Harnessing Explainable Artificial Intelligence (XAI) based SHAPLEY Values and Ensemble Techniques for Accurate Alzheimer's Disease Diagnosis

Bala Krishnan Raghupathy
Department of CSE, SASTRA Deemed to be University, India
balakrishnan@src.sastra.edu

Manyam Rajasekhar Reddy
School of Computing, Amrita Vishwa Vidyapeetham, Amaravati Campus, India
m_rajasekharreddy@av.amrita.edu

Prasad Theeda
Monash University, Malaysia Campus | Vellore Institute of Technology, Vellore, India
prasad.theeda@monash.edu

Elangovan Balasubramanian
Department of CSE, Koneru Lakshmaiah Education Foundation, India.
elan77@gmail.com

Rajesh Kumar Namachivayam
Department of CSE, SASTRA Deemed to be University, India
rajeshkumar.rb@gmail.com

Manikandan Ganesan
School of Computing, SASTRA Deemed to be University, India
manikandan@it.sastra.edu (corresponding author)

Received: 15 November 2024 | Revised: 27 December 2024 and 8 January 2025 | Accepted: 11 January 2025

DOI: https://doi.org/10.48084/etasr.9619

ABSTRACT

Machine Learning (ML) is a dynamic method for managing extensive datasets to uncover significant patterns and hidden insights. ML has revolutionized numerous industries, from healthcare to finance, and from entertainment to transportation. Ensemble classifiers combined with Explainable AI (XAI) have surfaced as a significant asset in the field of Alzheimer's Disease (AD) diagnosis. Boosting EC techniques coupled with Shapley Additive Explanations (SHAP) offers a powerful approach to AD diagnosis. This paper investigates boosting ensemble ML schemes, such as XGBoost, LightGBM, and Gradient Boosting (GB), for AD diagnosis and SHAP for feature selection. The proposed scheme achieved efficient results, with an accuracy of more than 94% with minimum features for the detection process.

Keywords-machine learning; SHAP; ensemble classifiers; explainable AI; Alzheimer's disease

I. INTRODUCTION

The digital era has led to a new era of healthcare, where Machine Learning (ML) is transforming the methods of diagnosing, treating, and preventing diseases. ML algorithms can evaluate large amounts of patient data, including medical records, genetic information, and imaging scans, allowing trend detection and outcome prediction with impressive precision. This enables early disease prediction and detection, more personalized treatment plans, and improved patient outcomes.

The diagnosis of Alzheimer's Disease (AD) poses a substantial problem for global healthcare systems, affecting millions worldwide, with over 55 million individuals diagnosed with dementia in 2020. The number of such individuals is expected to roughly double every two decades, reaching 78 million by 2030 and 139 million by 2050. This disease not only affects individuals but also imposes a significant economic burden on healthcare systems around the world. A timely and precise diagnosis is essential for optimal treatment and possible therapeutic measures. AD is a growing concern in India, with significant implications for public health and the economy. Recent estimates indicate that approximately 8.8 million Indians aged 60 and above currently live with dementia, which translates to a prevalence rate of 7.4% among the age group of 60 and older. This rate is expected to increase dramatically, with projections suggesting that the number of people suffering from dementia in India could reach 11.4 million by 2050. The increase in dementia cases is particularly pronounced in rural areas among women. As the elderly population in India is projected to reach 20% of the total population by 2031, the burden of AD is expected to escalate, highlighting the urgent need for effective healthcare strategies and support systems to address this silent epidemic.

ML has become a transformative instrument in the expansive domain of Artificial Intelligence (XAI) for the effective processing of AD datasets. Traditional diagnostic methods often rely on subjective evaluations and may not be sensitive enough to early-stage AD detection. ML algorithms analyze complex patterns within large datasets, including neuroimaging data, genetic information, and cognitive assessments. Taking advantage of the above-stated data on AD, the available ML models can identify subtle changes in brain structure and function that can precede clinical symptoms, allowing earlier and more precise diagnoses. Early detection of AD is crucial in healthcare, as it facilitates prompt therapies that can decelerate disease progression and improve the quality of life for patients.

Ensemble methods, such as bagging, boosting, and stacking, are powerful techniques that integrate many models to enhance the overall prediction efficacy of ML applications. Although bagging and stacking can be useful, boosting frequently emerges as the superior option, particularly in fields such as healthcare, where reliable and precise predictions are crucial. The sequential learning process of boosting principles allows to iteratively execute classifiers and permits focus on misclassified instances, resulting in more robust models. This makes it ideal for complex, noisy, and sensitive healthcare data, such as AD datasets, where subtle patterns and nuances can be challenging to achieve efficient results.

Ensemble methods, which combine multiple ML models, have shown particular promise in healthcare applications, including AD detection. By leveraging the strengths of diverse ML and DL models, ensemble methods can enhance accuracy, robustness, and generalizability. Boosting is a type of ensemble that iteratively improves the performance of weak learners, resulting in highly accurate and reliable predictions. This makes them well suited for the complex and nuanced nature of AD diagnosis, where multiple factors contribute to disease progression. Boosting algorithms are a cornerstone of ensemble classifiers that have been proven to be transformative in healthcare applications. Table I presents the importance of ML and AI in the domain of healthcare.

TABLE I. IMPORTANCE OF ML AND AI IN HEALTHCARE

	Feature	Description
1	Early disease detection	ML algorithms can scrutinize extensive datasets, allowing the identification or detection of early disease indicators.
2	Individualized healthcare	ML algorithms can assess patient data to tailor treatments to the specific requirements of each unique patient.
3	Accurate diagnosis	AI-powered tools can analyze sensitive medical data, reducing human error and improving diagnostic precision.
4	Research and development of medical supplies	AI can expedite pharmaceutical development by forecasting molecular interactions.
5	Efficient clinical trials	ML can lead to faster and more cost-effective trials.
6	Enhanced treatment for patients	AI-powered chatbots and virtual assistants can deliver support on time.
7	Administrative efficiency	ML supports automating tasks such as appointment scheduling, claims, processing, and so on.

The following are the main contributions of this study:

- Incorporates Shapley Additive Explanation (SHAP) principles to ensure accurate feature selection.
- Compares multiple ML algorithms to improve detection accuracy with and without feature selection.
- Evaluates the efficiency of the proposed scheme using various computational metrics and error rates.
- The proposed model can help healthcare service providers handle their raw data about patients and effectively focus on decision-making with observational outcomes.

In [1], various ML approaches, such as Support Vector Machines (SVM), Random Forests (RF), and neural networks, were used to predict and examine dementia progression using neuroimaging, genetics, and clinical data. Several data types, including genetic, biomarker, and neuroimaging, were integrated to increase predictive accuracy. In [2], advanced diagnostic methods for AD were examined, using ML techniques to improve early detection of AD. This work utilized neural networks, SVM, and clustering in conjunction with MRI and PET scans to identify different stages of the disease. The ability of ML models to process large datasets makes them ideal for examining blood proteins.

In [3], a novel scheme was proposed for AD diagnosis, combining ML principles with biomarkers and showing improved sensitivity and specificity in the diagnosis process. In [4], Naive Bayes (NB), SVM, and K-Nearest Neighbor (KNN) methods were used to detect early-stage AD using gene expression data. The patients were classified according to their gene expression profiles. Information Gain (IG) was utilized to discover 44 essential features among 16,000 genes. The dataset included only to gene expression data, limiting the models' generalizability.

In [5], Logistic Regression (LR) and SVM were combined to predict heart disease, using improved auto-categorical particle swarm optimization for feature selection and achieving up to 33% performance improvement. In [6], a voting scheme was proposed for early AD diagnosis, using NB, RF, XGBoost, and GB on the OASIS dataset and achieving 96% accuracy. In [7], Auto-WEKA, Decision Table / NB (DTNB), and multi-objective evolutionary fuzzy classifiers were used to predict cardiovascular disease. In [8], principal component analysis was used for feature selection to improve the prediction accuracy of coronary heart disease diagnosis, achieving higher accuracy with AdaBoost. In [9], a robust framework was introduced for AD detection and staging. This was achieved by MRMR feature selection and RF classification, achieving an overall test accuracy of 93.11%. In [10], an extensive review of AD detection using pre-trained models was presented. This survey comprehensively presented the existing literature on various methods and approaches for AD detection.

II. PROPOSED SYSTEM

Figure 1 shows the workflow of the proposed model.

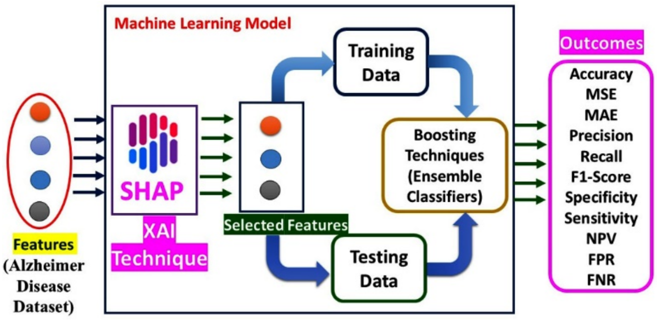

Fig. 1. Workflow of the proposed XAI-based ensemble classifier for AD detection.

A. *Dataset Description*

The proposed system used the AD dataset in [11], which offers a rich source comprising 2149 patient records, each containing a wealth of information about patients, with 36 features that include demographics, lifestyle factors, medical history, clinical measurements, cognitive assessments, and diagnosis. This dataset provides a comprehensive view of crucial data for understanding and investigating AD. The dataset often contains research notes from clinicians and researchers, providing valuable insight into the disease's progression and potential treatments. It holds a total count of 1389 non-diseased and 760 diseased patient details.

B. *Machine Learning (ML) with Explainable Artificial Intelligence (XAI)*

The integration of XAI, namely SHAP (SHapley Additive exPlanations), along with boosting ensemble classifiers such as XGBoost, LightGBM, and GB is pivotal for the AD diagnosis process. XAI-enhanced ML models can provide not only high accuracy but also effective interpretability, which is highly essential in the diagnosis of healthcare data to gain clinician trust and understanding. SHAP values can help demystify model decisions by attributing significance to each feature, ensuring transparency, and aiding in the identification of key biomarkers.

C. *Execution*

The AD dataset contains numeric and non-numeric data that were preprocessed. In this phase, non-numeric features were converted to numeric ones using label encoding. The encoded data was input to SHAP to detect efficient features for the diagnostic process. The essential features extracted from SHAP, stated as the preprocessed dataset, were then considered for evaluation. The preprocessed features were divided into 80:20 to train and test the boosting-based ensemble classifiers XGBoost, LightGBM, and GB.

III. EXPERIMENTAL ANALYSIS

The models were implemented in Windows 11 64-bit using Python 3.13 on Jupyter Notebook 7.2.2. Accuracy, testing time, MSE, Precision, Recall, F1-score, False Positive Rate (FPR), and False Negative Rate (FNR) were some of the metrics used to evaluate the performance of the models.

$$Accuracy = \frac{TP+TN}{FN+TN+FP+TP} \tag{1}$$

$$Precision = \frac{TP}{FP+TP} \tag{2}$$

$$Recall = \frac{TP}{FN+TP} \tag{3}$$

$$F1-Score = \frac{2\,(Precision*Recall)}{(Precision+Recall)} \tag{4}$$

$$Specificity = \frac{TN}{FP+TN} \tag{5}$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP+TN} \tag{6}$$

$$False\ Negative\ Rate\ (FNR) = \frac{FN}{FN+TP} \tag{7}$$

Table II presents the results of the boosting classifiers without applying SHAP, considering all 36 features. Figure 2 depicts the Receiver Operating Characteristic (ROC) curve as well as the Precision-Recall curve for the observations in Table II.

TABLE II. RESULTS WITHOUT SHAP

Metrics	**Features count: 36**		
	Classifiers		
	XGBoost	**LightGBM**	**GB**
Accuracy	94.65%	94.88%	94.64%
Testing Time	0.52 s	0.727 s	2.155 s
Mean Squared Error	0.053	0.051	0.053
Precision	91.10%	92.20%	91.10%
Recall	93.40%	93.40%	94.10%
F1-score	92.50%	92.80%	92.60%
Specificity	91.61%	92.20%	91.08%
(FPR)	0.83%	0.77%	0.89%
(FNR)	0.36%	0.36%	0.32%

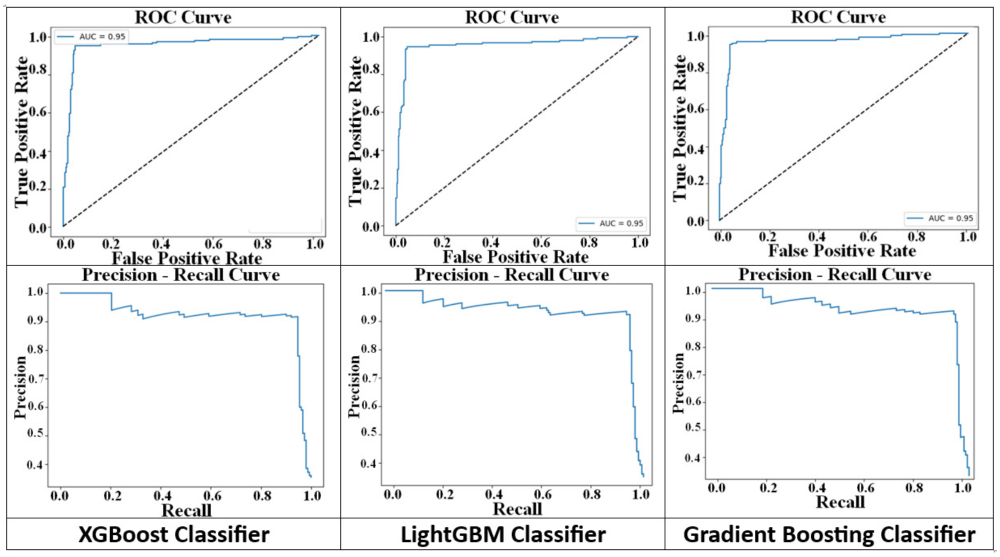

Fig. 2. Evolution of ROC and Precision-Recall curve for XGBoost, LightGBM, and GB classifiers without SHAP.

Figure 3 presents the results of the feature selection using SHAP on the AD dataset. This approach indicated 29 features for the observations. Table III presents the results of the boosting models, after selecting 29 features using SHAP. Figure 4 displays the ROC and the Precision-Recall curves for the observations presented in Table III.

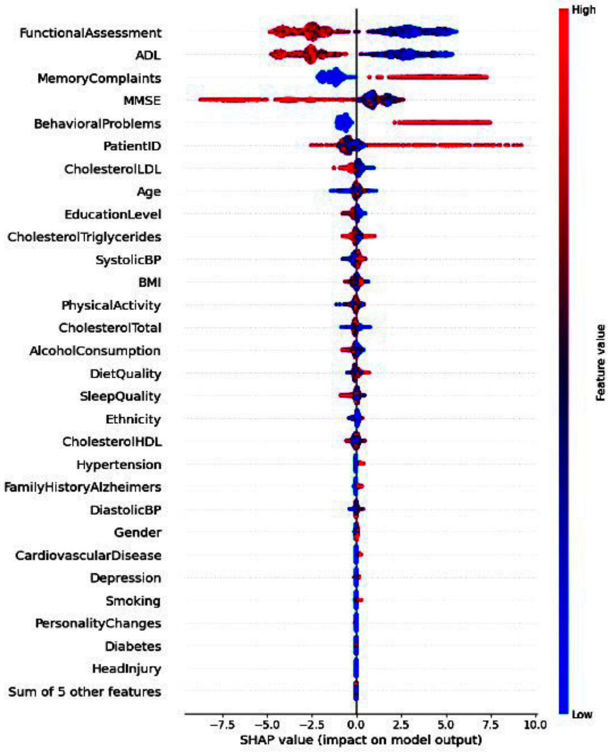

Fig. 3. Results after feature selection by SHAP.

TABLE III. RESULTS WITH SHAP

Metrics	**Features count: 29**		
	Classifiers		
	XGBoost	**LightGBM**	**GB**
Accuracy	94.88%	95.11%	96.91%
Testing Time	0.541 s	0.206 s	0.77 s
Mean Squared Error	0.0511	0.048	0.05
Precision	91.60%	92.80%	91.70%
Recall	94.70%	93.40%	94.10%
F1-Score	92.90%	93.10%	92.90%
Specificity	91.13%	92.80%	91.66%
Sensitivity	97.05%	96.38%	96.71%
False Positive Rate (FPR)	0.88%	0.71%	0.83%
False Negative Rate (FNR)	0.29%	0.36%	0.32%

The proposed boosting scheme with SHAP for feature selection was compared with basic ML classifiers such as NB, LR, DT, KNN, and RF, as shown in Table IV. These results highlight the competitive performance of the proposed approach, showing its suitability for decision-making in healthcare applications.

TABLE IV. ACCURACY OF VARIOUS CLASSIFIERS WITH SHAP FEATURE EXTRACTION

S. No	**Classifiers**	**Accuracy Percentage**
1	**NB**	91.34%
2	**LR**	89.71%
3	**DT**	92.11%
4	**KNN**	93.43%
5	**RF**	93.48%
6	**XGBoost**	94.88%
7	**LightGBM**	95.11%
8	**GB**	96.91%

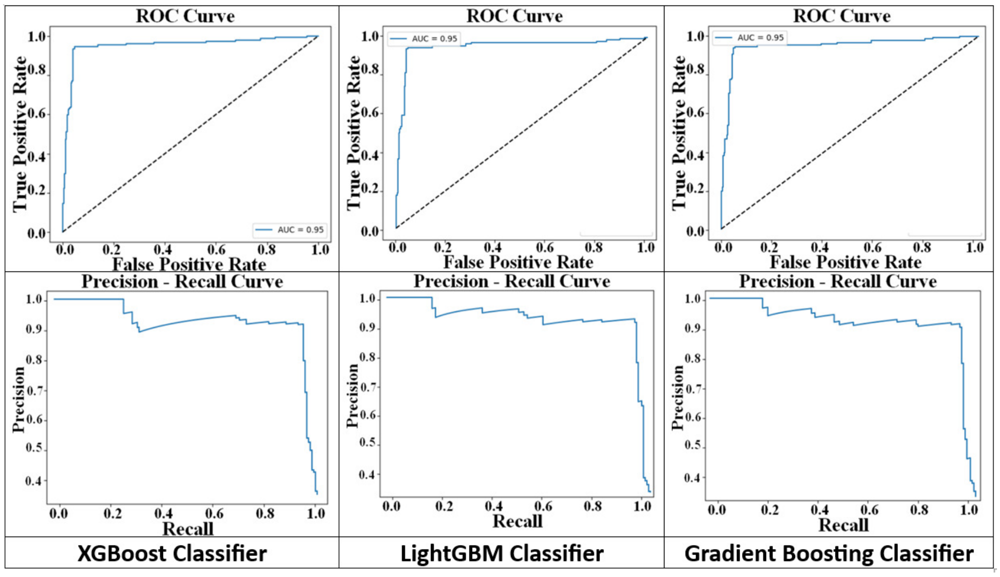

Fig. 4. Evolution of ROC and Precision-Recall curves for XGBoost, LightGBM, and GB classifiers with SHAP.

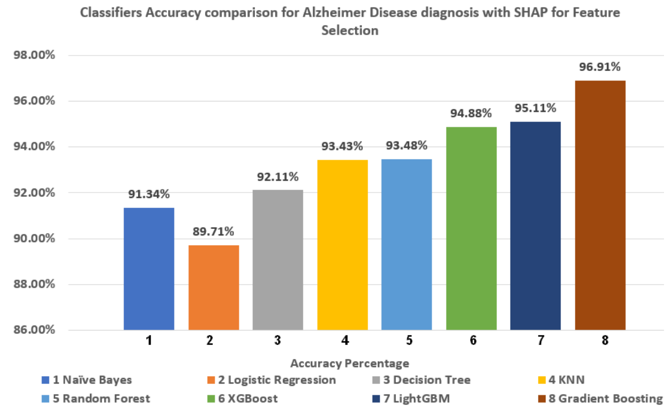

Fig. 5. Comparison of various classifiers with SHAP.

IV. CONCLUSION

This study investigated using SHAP for feature selection before using ML models for classification. Ensemble boosting-based techniques such as XGBoost, LightGBM, and GB combined with SHAP provide an effective approach to AD diagnosis. The proposed approach provides an efficient output with an accuracy of more than 94% using minimal features for the detection procedure. According to experimental observations, the proposed scheme achieved higher accuracy in AD diagnosis with less testing time, demonstrating its suitability among traditional classifiers.

REFERENCES

[1] R. Li, "Data Mining and Machine Learning Methods for Dementia Research," in *Biomarkers for Alzheimer's Disease Drug Development*, R. Perneczky, Ed. Springer, 2018, pp. 363–370.

[2] Z. Fan, F. Xu, X. Qi, C. Li, and L. Yao, "Classification of Alzheimer's disease based on brain MRI and machine learning," *Neural Computing and Applications*, vol. 32, no. 7, pp. 1927–1936, Apr. 2020, https://doi.org/10.1007/s00521-019-04495-0.

[3] C. H. Chang, C. H. Lin, and H. Y. Lane, "Machine Learning and Novel Biomarkers for the Diagnosis of Alzheimer's Disease," *International Journal of Molecular Sciences*, vol. 22, no. 5, Jan. 2021, Art. no. 2761, https://doi.org/10.3390/ijms22052761.

[4] S. T. Ahmed and S. M. Kadhem, "Alzheimer's disease prediction using three machine learning methods," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 3, pp. 1689-1697, Sep. 2022, https://doi.org/10.11591/ijeecs.v27.i3.pp1689-1697.

[5] A. K. Dubey, A. K. Sinhal, and R. Sharma, "An Improved Auto Categorical PSO with ML for Heart Disease Prediction," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8567–8573, Jun. 2022, https://doi.org/10.48084/etasr.4854.

[6] K. M. M. Uddin, M. J. Alam, Jannat-E-Anawar, M. A. Uddin, and S. Aryal, "A Novel Approach Utilizing Machine Learning for the Early Diagnosis of Alzheimer's Disease," *Biomedical Materials & Devices*, vol. 1, no. 2, pp. 882–898, Sep. 2023, https://doi.org/10.1007/s44174-023-00078-9.

[7] S. M. Alanazi and G. S. M. Khamis, "Optimizing Machine Learning Classifiers for Enhanced Cardiovascular Disease Prediction," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12911–12917, Feb. 2024, https://doi.org/10.48084/etasr.6684.

[8] M. J. Ghrabat *et al.*, "Utilizing Machine Learning for the Early Detection of Coronary Heart Disease," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17363–17375, Oct. 2024, https://doi.org/10.48084/etasr.8171.

[9] H. K. Ravikiran, R. Deepak, H. A. Deepak, M. S. Prapulla Kumar, S. Sharath, and G. H. Yogeesh, "A robust framework for Alzheimer's disease detection and staging: incorporating multi-feature integration, MRMR feature selection, and Random Forest classification," *Multimedia Tools and Applications*, Aug. 2024, https://doi.org/10.1007/s11042-024-19875-5.

[10] M. H. M. Khan, P. Reesaul, M. M. Auzine, and A. Taylor, "Detection of Alzheimer's disease using pre-trained deep learning models through transfer learning: a review," *Artificial Intelligence Review*, vol. 57, no. 10, Sep. 2024, Art. no. 275, https://doi.org/10.1007/s10462-024-10914-z.

[11] R. El Kharoua, "Alzheimer's Disease Dataset." Kaggle, https://doi.org/10.34740/kaggle/dsv/8668279.