal, S., Azam, S., Hasib, K. Md., Karim, A. and Jonkman, M. (2023), "Dementia prediction using machine learning", *Procedia Comput. Sci.*, 219, 1297-1308. https://doi.org/10.1016/j.procs.2023.01.414.

[17] Kotturu, P.K. and Kumar, A. (2020), "Comparative study on machine learning models for early diagnose of Alzheimer's disease: Multi correlation method", *Proceedings of the International Conference on Communication and Electronics Systems (ICCES)*, 778-783. https://doi.org/10.1109/ICCES48766.2020.9137872.

[18] Grueso, S. and Viejo-Sobera, R. (2021), "Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: A systematic review", *Alzheimer's Res. Ther.*, 13(1), 1-29. https://doi.org/10.1186/s13195-021-00900-w.

[19] Harika, S., Yamini, T., Nagasaikamesh, T., Basha, S.H., Kumar, S.S. and DurgaKameswari, S.S. (2022), "Alzheimer's disease detection using different machine learning algorithms", *Int. J. Res. Appl. Sci. Eng. Technol.*, 10(10), 62-66. https://doi.org/10.22214/IJRASET.2022.46937.

[20] Javeed, A., et al. (2023), "Early prediction of dementia using feature extraction battery (FEB) and optimized support vector machine (SVM) for classification", *Biomedicines*, 11(2), 439. https://doi.org/10.3390/biomedicines11020439.

[21] Kavitha, C., Mani, V., Srividhya, S.R., Khalaf, O.I. and Tavera Romero, C.A. (2022), "Early-stage Alzheimer's disease prediction using machine learning models", *Front. Public Health*, 10, 853294. https://doi.org/10.3389/fpubh.2022.853294.

[22] Speiser, J.L., Miller, M.E., Tooze, J. and Ip, E. (2019), "A comparison of random forest variable selection methods for classification prediction modeling", *Expert Syst. Appl.*, 134, 93-101. https://doi.org/10.1016/j.eswa.2019.05.028.

[23] Wang, H., et al. (2022), "Develop a diagnostic tool for dementia using machine learning and non-imaging features", *Front. Aging Neurosci.*, 14, 945274. https://doi.org/10.3389/fnagi.2022.945274.

[24] Zapata-Paulini, J. and Cabanillas-Carbonell, M. (2024), "Evaluation of machine learning algorithms in the early detection of Parkinson's disease: A comparative study", *Indonesian J. Electr. Eng. Comput. Sci.*, 35(1), 222-237. https://doi.org/10.11591/ijeecs.v35.i1.pp222-237.
[25] Rajayyan, S. and Mustafa, S.M.M. (2023), "Prediction of dementia using machine learning model and performance improvement with cuckoo algorithm", *Int. J. Electr. Comput. Eng.*, 13(4), 4623-4632. https://doi.org/10.11591/ijece.v13i4.pp4623-4632.
[26] Sarica, A., Cerasa, A. and Quattrone, A. (2017), "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review", *Front. Aging Neurosci.*, 9, 329. https://doi.org/10.3389/fnagi.2017.00329.
[27] Bin-Hezam, R. and Ward, T.E. (2019), "A machine learning approach towards detecting dementia based on its modifiable risk factors", *Int. J. Adv. Comput. Sci. Appl.*, 10(8), 148-158. https://doi.org/10.14569/IJACSA.2019.0100820.
[28] Uddin, K.M.M., Alam, M.J., E-Anawar, J., Jannat, U., Uddin, M.A. and Aryal, S. (2023), "A novel approach utilizing machine learning for the early diagnosis of Alzheimer's disease", *Biomed. Mater. Devices*, https://doi.org/10.1007/s44174-023-00078-9.
[29] Bansal, M., Goyal, A. and Choudhary, A. (2022), "A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short-term memory algorithms in machine learning", *Decision Analytic J.*, 3, 100071. https://doi.org/10.1016/j.dajour.2022.100071.
[30] Kreuzberger, D., Kuhl, N. and Hirschl, S. (2022), "Machine learning operations (MLOps): Overview, definition, and architecture", *IEEE Access*, 11, 31866-31879. https://doi.org/10.1109/ACCESS.2023.3262138.
[31] Lee, J., et al. (2023), "Prevalence of dementia in India: National and state estimates from a nationwide study", *Alzheimer's Dement.*, 19(7), 2898-2912. https://doi.org/10.1002/alz.12928.
[32] Li, D., Jia, J., Zeng, H., Zhong, X., Chen, H. and Yi, C. (2024), "Efficacy of exercise rehabilitation for managing patients with Alzheimer's disease", *Neural Regen. Res.*, 19(10), 2175-2188. https://doi.org/10.4103/1673-5374.391308.
[33] Montgomery, W., Ueda, K., Jorgensen, M., Stathis, S., Cheng, Y. and Nakamura, T. (2017), "Epidemiology, associated burden, and current clinical practice for the diagnosis and management of Alzheimer's disease in Japan", *Clinicoecon. Outcomes Res.*, 10, 13-28. https://doi.org/10.2147/CEOR.S146788.
[34] Souchet, B., Michaïl, A., Billoir, B. and Braudeau, J. (2023), "Biological diagnosis of Alzheimer's disease based on amyloid status: An illustration of confirmation bias in medical research", *Int. J. Mol. Sci.*, 24(24), 17544. https://doi.org/10.3390/ijms242417544.
[35] Tahami Monfared, A.A., Byrnes, M.J., White, L.A. and Zhang, Q. (2022), "Alzheimer's disease: Epidemiology and clinical progression", *Neurol. Ther.*, 11(2), 553-569. https://doi.org/10.1007/s40120-022-00338-8.
[36] Band, S., Yarahmadi, A., Hsu, C.-C., Biyari, M., Sookhak, M., Ameri, R., et al. (2023), "Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods", *Inform. Med. Unlocked*, 40, 101286. https://doi.org/10.1016/j.imu.2023.101286.
[37] Tang, X. and Liu, J. (2021), "Comparing different algorithms for the course of Alzheimer's disease using machine learning", *Ann. Palliat. Med.*, 10(9), 9715-9724. https://doi.org/10.21037/apm-21-2013.

[38] Vinet-Couchevellou, M. (2022), “Subjective dementification as an effect of Alzheimer’s ideology”, *L’Evolution Psychiatrique*, 87(3), 631-643. https://doi.org/10.1016/j.evopsy.2021.09.005.
[39] Hosam El-Sofany, Bouallegue, B. and Abd El-Latif, Y.M. (2024), “A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method”, *Sci. Rep.*, 14, Article 74656. https://doi.org/10.1038/s41598-024-74656-2.
[40] Nichols, E., et al. (2019), “Global, regional, and national burden of Alzheimer’s disease and other dementias, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016”, *Lancet Neurol.*, 18(1), 88-106. https://doi.org/10.1016/S1474-4422(18)30403-4.
[41] Hosam El-Sofany (2024), “Predicting heart diseases using machine learning and different data classification techniques”, IEEE Access, 12, 106146-106160. https://doi.org/10.1109/ACCESS.2024.3437181.
[42] Alroobaea, R., et al., 2021. Alzheimer’s Disease Early Detection Using Machine Learning Techniques. https://doi.org/10.21203/RS.3.RS-624520/V1.
[43] Hosam El-Sofany, Belgacem Bouallegue, Yasser M. Abd El-Latif, "A proposed biometric authentication hybrid approach using iris recognition for improving cloud security", Heliyon, https://doi.org/10.1016/j.heliyon.2024.e36390, 16 August 2024.
[44] Castellano, G., Esposito, A., Lella, E., Montanaro, G., and Vessio, G. (2024). Automated detection of Alzheimer’s disease: a multi-modal approach with 3D MRI and amyloid PET. Sci. Rep. 14:5210. doi: 10.1038/s41598-024-56001-9
[45] Hosam El-Sofany, Samir A. El-Seoud, Omar H. Karam, Belgacem Bouallegue, Abdelmoty M. Ahmed. "A Proposed Secure Framework for Protecting Cloud-Based Educational Systems from Hacking". Egyptian Informatics Journal, https://doi.org/10.1016/j.eij.2024.100505, Volume 27, September 2024.