Journal of Information Analytics, 1(2), 1-13

Journal of Information Analytics

Submit 25.12.2025
Revision 09.01.2026
Accept 09.01.2026
Publish 10.01.2026

Research Article

Open Access

An Optimized Ensemble Learning Approach for Accurate Alzheimer's Disease Classification

Ahmet Aydın

Vocational School of Information Technologies, Adana Alparslan Türkeş Science and Technology University, Adana, Türkiye

Abstract

Effective clinical management of Alzheimer's disease (AD) fundamentally relies on the precision and timing of initial diagnosis. This research introduces a robust machine learning framework designed to categorize AD stages by synthesizing a diverse array of demographic, clinical, and lifestyle indicators. Unlike conventional approaches, our methodology integrates a multi-stage pipeline ranging from rigorous data preprocessing to advanced ensemble learning and strategic decision threshold calibration. While various baseline models were scrutinized, the Random Forest architecture emerged as the most resilient classifier. To maximize its diagnostic utility, we refined the model through stratified cross-validation and randomized hyperparameter exploration, fine-tuning the balance between sensitivity and specificity. The resulting system yielded high-performance metrics, notably an accuracy of 95.3% and an AUC of 0.94, underscoring its generalizability to unseen data. Crucially, feature importance mapping identified Functional Assessment, ADL scores, and MMSE results as the primary drivers of classification, bridging the gap between computational outputs and established geriatric insights. These findings suggest that the optimized Random Forest framework functions not merely as a predictive tool, but as a transparent and interpretable decision-support mechanism for clinicians.

Keywords: Alzheimer's disease, Machine learning, Random Forest, Clinical decision support,

Feature importance

Citation: Aydin, A., (2025). An Optimized Ensemble Learning Approach for Accurate Alzheimer's Disease Classification. *Journal of Information Analytics*, 1(2), 1-13.

Corresponding Author: Ahmet Aydın aaydin@atu.edu.tr

1. Introduction

Alzheimer's disease (AD) represents a debilitating and progressive neurodegenerative condition that fundamentally impairs memory and cognitive autonomy, placing an escalating strain on global healthcare infrastructures and caregiving networks. Given the insidious nature of AD progression, the ability to secure an early and precise diagnosis is paramount not only for slowing symptomatic decline but also for optimizing the strategic window for clinical intervention and patient support. Traditional diagnostic pathways, however, frequently necessitate exhaustive neuropsychological assessments and expert clinical synthesis, a process that can be inherently subjective, resource-intensive, and difficult to implement at scale.

The paradigm shift toward data-driven methodologies has positioned machine learning (ML) as a formidable ally in clinical decision-making. By deciphering intricate patterns within multifaceted datasets encompassing demographic, cognitive, and lifestyle-related variables ML architectures can detect subtle pathological signatures that might elude conventional clinical observation. Consequently, supervised classification models have emerged as a cornerstone for modern disease prognosis and risk stratification.

Nevertheless, the integration of ML into Alzheimer's diagnostics is not without significant hurdles. Achieving robust generalization across heterogeneous patient populations, balancing the critical trade-off between sensitivity and specificity, and ensuring model interpretability remain major obstacles. A high-performing model that operates as a "black box" often faces skepticism in clinical settings, where understanding the why behind a prediction is as vital as the prediction itself.

In this landscape, ensemble learning techniques specifically Random Forest architectures provide a compelling equilibrium between raw predictive power and structural transparency. By synthesizing the outputs of diverse decision trees, Random Forests effectively navigate non-linear feature interactions while mitigating the risks of overfitting. Furthermore, their capacity for intrinsic feature importance ranking allows for a clearer understanding of the clinical and cognitive drivers behind a diagnosis (Aydin et al., 2021).

Motivated by these considerations, this study introduces an optimized Random Forest framework tailored for the classification of Alzheimer's disease. Our approach prioritizes clinical utility by emphasizing sensitivity and model explainability through a rigorous experimental pipeline. The primary contributions of this research are as follows:

- Multidimensional Framework: Development of an integrated ML diagnostic system utilizing a holistic set of clinical, demographic, and lifestyle features.
- Systematic Benchmarking: A comprehensive evaluation of baseline models to establish a rigorous performance baseline.
- Architectural Optimization: Refinement of the Random Forest model through stratified cross-validation and hyperparameter search to ensure resilience across datasets.
- Sensitivity-Centric Tuning: Application of decision threshold optimization to minimize false negatives, a critical factor in geriatric clinical care.

- Interpretability Mapping: Execution of feature importance analysis to align computational results with clinical knowledge, thereby enhancing the framework's transparency.

2. Literature Review

The surge in clinical, neuroimaging, and lifestyle-related datasets has catalyzed a significant shift toward machine learning (ML) applications in Alzheimer's Disease (AD) diagnostics. While classical diagnostic protocols remain the gold standard, the inherent need for expert-led, resource-heavy evaluations has fueled the demand for automated decision-support systems capable of facilitating early-stage detection (Aydin & Avaroğlu, 2024).

Historically, research in this field was anchored in traditional statistical modeling and basic ML algorithms, primarily using cognitive scores and demographic data to identify AD markers. However, these pioneering efforts often struggled with constrained datasets and narrow feature sets, which limited their ability to generalize across broader populations. As computational power evolved, the focus shifted toward more sophisticated architectures, such as Support Vector Machines (SVM), neural networks, and ensemble methods. While these models excel at deciphering non-linear complexities within clinical data, their "black-box" nature often creates a tension between high predictive accuracy and the interpretability required for clinical integration.

Recent scholarly contributions have increasingly favored hybrid and multimodal strategies to overcome these interpretative and performance-related hurdles. For instance, Raza et al. (2024) demonstrated the efficacy of merging handcrafted MRI features with deep learning representations to refine classification on the ADNI dataset. Pushing the boundaries of traditional computing, Belay et al. (2024) integrated quantum machine learning with ensemble deep learning, achieving remarkable AUC and accuracy metrics. Furthermore, the work of Sheng et al. (2024) underscores the necessity of data fusion, revealing that a synthesis of MRI, PET, and cerebrospinal fluid (CSF) biomarkers consistently outperforms single-modality frameworks by leveraging complementary biological signals.

Systematic reviews also reflect these evolving trends. Both Kaur et al. (2024) and Hechkel & Helali (2025) observe that CNN-based architectures and transfer learning particularly when applied to neuroimaging currently lead the field in diagnostic performance. Furthermore, Singh et al. (2024) highlighted the capacity of deep learning to predict the transition from Mild Cognitive Impairment (MCI) to clinical AD, though they cautioned that issues regarding cohort heterogeneity and model transparency persist. Beyond mere diagnosis, ML is also reshaping therapeutic discovery; Tarikul Islam et al. (2024) utilized an in silico and ML-driven screening process to identify potential acetylcholinesterase inhibitors, showcasing the versatility of predictive modeling in AD treatment research.

Despite this progress, a critical synthesis of existing literature reveals several persistent gaps. A significant portion of current research prioritizes raw accuracy over clinical sensitivity and lacks rigorous baseline comparisons or systematic threshold optimization. Many models remain unoptimized for real-world deployment where the cost of a false negative is high. Addressing these deficiencies, the present study introduces an optimized Random Forest framework that seeks to harmonize predictive robustness with clinical interpretability and strategic decision-tuning.

3. Methodology

The methodological framework of this research encompasses a multi-layered approach to Alzheimer's disease classification, integrating rigorous data preparation, systematic model selection, and advanced optimization protocols. To ensure that the resulting diagnostic tool is both technically robust and clinically viable, the pipeline transitions from granular data preprocessing to high-level ensemble learning and decision threshold calibration. Each phase of the methodology was curated to maximize predictive fidelity while maintaining the structural transparency necessary for healthcare applications. The following subsections detail the dataset characteristics, the algorithmic architectures explored, and the specific metrics utilized to validate the framework's diagnostic integrity.

3.1. Dataset Description and Data Preprocessing

The empirical foundation of this research is based on a structured clinical dataset (Kharoua, 2024), which captures a holistic profile of patient health through demographic, physiological, and behavioral lenses. The feature space is multidimensional, encompassing fundamental metrics such as age and body mass index (BMI), alongside cardiovascular indicators like blood pressure and cholesterol levels. To capture the multifaceted nature of Alzheimer's pathology, the dataset also incorporates lifestyle variables specifically sleep hygiene and dietary patterns complemented by rigorous cognitive assessments, functional evaluation scores, and observed behavioral symptoms. The diagnostic objective is framed as a binary classification task, distinguishing between confirmed Alzheimer's cases and healthy controls.

Regarding data integrity, the initial audit revealed a complete set of records with no missing entries, which obviated the need for statistical imputation techniques. To maintain the scientific validity of the predictive process and prevent artificial performance inflation (data leakage), all unique identifiers and non-contributory metadata were systematically stripped from the feature matrix. This refinement ensures that the subsequent machine learning models derive their predictive power solely from clinically relevant patterns rather than incidental data artifacts.

3.2. Baseline Machine Learning Models

To develop a rigorous performance benchmark, this study initially scrutinized several foundational machine learning architectures. The candidate pool included Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, and Random Forest classifiers. To maintain the integrity of the comparative analysis and ensure a level playing field, each model was trained on an identical subset of data and validated against the same unseen test set under strictly synchronized experimental conditions.

The preliminary comparative analysis indicated a clear performance hierarchy, with ensemble-based methodologies specifically the Random Forest architecture consistently outperforming individual classifiers. This superiority stems from the Random Forest's capacity to synthesize the outputs of multiple decision trees, a mechanism that allows it to map the intricate, non-linear interactions inherent in clinical and cognitive datasets. Furthermore, the ensemble averaging property of the model serves as a natural defense against overfitting, which is a common pitfall for simpler, single-model architectures.

Beyond its predictive accuracy, the Random Forest was selected for its intrinsic interpretability. Its built-in feature importance rankings provide a "glass-box" view into the diagnostic process, aligning with the

transparency requirements essential for high-stakes clinical decision-making. Consequently, the Random Forest was adopted as the architectural backbone of the proposed diagnostic framework.

3.3. Model Optimization, Evaluation, and Interpretability

To fortify the model's resilience and its ability to generalize across diverse patient profiles, we implemented a sophisticated experimental optimization protocol. This phase utilized a randomized hyperparameter search integrated with stratified cross-validation. We systematically calibrated critical architectural parameters, such as the ensemble size (number of trees), maximum tree depth, and minimum sample split requirements, alongside class weighting adjustments. The optimization objective was centered on the F1-score, ensuring a harmonized equilibrium between precision and recall a prerequisite for high-stakes clinical diagnostics where neither false alarms nor missed cases are acceptable.

To neutralize potential data partitioning biases and secure more reliable performance benchmarks, a stratified 5-fold cross-validation scheme was consistently applied. Beyond standard parameter tuning, we introduced a decision threshold optimization phase. By shifting away from the conventional default probability thresholds, we empirically scrutinized a range of operating points to identify the threshold that yielded the optimal F1-score. This strategic refinement was specifically designed to bolster clinical sensitivity, thereby minimizing the risk of false-negative outcomes without compromising overall predictive precision.

The integrity of the final model was validated through a multidimensional suite of metrics, including accuracy, precision, recall, F1-score, AUC, and Average Precision (AP). Within this evaluative framework, recall was prioritized as the primary indicator, reflecting our clinical commitment to maximizing the detection rate of Alzheimer's cases. These quantitative results were further corroborated by visual diagnostics, including confusion matrices, ROC curves, and precision–recall trajectories.

Finally, to bridge the gap between computational prediction and clinical practice, we conducted an in-depth feature importance analysis. By extracting the inherent importance scores from the optimized Random Forest ensemble, we pinpointed the most influential demographic, cognitive, and clinical drivers of the model's decisions. This focus on interpretability ensures that the framework's outputs are not only accurate but also transparent and clinically coherent, facilitating their potential adoption in real-world geriatric care.

4. Results

This section provides a detailed synthesis of the experimental findings derived from the proposed diagnostic framework. The analysis is structured into three distinct phases: an initial benchmarking of baseline classifiers, a rigorous evaluation of the optimized Random Forest architecture, and an interpretability study focused on feature importance.

The computational infrastructure utilized for these experiments consisted of a workstation powered by an Intel-based multi-core processor, 16 GB of RAM, and an NVIDIA GeForce RTX 3050 GPU to facilitate efficient processing. From a software perspective, the framework was constructed within a Python

environment, leveraging the scikit-learn library for the implementation of both algorithms and evaluative protocols.

To maintain the statistical integrity of our findings, we employed a stratified train-test split, which ensured that the disease prevalence and class distributions remained consistent across both subsets. A uniform preprocessing pipeline and a fixed random seed were applied to all models to guarantee the reproducibility of the results. While baseline models were assessed using their standard hyperparameter configurations, the Random Forest classifier underwent a more intensive refinement process, incorporating grid-based hyperparameter tuning and five-fold stratified cross-validation. The final model performance was quantified through a robust array of metrics namely accuracy, precision, recall, F1-score, and ROC-AUC with the decision threshold strategically calibrated to maximize the F1-score for clinical viability.

4.1. Baseline Model Performance

To construct a rigorous benchmark for our proposed framework, we initially conducted a comparative assessment of several foundational machine learning architectures. This baseline evaluation encompassed Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, and the Random Forest classifier. To eliminate potential biases and ensure the validity of our comparisons, all models were subjected to an identical experimental regime, utilizing synchronized data partitions and standardized evaluative criteria.

The empirical data, summarized in Table 1, highlights a distinct hierarchy in model efficacy. Among the candidates, the Random Forest classifier emerged as the most robust baseline, consistently outperforming its counterparts across all primary performance indicators. Its ability to maintain superior accuracy and precision levels relative to simpler models underscores the necessity of ensemble-based approaches in handling the complexities of Alzheimer's diagnostic data.

Table 1. Baseline Model Performance

Model	Accuracy	Precision	Recall	F1-score	AUC
Logistic Regression	0.816	0.742	0.737	0.739	0.885
K-Nearest Neighbors	0.707	0.627	0.421	0.504	0.738
Decision Tree	0.900	0.847	0.875	0.861	0.894
Random Forest	0.944	0.944	0.895	0.919	0.940

The performance metrics for the initial candidate models are detailed in Table 1. A comparative review of these results reveals that the Random Forest classifier established the most effective performance ceiling, surpassing Logistic Regression, K-Nearest Neighbors, and Decision Tree architectures in terms of accuracy, F1-score, and AUC. This divergence in performance suggests that the multifaceted nature of clinical and cognitive data is best captured through the collective intelligence of ensemble-based methods.

The superiority of the Random Forest model as a baseline is not merely numerical but reflects its structural resilience in mapping high-dimensional feature interactions. Unlike simpler linear or distance-based classifiers, the ensemble approach effectively managed the inherent noise and non-linearities within the patient records. These empirical observations provided a compelling rationale for adopting the Random Forest as the foundational architecture for the subsequent optimization and threshold-tuning phases of this research.

4.2. Performance of the Optimized Random Forest Model

Building upon the initial baseline assessments, we subjected the Random Forest architecture to an intensive refinement process, integrating hyperparameter optimization, stratified cross-validation, and strategic decision threshold calibration. To gauge the efficacy of these interventions, the final model was deployed on a previously unseen, independent test set to validate its generalization capacity within a simulated clinical environment.

The empirical results underscore the significant performance dividends yielded by this optimization strategy. The finalized model attained an accuracy of 95.3%, supported by a precision of 94.6% and a recall of 92.1%, resulting in a robust F1-score of 93.3%. Furthermore, the model's AUC of 0.94 and Average Precision (AP) of 0.93 reflect an exceptional discriminative capability and high reliability in its probabilistic estimations.

As detailed in Table 2, a comparative synthesis of the baseline metrics against the optimized parameters clearly demonstrates the incremental gains achieved. By fine-tuning the decision threshold, we successfully elevated the model's sensitivity, ensuring that it remains resilient and precise even when faced with the inherent complexities of Alzheimer's diagnostic features.

Table 2. Performance Comparison of Baseline Models and the Optimized Random Forest

Model	Accuracy	Precision	Recall	F1-score	AUC	AP
Logistic Regression	0.816	0.742	0.737	0.739	0.885	–
K-Nearest Neighbors	0.707	0.627	0.421	0.504	0.738	–
Decision Tree	0.900	0.847	0.875	0.861	0.894	–
Random Forest (Baseline)	0.944	0.944	0.895	0.919	0.940	–
Random Forest (Optimized)	0.953	0.946	0.921	0.933	0.940	0.930

Table 2 illustrates the performance improvements achieved through optimization of the Random Forest model. Compared to the baseline Random Forest, the optimized model shows an increase in accuracy from 94.4% to 95.3%, precision from 94.4% to 94.6%, recall from 89.5% to 92.1%, and F1-score from 91.9% to 93.3%. In addition, the optimized model maintains a high AUC of 0.94 while achieving an average precision of 0.93, indicating enhanced discriminative capability and improved sensitivity. These gains demonstrate that hyperparameter tuning and decision threshold optimization effectively reduce false negatives and strengthen the model's suitability for clinical Alzheimer's disease diagnosis, as illustrated in Figure 1.

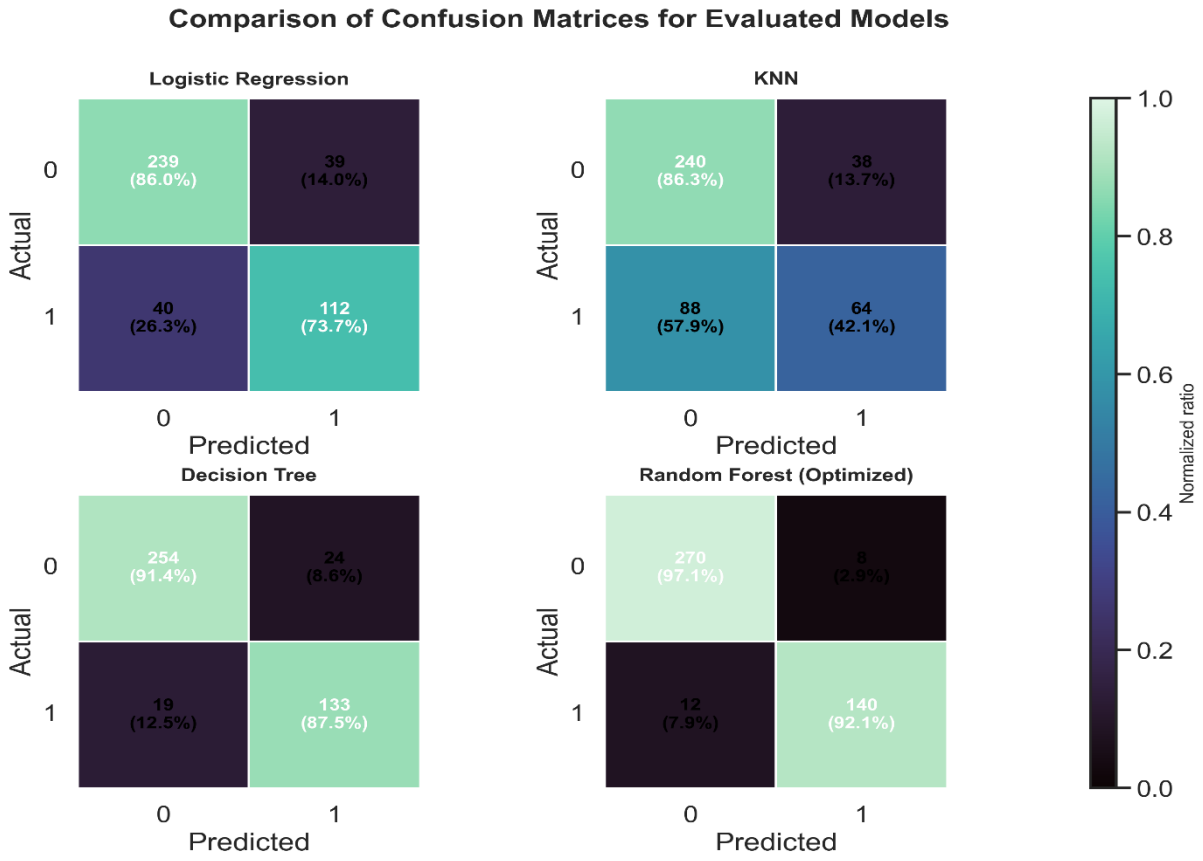

Figure 1. Comparison of Confusion Matrices for Evaluated Models

A comparative confusion matrix analysis is presented in Figure 1, which contrasts the error distributions of the evaluated models. As shown in Figure 1, the optimized Random Forest correctly classified 270 true negatives and 140 true positives, while producing only 8 false positives and 12 false negatives. The relatively low number of false negative cases demonstrates that the optimized framework is particularly effective in identifying Alzheimer's disease cases, which is critical for clinical diagnostic applications, as further supported by the ROC curve analysis shown in 2.

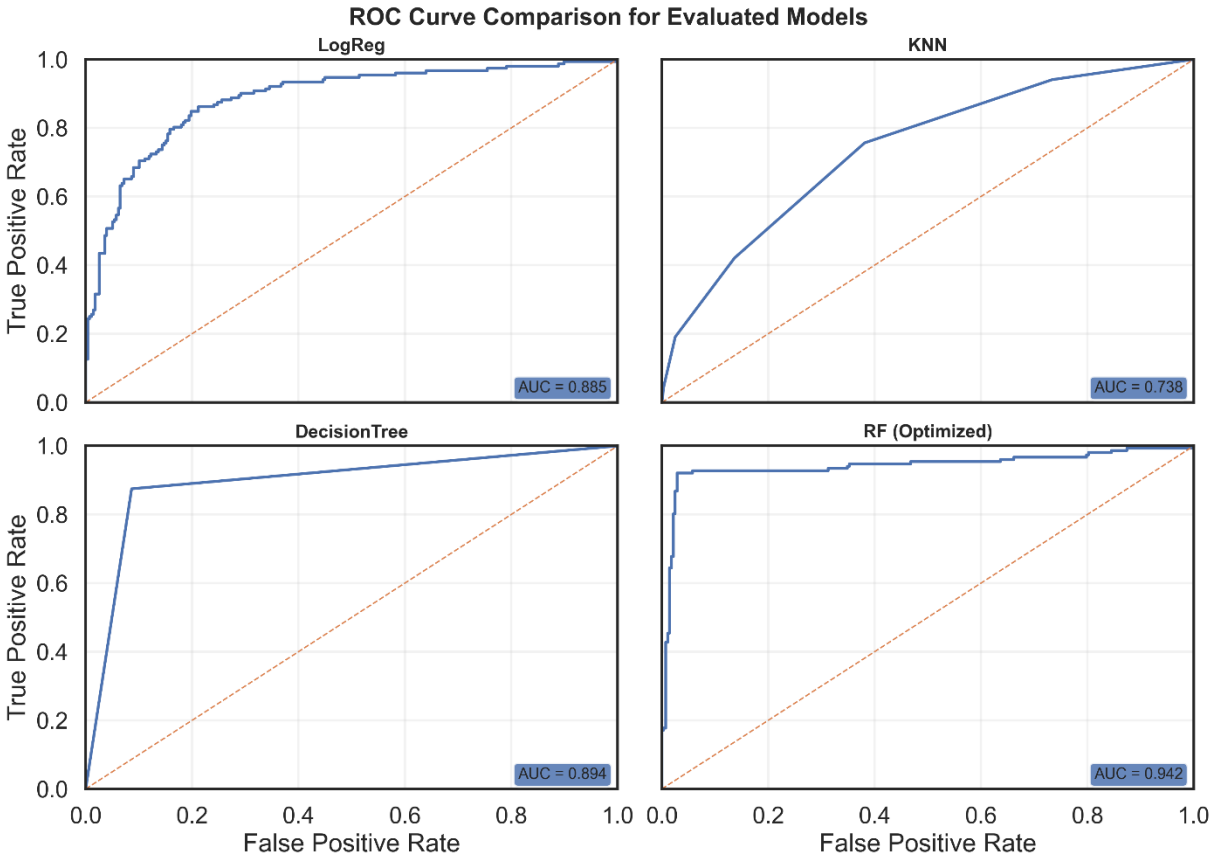

Figure 2. ROC Curve Comparison of Machine Learning Models

The precision–recall curve comparison of the evaluated machine learning models is presented in Figure 3 below.

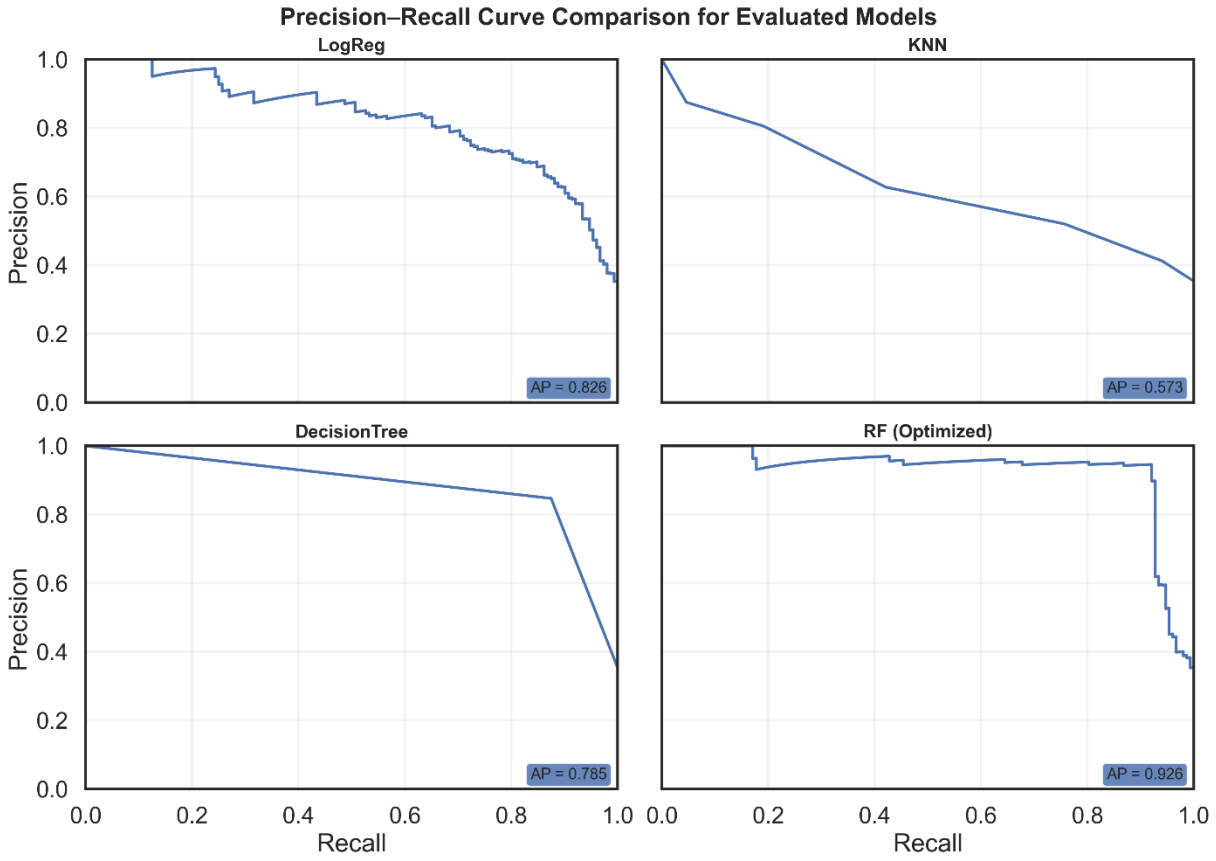

Figure 3. Precision–Recall Curve Comparison of Machine Learning Models

The ROC curves are compared in Figure 2, where the optimized Random Forest exhibits the highest AUC among the evaluated models, confirming strong class separability across decision thresholds. Similarly, the precision–recall (PR) curves in Figure 3 demonstrate that the optimized model maintains a favorable precision–recall trade-off, supporting its suitability for clinical decision-support scenarios where high sensitivity is required.

4.3. Feature Importance and Model Interpretability

In addition to predictive performance, the interpretability of the proposed framework was examined through feature importance analysis derived from the optimized Random Forest model. The resulting feature importance distribution, which reflects the relative contribution of each feature to the classification decision, is presented in Figure 4.

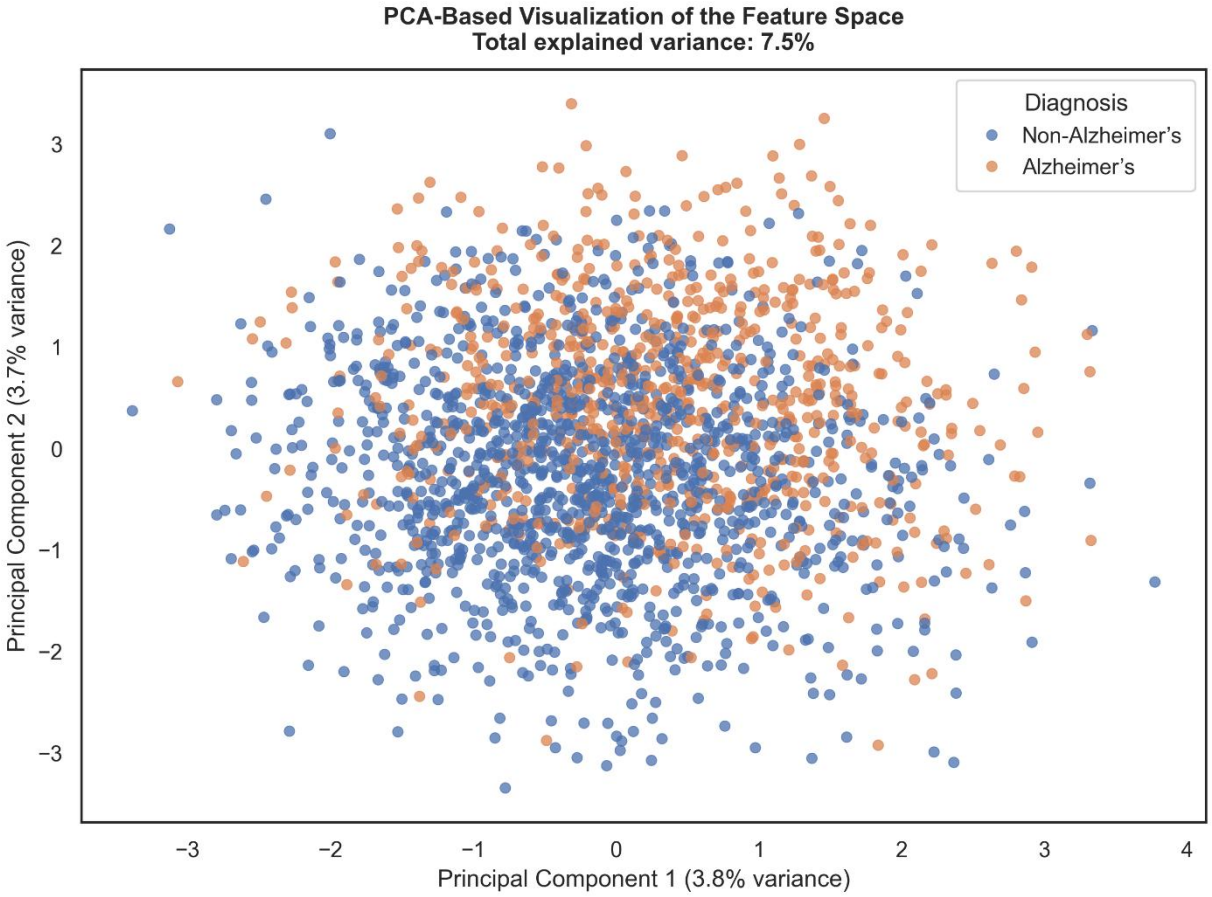

Figure 4. Feature Importance of the Optimized Random Forest Model

As shown in Figure 4, features related to functional and cognitive assessments are among the most prominent in the classification process. In particular, Functional Assessment, Activities of Daily Living (ADL), Mini-Mental State Examination (MMSE) scores, memory-related complaints, and behavioral problems exhibit the highest feature importance values. This indicates that measures associated with cognitive function and daily living activities are more discriminative for Alzheimer's disease classification within the evaluated dataset.

In contrast, physiological and lifestyle-related variables, including cholesterol measures, sleep quality, body mass index, alcohol consumption, and blood pressure, display comparatively lower feature importance scores. This suggests that, in the context of the current dataset, these variables contribute less to the classification decision than cognitive and functional indicators.

Overall, the feature importance analysis shows that the optimized Random Forest model primarily relies on functional and cognitive features when making classification decisions. This observation supports the interpretability of the proposed framework and demonstrates consistency between the model's decision patterns and commonly used clinical assessment indicators.

5. Conclusion

In this study, an optimized Random Forest–based machine learning framework was proposed for the classification of Alzheimer's disease using demographic, clinical, cognitive, and lifestyle-related features. A comprehensive experimental pipeline was designed, incorporating baseline model comparison,

ensemble learning, hyperparameter optimization, stratified cross-validation, and decision threshold tuning to achieve robust and clinically meaningful performance.

The experimental results demonstrated that the optimized Random Forest model achieved high classification accuracy and strong clinical sensitivity, effectively reducing false negative predictions while maintaining high precision. The use of decision threshold optimization further enhanced the model's suitability for clinical decision-support applications, where early and reliable detection of Alzheimer's disease is of critical importance.

In addition to predictive performance, the proposed framework emphasizes interpretability through feature importance analysis. The results indicate that functional and cognitive assessment–related features play a dominant role in Alzheimer's disease classification, aligning the model's decision-making process with established clinical knowledge. This interpretability strengthens the potential applicability of the framework in real-world clinical settings.

Overall, the findings of this study suggest that optimized ensemble learning approaches can provide effective, interpretable, and reliable tools for Alzheimer's disease diagnosis. Future work may focus on validating the proposed framework on external datasets, extending the approach to multi-stage disease classification, and incorporating longitudinal data to further enhance early detection and disease progression analysis.

References

Aydin, A., & Avaroğlu, E. (2024). Contact classification for human–robot interaction with densely connected convolutional neural network and convolutional block attention module. Signal, Image and Video Processing, 18(5), 4363-4374. https://doi.org/10.1007/s11760-024-03078-4

Aydin, A., Salur, M. U., & Aydin, İ. (2021). Fine-tuning convolutional neural network based railway damage detection. IEEE EUROCON 2021-19th International Conference on Smart Technologies, 216-221. https://ieeexplore.ieee.org/abstract/document/9535585/

Hechkel, W., & Helali, A. (2025). Unveiling Alzheimer's disease early: A comprehensive review of machine learning and imaging techniques. Archives of Computational Methods in Engineering, 32(1), 471-484. https://doi.org/10.1007/s11831-024-10179-3

Jenber Belay, A., Walle, Y. M., & Haile, M. B. (2024). Deep Ensemble learning and quantum machine learning approach for Alzheimer's disease detection. Scientific Reports, 14(1), 14196. https://doi.org/10.1038/s41598-024-61452-1

Kaur, A., Mittal, M., Bhatti, J. S., Thareja, S., & Singh, S. (2024). A systematic literature review on the significance of deep learning and machine learning in predicting Alzheimer's disease. Artificial Intelligence in Medicine, 154, 102928. https://doi.org/10.1016/j.artmed.2024.102928

Rabie El Kharoua. (2024). Alzheimer's Disease Dataset. https://doi.org/10.34740/KAGGLE/DSV/8668279

Raza, H. A., Ansari, S. U., Javed, K., Hanif, M., Mian Qaisar, S., Haider, U., Pławiak, P., & Maab, I. (2024). A proficient approach for the classification of Alzheimer's disease using a hybridization of machine learning and deep learning. Scientific Reports, 14(1), 30925. https://doi.org/10.1038/s41598-024-81563-z

Sheng, J., Zhang, Q., Zhang, Q., Wang, L., Yang, Z., Xin, Y., & Wang, B. (2024). A hybrid multimodal machine learning model for Detecting Alzheimer's disease. Computers in Biology and Medicine, 170, 108035. https://doi.org/10.1016/j.compbiomed.2024.108035

Singh, S. G., Das, D., Barman, U., Saikia, M. J., Singh, S. G., Das, D., Barman, U., & Saikia, M. J. (2024). Early Alzheimer's disease detection: a review of machine learning techniques for forecasting transition from mild cognitive impairment. Diagnostics, 14(16). https://doi.org/10.3390/diagnostics14161759

Tarikul Islam, M., Aktaruzzaman, M., Saif, A., Riyad Hasan, A., Hasan Sourov, M. M., Sikdar, B., Rehman, S., Tabassum, A., Abeed-Ul-Haque, S., Hasan Sakib, M., Alam Muhib, M. M., Ahasan Setu, M. A., Tasnim, F., Rayhan, R., M. Abdel-Daim, M., & Obayed Raihan, M. (2024). Identification of acetylcholinesterase inhibitors from traditional medicinal plants for Alzheimer's disease using in silico and machine learning approaches. RSC Advances, 14(47), 34620-34636. https://doi.org/10.1039/D4RA05073H