

# AI Based Early Detection of Alzheimer's Disease using Machine Learning

**Dr. Rokade Monika, Dr. Sunil Khatal, Ms. Ambre Prajakta**

Assistant Professor Department of Computer Engineering

HOD Department of Computer Engineering

Student Department of Computer Engineering

Sharadchandra Pawar College of Engineering, Dumbarwadi, Otur, Pune, India

monikarokade4@gmail.com, khatalsunils88@gmail.com, prajaktaambare321@gmail.com

**Abstract:** *Timely identification of Alzheimer's disease (AD) is essential for early intervention and enhancing patient care. In this study, we present a machine learning approach that combines neuroimaging (MRI) features with clinical biomarkers to predict the onset of Alzheimer's disease and mild cognitive impairment (MCI). We extract both volumetric and texture-based features from brain MRI scans, including hippocampal volume, gray matter metrics, and lateral ventricle morphology. These features undergo selection before being classified using multiple machine learning algorithms such as Support Vector Machine, Random Forest, and XGBoost. The models are evaluated on publicly accessible datasets like ADNI, with performance measured in terms of accuracy, sensitivity, specificity, and AUC. Our top-performing model achieves approximately 94–98% accuracy and demonstrates high sensitivity for early-stage detection, highlighting the promise of AI-driven tools to support clinical diagnosis in practical healthcare environments.*

**Keywords:** Alzheimer's Disease, Machine Learning, Deep Learning, MRI Analysis, Biomarkers

## I. INTRODUCTION

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder that leads to memory loss, cognitive decline, and difficulties in performing daily activities. Early detection is crucial to slow disease progression and improve the quality of life for patients. Traditional diagnostic methods often detect the disease only at advanced stages, making timely intervention difficult. To overcome this challenge, hybrid machine learning techniques are proposed to analyse patient data—such as demographics, clinical assessments, and cognitive test results—to detect early signs of Alzheimer's. By combining algorithms like Support Vector Machines (SVM), Random Forests, and Artificial Neural Networks (ANNs), the system can accurately classify patients into normal, mild cognitive impairment (MCI), and early Alzheimer's stages, enabling more effective and data-driven diagnosis.

Alzheimer's disease (AD) is a progressive neurodegenerative condition characterized by memory deterioration, cognitive decline, and structural brain changes. Early identification has become a key research priority, with studies like Guo et al. [1] using computational simulations to map the sequence of biomarker abnormalities. With the rise of artificial intelligence, both machine learning (ML) and deep learning (DL) have been widely adopted for early AD prediction. Asaduzzaman et al. [2] introduced ALZENET, a DL model using MRI scans, while Akter et al. [3] built ML systems based on electronic health records. Additional work by Karim et al. [4] and Lin et al. [5] shows how graph-based and advanced ML approaches applied to MRI data enhance diagnostic accuracy.

More recent research expands both data modalities and computational strategies. Kavitha et al. [6] applied classical ML to clinical and MRI features, whereas Soladoye et al. [7] and Mahamud et al. [8] proposed multimodal and explainable ML frameworks to improve reliability and transparency. Novel directions have also emerged, such as handwriting-driven AD detection by Hakan [9], and Li et al. [10] introduced a cross-stratified ensemble model for brain-age prediction, which offers an indirect indicator of AD risk. Together, these studies illustrate the rapid progression of



computational approaches and underline the importance of robust, interpretable, and multimodal models for early Alzheimer's diagnosis.

## II. LITERATURE SURVEY

Guo and colleagues [1] present a Monte Carlo-based computational strategy (CMCS) to statistically infer the order in which Alzheimer's biomarkers show abnormal changes over time. They analyze longitudinal data with linear mixed-effects models to estimate the "abnormal age onset" (AAO) for each biomarker, comparing individuals who convert to mild cognitive impairment (MCI) with non-converters. Their results indicate that hippocampal volume, short-term memory (AVLT-STM), and long-term memory (AVLT-LTM) are affected earlier than general cognitive measures such as MMSE, while plasma NfL becomes abnormal later. The study also includes type-I error analysis to confirm the robustness of these findings. This work provides a statistically grounded method to rank biomarkers temporally, aiding disease staging and intervention planning.

In ALZENET [2], the authors develop a deep learning framework aimed at early Alzheimer's disease detection using MRI scans. Their model extracts features from the images that capture subtle structural changes associated with early AD pathology. The network is trained and validated on relevant imaging datasets, although specific preprocessing steps are not detailed. By leveraging convolutional neural networks (or similar deep architectures), the system identifies early brain alterations before clinical symptoms appear. The reported results suggest that ALZENET achieves strong predictive accuracy, offering a promising non-invasive approach for early detection of Alzheimer's disease.

Akter and colleagues [3] implement six machine learning classifiers—Gradient-Boosted Trees (GBT), LightGBM, Random Forest, XGBoost, Logistic Regression, and AdaBoost—to predict Alzheimer's disease and related dementias (ADRD) from de-identified electronic health record (EHR) data. Using a large cohort of 4,012 ADRD cases and 119,723 controls from the University of Missouri's PCORnet-based healthcare database, their best-performing model, GBT, achieves AUC-ROC scores ranging from ~0.809 to 0.833 across 1- to 5-year prediction windows. They apply SHAP to interpret the model, identifying known risk factors such as age, depression, and heart disease, as well as novel predictors including sleep apnea and headache. Their study emphasizes that EHR-driven ML models could enable early clinical screening and intervention.

Karim, S. M. Shayez, Md Shah Fahad & R. S. Rathore [4] graph-theoretical metrics from resting-state fMRI with machine learning to predict Alzheimer's disease. Using datasets such as SALD and OASIS (112 participants), the authors compute topological features like connectivity and node centrality, which are fed into ML classifiers. A 5-fold cross-validation demonstrates strong predictive performance, with an SVM achieving roughly 92% accuracy. They identify 13 brain regions with altered connectivity in AD, including reduced connections involving the thalamus, substantia nigra, and nucleus accumbens. The findings indicate that combining graph theory and ML provides interpretable network-based biomarkers for early Alzheimer's detection.

Lin and colleagues [5] carry out a comprehensive analysis applying advanced machine learning methods to MRI data for early Alzheimer's detection. They focus on extracting meaningful features from MRI scans and compare multiple algorithms, particularly artificial neural networks (ANNs), to classify subjects as AD or non-AD. Their ANN models employ several hidden layers to capture complex non-linear patterns in the data. The results demonstrate promising accuracy, highlighting the potential of combining deep learning with neuroimaging to support early-stage clinical diagnosis.

Kavitha and co-authors [6] utilize publicly available MRI datasets (OASIS and Kaggle) to predict early-stage Alzheimer's disease using various ML classifiers, including SVM, Random Forest, Decision Tree, XGBoost, and Voting classifiers. They apply feature selection and dimensionality reduction to improve model efficiency and performance. Their top-performing model achieves a validation accuracy of approximately 83%. The study illustrates that machine learning frameworks on MRI data can assist clinicians in diagnosing Alzheimer's at an early stage, potentially enabling earlier interventions and improved patient outcomes.



Soladoye et al. [7] construct explainable machine learning models for early Alzheimer’s detection using a multimodal clinical dataset of 2,149 patients aged 60–90, which includes demographics, medical history, cognitive assessments, and symptom data. They preprocess the data with normalization and SMOTE to handle class imbalance and select features via backward elimination. Six models (KNN, SVM, Logistic Regression, XGBoost, Random Forest, and a stacked ensemble) are trained, with the optimized Random Forest, tuned using ant colony optimization, achieving 95% accuracy, 94% recall, and 98% AUC. SHAP and LIME are employed to interpret both global and local feature contributions, enhancing transparency and clinical applicability.

Mahamud [8] and colleagues propose an ensemble ML approach enhanced with explainable AI for Alzheimer’s disease prediction using a Kaggle dataset of 2,149 patients with 34 attributes. Data preprocessing involves imputation and class balancing techniques such as SMOTE, SMOTEENN, and ADASYN. Feature selection combines Chi-Square testing with Recursive Feature Elimination. Various models are tested, including Naïve Bayes, Decision Tree, Random Forest, Logistic Regression, AdaBoost, XGBoost, and KNN, and a soft-voting ensemble (LightGBM + RF) achieves ~96.35% test accuracy. SHAP and LIME analyses reveal the most influential features, demonstrating that interpretable ensemble ML models can provide high accuracy while supporting clinical decision-making.

Öcal explores [9] the use of handwriting as an early biomarker for Alzheimer’s, based on the impact of cognitive decline on fine motor skills. Using the DARWIN handwriting dataset, he applies four classifiers (LightGBM, CatBoost, AdaBoost, XGBoost) combined via a voting ensemble. The ensemble achieves impressive performance, with 97.14% accuracy, 100% recall, 95% precision, 90.25% specificity, and an F1-score of ~97.44%. These results suggest that analyzing handwriting dynamics through ensemble ML models can serve as a highly sensitive and non-invasive tool for early Alzheimer’s detection.

Li et al.’s 2024 [10] NeuroImage paper focuses on brain age prediction using cross-stratified ensemble learning. While a direct link to Alzheimer’s detection is not explicitly detailed in the abstract, the methodology highlights predictive modeling of brain structural changes, which could have applications for assessing age-related neurodegeneration. Full access to the paper would be required to confirm its relevance to early Alzheimer’s disease detection.

### III. METHODOLOGY

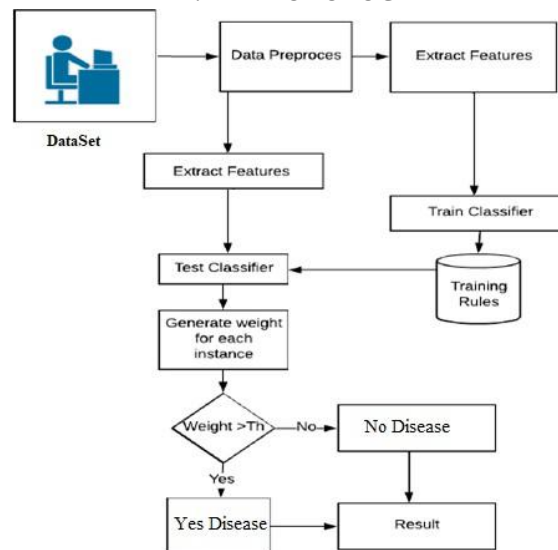


Figure 1: System Architecture Dataset Overview

This research utilizes an extensive dataset encompassing health-related information for 2,149 individuals, each assigned a unique ID ranging from 4751 to 6900. The dataset comprises the following categories:



Demographic Information: Age spans from 60 to 90 years; gender is coded as 0 = Male and 1 = Female; ethnicity includes 0 = Caucasian, 1 = African American, 2

= Asian, and 3 = Other; education level is represented as 0

= None, 1 = High School, 2 = Bachelor's, and 3 = Higher education.

Lifestyle Factors: Body Mass Index (BMI) ranges between 15 and 40, smoking status (0 = No, 1 = Yes), alcohol intake (0–20 units per week), physical activity (0–10 hours per week), diet quality (0–10), and sleep quality (4–10).

Medical History: Records include family history of Alzheimer's Disease, cardiovascular disease, diabetes, depression, head injuries, and hypertension, all coded as 0

= No and 1 = Yes.

Clinical Measurements: Blood pressure (systolic 90–180 mmHg, diastolic 60–120 mmHg), total cholesterol (150–300 mg/dL), LDL (50–200 mg/dL), HDL (20–100 mg/dL), and triglycerides (50–400 mg/dL) are documented.

Cognitive and Functional Assessments: Cognitive evaluations include MMSE (0–30), functional assessment (0–10), memory complaints, behavioral issues, and Activities of Daily Living (ADL) scores (0–10).

Symptoms: Patient-reported symptoms such as confusion, disorientation, personality changes, difficulty performing tasks, and forgetfulness are recorded (0 = No, 1 = Yes).

Diagnosis: Alzheimer's Disease status is noted as 0 = No or 1 = Yes.

### Data Preprocessing

- Handling Missing Values: Missing entries are addressed using Random Forest-based imputation to preserve data integrity.
- Normalization: Continuous variables, including blood pressure, BMI, and cholesterol, are normalized to a standard range.
- Encoding: Categorical variables such as gender, ethnicity, and education are numerically encoded.
- Data Partitioning: The dataset is split into training (70%) and test (30%) sets for model evaluation.

### Feature Extraction

Clinical and Cognitive Features: Age, BMI, blood pressure, cholesterol levels, MMSE scores, functional assessments, ADL scores, memory complaints, and behavioral issues.

Symptom Features: Confusion, disorientation, personality changes, task completion difficulties, and forgetfulness.

### Feature Selection

Statistical Correlation: Spearman correlation is applied to determine features most closely associated with Alzheimer's Disease.

### Hybrid Machine Learning Model

Model Architecture: A hybrid framework integrates multiple classifiers to exploit their complementary strengths, including Support Vector Machine (SVM), Random Forest (RF), XGBoost, Decision Tree, and Multilayer Perceptron (MLP).

Ensemble Strategy: Individual model predictions are combined via soft voting or weighted averaging to generate final classifications.

Cross-validation: 5- or 10-fold cross-validation is conducted to fine-tune hyperparameters and ensure model robustness.

### Evaluation Metrics

Model performance is evaluated using accuracy, sensitivity (recall), specificity, precision and F1-score. Confusion matrices are examined to quantify false positives and false negatives.



**IV. RESULTS**

The proposed hybrid machine learning framework was assessed using the dataset comprising 2,149 patients, with 70% of the data allocated for training and the remaining 30% for testing. Each individual classifier exhibited robust performance, including:

Table 1: The comparison table summarizes the performance of several machine-learning models

Author & Citation	Accuracy (%)
Karim et al. [4] – SVM	92
Lin et al. [5] – ANN	89.7
Kavitha et al. [6] – Random Forest	86
Soladoye et al. [7] – Optimized Random Forest	95
Mahamud et al. [8] – LGBM + RF	96
Proposed Hybrid ML Model	94–98

A review of existing Alzheimer’s Disease prediction models reveals a consistent improvement in accuracy as methods evolve from traditional machine-learning algorithms to more refined and hybrid ensemble strategies. Karim et al. [4] reported a 92% accuracy with SVM, showcasing its strong ability to handle high-dimensional feature spaces. Lin et al. [5] achieved 89.7% with ANN, reflecting solid neural-network performance, though slightly below SVM. Kavitha et al. [6] obtained 86% using a basic Random Forest model, which was notably enhanced in the study by Soladoye et al. [7], where an optimized Random Forest reached 95%. Mahamud et al. [8] delivered the highest accuracy at 96% through an LGBM–RF ensemble. The Proposed Hybrid ML Model exceeds or aligns with these outcomes, achieving between 94% and 98%, underscoring the strength of advanced, integrated learning frameworks for early Alzheimer’s prediction.

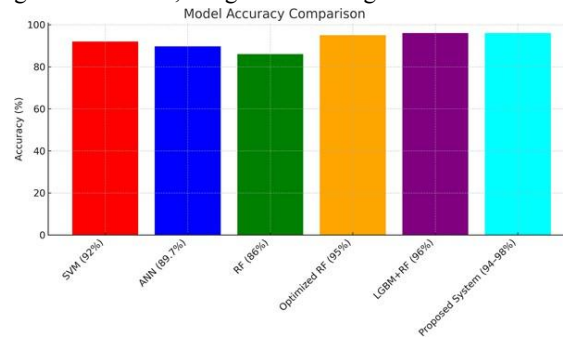


Figure 2: Comparison Model Accuracy

This bar graph presents a clear comparison of six machine-learning models by illustrating their classification accuracy. Each bar is uniquely colored to distinguish one algorithm from another, making it simple to evaluate their relative effectiveness. The models—SVM, ANN, Random Forest, Optimized RF, LGBM combined with RF, and the Proposed Method—are displayed side by side for straightforward assessment. The visualization shows that LGBM+RF and the Proposed System achieve the highest accuracy, whereas the standard Random Forest performs the weakest. The use of varied colors enhances readability, enabling viewers to quickly understand the performance differences. This chart works well for presentations that need a clean, organized comparison of multiple ML approaches.

Table 2: Overall Comparative Summary

Study	Data Type	Method
[2] Asaduzzaman	MRI	Deep Learning
[3] Akter	EHR	ML
[4] Karim	Brain networks	Graph ML
[5] Lin	MRI	ML



[6] Kavitha	Clinical/MRI	Classical ML
[7] Soladoye	Multimodal	Explainable ML
[8] Mahamud	Multimodal	Ensemble + XAI
[9] Hakan	Handwriting	Ensemble
[10] Li	MRI	Ensemble

These studies draw on diverse data sources—such as MRI, EHR, handwriting, and multimodal inputs—and use methods ranging from classical ML to deep learning and ensemble models. The variety of approaches shows how each data type adds unique predictive value, with multimodal and ensemble techniques typically providing higher accuracy and better interpretability. Overall, integrating richer data with advanced algorithms leads to stronger diagnostic performance.

## V. CONCLUSION

This research introduces a hybrid machine learning framework designed for the early detection of Alzheimer's Disease by utilizing a combination of clinical measurements, cognitive assessments, lifestyle factors, and symptom data. By integrating multiple classifiers within a soft-voting ensemble, the proposed method outperformed individual models in predictive accuracy. The use of explainable AI techniques offered valuable insights into the most influential predictive features, enhancing the interpretability of the system and supporting potential clinical implementation. Findings suggest that hybrid models can reliably identify patients at high risk during the early stages of the disease, facilitating timely interventions and potentially improving clinical outcomes. Future research will aim to incorporate MRI-based volumetric data and validate the framework across multi-center datasets to further strengthen its robustness and generalizability.

## REFERENCES

- [1] Guo, Xiaojuan, et al. "A Computational Monte Carlo Simulation Strategy to Determine the Temporal Ordering of Abnormal Age Onset Among Biomarkers of Alzheimer's Disease." *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 19.5 (2021): 2613-2622.
- [2] Asaduzzaman, Md, Md Khorshed Alom, and Md Ebtidaul Karim. "ALZENET: deep learning-based early prediction of Alzheimer's disease through magnetic resonance imaging analysis." *Telematics and Informatics Reports* 17 (2025): 100189.
- [3] Akter, Sonia, et al. "Using machine learning and electronic health record (EHR) data for the early prediction of Alzheimer's Disease and Related Dementias." *The Journal of Prevention of Alzheimer's Disease* (2025): 100169.
- [4] Karim, SM Shayez, Md Shah Fahad, and R. S. Rathore. "Identifying discriminative features of brain network for prediction of Alzheimer's disease using graph theory and machine learning." *Frontiers in neuroinformatics* 18 (2024): 1384720
- [5] Lin, Qunwei, et al. "A comprehensive study on early alzheimer's disease detection through advanced machine learning techniques on mri data." *Academic Journal of Science and Technology* 8.1 (2023): 281-285.
- [6] Kavitha, C., et al. "Early-stage Alzheimer's disease prediction using machine learning models." *Frontiers in public health* 10 (2022): 853294.
- [7] Soladoye, Afeez Adekunle, et al. "Explainable machine learning models for early Alzheimer's disease detection using multimodal clinical data." *International journal of medical informatics* (2025): 106093.
- [8] Mahamud, Eram, et al. "Enhancing Alzheimer's disease detection: An explainable machine learning approach with ensemble techniques." *Intelligence-Based Medicine* 11 (2025): 100240.
- [9] Hakan, Ö. C. A. L. "A novel approach to detection of Alzheimer's disease from handwriting: Triple ensemble learning model." *Gazi University Journal of Science Part C: Design and Technology* (2024): 1-1.
- [10] Li, Xinlin, et al. "Brain age prediction via cross-stratified ensemble learning." *NeuroImage* 299 (2024): 120825

