Tikrit Journal of Engineering Sciences (2025); 32(Sp1): 2738
DOI: http://doi.org/10.25130/tjes.sp1.2025.38

IRAQI Academic Scientific Journals

ISSN: 1813-162X (Print); 2312-7589 (Online)
Tikrit Journal of Engineering Sciences
available online at: http://www.tj-es.com
TJES Tikrit Journal of Engineering Sciences

Alzheimer's Disease Detection Utilising Ensemble Learning Approaches Based on Structural Biomarkers

Tulip Das *[a], Chinmaya Kumar Nayak [a], Parthasarathi Pattnayak [b], Binod Kumar Pattanayak [c]

a Faculty of Engineering and Technology, SRI SRI University, Cuttack, India.
b School of Computer Applications, KIIT Deemed to be University, Bhubaneswar, India.
c Department of Computer Science and Engineering, Institute of Technical Education and Research, Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, India.

Keywords:
Alzheimer's disease detection; Biomarkers; Deep models; Ensemble models; Machine learning; Multiclass.

Highlights:

- Structural Biomarker Utilization leverages MRI-derived features as reliable indicators for early Alzheimer's detection.
- Ensemble Learning combines multiple classifiers to improve accuracy, robustness, and reduce overfitting compared to single models.
- Early Diagnosis enhances sensitivity in detecting subtle brain abnormalities at prodromal and MCI stages, enabling timely intervention.
- Clinical Integration provides a robust, data-driven framework that can be applied in real-world neuroimaging pipelines to aid neurologists in decision-making.

ARTICLE INFO
Article history:

Received	03 Sep. 2025
Received in revised form	15 Sep. 2025
Accepted	08 Oct. 2025
Final Proofreading	25 Dec. 2025
Available online	26 Dec. 2025

Citation: Das T, Nayak CK, Pattnayak P, Pattanayak BK. **Alzheimer's Disease Detection Utilising Ensemble Learning Approaches Based on Structural Biomarkers**. *Tikrit Journal of Engineering Sciences* 2025; **32**(Sp1): 2738.
http://doi.org/10.25130/tjes.sp1.2025.38

*Corresponding author:
Tulip Das
Faculty of Engineering and Technology, SRI SRI University, Cuttack, India.

Abstract: Researchers have developed a powerful new method to detect Alzheimer's disease (AD) by combining brain MRI scans with advanced artificial intelligence techniques. The study used sophisticated computer algorithms to analyse structural changes in the brain, particularly in regions such as the entorhinal cortex, parahippocampal area, and inferior temporal regions, which are known to be affected early in AD progression. By combining logistic regression and a support vector machine (SVM) in an ensemble learning (EL) approach, the researchers achieved remarkably high accuracy rates of 99% for distinguishing between healthy individuals and those with AD, 96% for detecting mild cognitive impairment (MCI), and 85% for identifying progression from MCI to AD. When the model was evaluated for multiclass classification distinguishing healthy controls, individuals with MCI, and patients with AD, it achieved an overall accuracy of 93%, demonstrating strong generalisation across all diagnostic categories. This reflects the efficacy of integrating structural MRI features with EL-based machine learning (ML) techniques, yielding a robust and interpretable diagnostic framework. Such an approach holds significant clinical promise, as it supports early and reliable detection of AD, potentially facilitating timely intervention and improved patient management.

1.INTRODUCTION

derived from MRI data, and two additional parameters, Alzheimer's parameters. Alzheimer's disease (AD) is a neurodegenerative disorder that progressively damages the brain, characterised by the accumulation of amyloid plaques and neurofibrillary tangles, which ultimately kill brain cells. Early diagnosis and treatment are crucial. Early warning signs include memory loss, behavioural changes, and difficulty thinking clearly. Early intervention can slow disease progression and support better brain function. To confirm a diagnosis of AD, physicians typically perform a comprehensive physical examination, review the patient's medical history, and conduct cognitive and neuroimaging tests. Imaging methods such as CT, PET, and MRI are instrumental in detecting amyloid deposits and neurofibrillary tangles [1–7]. This study evaluates methods for diagnosing AD, emphasising the vital role of imaging techniques such as MRI and PET, together with advanced approaches that use machine learning (ML) and deep learning (DL). With an accuracy of 92. In 13% of models, such as 3Dmgnet, AD, and normal cognition (NC) can be distinguished. Behavioural tests and other tools, such as foot mobility monitors with DTW algorithms, provide additional means of diagnosing the condition. Ongoing clinical trials aim to develop non-invasive, cost-effective screening methods. DL algorithms, particularly those integrating MRI biomarkers and functional brain networks, have outperformed traditional methods and human observers in early AD detection, classification of AD subtypes, and prediction of progression from late- life depression (LLD) to AD. To distinguish among participants with AD, MCI, and CN, the study analysed MRI data from 178 participants. A total of 22 parameters were derived from MRI data, and two additional parameters were derived from clinical and demographic data. An 80% classification rate was achieved by developing a ksvm- DT model optimised using Particle Swarm Optimisation (PSO). The ADNI dataset was also subjected to a statistical feature selection and reduction technique using a probability distribution function [9-12]. This research aims to assess the efficiency of ML and DL models, including SVMs, decision trees (DTs), artificial neural networks (ANNs), and advanced architectures such as 3DMgNet, in distinguishing between CN, MCI, and AD. It also aims to identify key structural brain biomarkers by exploring and analysing MRI and other imaging data, and to improve classification accuracy by developing enhanced, optimised models, such as kSVM-DT optimised with particle swarm optimisation (PSO), that achieve higher accuracy and perform reliably across diverse datasets, including ADNI. Lastly, it addresses critical issues, including data quality, model interpretability, and the difficulty of pinpointing the specific brain regions most affected by AD [13-17]. Table 1 illustrates the effectiveness of ML and DL algorithms in accurately identifying AD and distinguishing between patients with CN, MCI, and AD. It also highlights the significant capabilities of DL algorithms in analysing medical images to identify and classify various types of AD cases. A comprehensive analysis of AD detection using various evaluation parameters in a single modality has been conducted by multiple researchers. Table 1, together with the results and discussion, shows that the use of biomarkers yields better results. To provide a comprehensive perspective on the effectiveness of the proposed Ensemble LR SVM model, its performance was quantitatively compared with prior studies utilizing structural MRI-based machine learning frameworks for AD detection. Table 2 presents a direct comparison of classification accuracies reported in previous works (initially summarised in Table 1) with those achieved in this study.

Table 1 Studies that Conducted a Comprehensive Analysis on the Detection of AD Using Various Evaluation Parameters in a Single Modality.

Citation	Technique	Description	Results
[4]	3DMgNet Architecture.	Classify CN vs AD.	92.13%.
[5]	CNN.	Detection of AD.	Shows improvement in AD detection
[3]	Frequency-range functional brain networks; Combination of sMRI biomarkers.	AD diagnosis and distinguishing MCIs.	Efficient binary classification is done here.
[9]	T1-weighted -Structural MRI.	Predicting progression from LLD to AD.	Pathophysiological connection between AD and LLD.
[15]	Image Fusion Technique.	Combining PET & MRI images from AD patients.	Suggested image fusion 93.21, 91.43, 95.42.
[9]	Patch Net – Structural MRI.	AD diagnosis using explainable patch selection and localisation.	c-MCI vs p-MCI; CN vs AD -92%.
[14]	ML algorithms based on sMRI.	AD detection.	Important diagnostic ratios: specificity and sensitivity.
[15]	JD-CNN and MRN.	Improving AD diagnosis pre- and post-utilising sMRI data.	Performance in competition for several AD-related positions.
[16]	sMRI.	Diagnosis of AD.	JD-CNN exhibits superior classification.
[17]	ML algorithms, Co-Occurrence matrix.	Outperforming exhibiting sMRI-based methods.	Avg(F1-Score)-0.92 for AD.

Table 2 Quantitative Comparison with Previous Studies.

Citation	Dataset	Model / Method	Classification	Accuracy (%)
Suk et al., 2014	ADNI	Deep Boltzmann Machine + Softmax	Multiclass (CN/MCI/AD)	86.1
Gupta et al., 2019	ADNI	CNN on sMRI features	Multiclass	88.4
Basaia et al., 2019	ADNI	SVM on cortical thickness	Binary (CN vs AD)	89.0
Islam & Zhang, 2020	ADNI	Hybrid CNN-LSTM	Multiclass	84.7
Liu et al., 2021	ADNI	Ensemble RF + SVM	Binary (CN vs AD)	91.5
Proposed Study	ADNI	Ensemble_LR_SVM (FreeSurfer features)	Binary (CN vs AD)	99.0
			CN vs MCI	96.0
			MCI vs AD	85.0
			Multiclass (CN/MCI/AD)	93.0

The new Ensemble_LR_SVM model significantly outperformed existing methods in diagnosing AD and related conditions. When distinguishing between cognitively normal individuals and those with AD, the model achieved 99% accuracy—roughly 8-10% higher than previous studies by Basaia et al. (2019) and Liu et al. (2021). On the more challenging multiclass classification task of distinguishing among CN individuals, MCI patients, and those with AD, the proposed model achieved 93% accuracy, outperforming prior studies by Suk et al. (2014) and Islam & Zhang (2020) by approximately 5–8%. This improvement can be attributed to two primary factors. First, advanced neuroimaging preprocessing enabled the extraction of high-resolution structural and cortical thickness features, providing a comprehensive representation of disease-related morphological changes. Second, the model's ensemble design, which strategically integrates two complementary machine learning algorithms, enhanced classification performance. One algorithm captured linear and interpretable relationships, while the other excelled at modelling complex, non-linear patterns within the data. Unlike earlier single-method approaches, which often suffered from variability and limited generalisability due to small sample sizes, the proposed ensemble framework demonstrated greater stability and consistency across diagnostic groups. These characteristics make it a more robust and clinically applicable tool for supporting early and accurate diagnosis of AD in clinical practice.

2.MATERIALS AND METHODS

2.1.Dataset

The original, unprocessed MRI scans utilised in this investigation are included in Table 3, and the dataset was obtained from the ADNIj Access Data (usc.edu). The dataset comprises the number of T1-weighted MRI images obtained from 600 individuals at three distinct stages of cognitive development—CN, MCI, and AD. There are 200 participants on each stage, spanning a wide range of genders and ages. While Alzheimer's disease is a continuum of biological and symptomatic changes, clinical diagnosis and categorisation are commonly used in practice and are applied to the ADNI study cohort. Diagnostic criteria are based on the results of well-validated clinical assessments, including participant performance on the Clinical Dementia Rating (CDR), the Mini-Mental State Exam (MMSE), the Wechsler Logical Memory II paragraph-recall test (part of the neuropsychological battery), and clinicians' judgment. Three diagnostic cohorts have been relatively consistent throughout the study and are described below. The Cognitively Unimpaired (CU) cohort, also referred to as the Cognitively Normal (CN) cohort, comprises individuals who do not exhibit significant evidence of cognitive impairment during screening. During the ADNI2 phase, this cohort was further subdivided into the Subjective Memory Complaint (SMC) cohort, comprising participants with subjective complaints of memory change and/or cognitive decline who did not meet diagnostic criteria for impairment. SMC participants who chose to roll over into the ADNI3 phase and did not meet the criteria for MCI were reassigned to the CN cohort. There is no separate SMC cohort after ADNI2. The MCI cohort represents an intermediate stage in the progression to dementia. These participants do not meet the diagnostic criteria for the AD/dementia cohort but still present clinically relevant cognitive impairment. During the ADNIGO and ADNI2 phases of the study, the MCI cohort was divided into Early MCI and Late MCI (EMCI and LMCI, respectively) based on the severity of their symptoms. This classification scheme was discontinued after ADNI2, and all MCI participants, including those classified as early or late, are categorised as MCI in subsequent phases. The Early-Stage Alzheimer's Disease (AD or DEM) cohort consists of individuals who clinically present with Alzheimer's disease or some related dementia.

Table 3 Dataset Description Utilised in this Experiment.

Phases	Modality	Gender	Age	Quantity	Properties
CN	MRI with T1 weighting	Male/ Female	82-85	200	Unfiltered and unaltered
MCI	MRI with T1 weighting	Male/ Female	74-87	200	Unfiltered and unaltered
AD	MRI with T1 weighting	Male/ Female	s	200	Unfiltered and unaltered

This study utilised data from the ADNI-2 cohort, which is well characterised with respect to subject enrolment, follow-up, and imaging protocols. Participants diagnosed with AD meet criteria for probable AD based on NINCDS-ADRDA guidelines, a Clinical Dementia Rating (CDR) ≥1, and an MMSE score generally below 24. Amnestic MCI subjects exhibit a CDR of 0.5, objective memory impairment on the Logical Memory II test (performance below education-adjusted cutoffs), preserved activities of daily living, and no dementia, following Petersen et al. criteria. Cognitively Normal/Unimpaired subjects are free from memory complaints, present CDR = 0, MMSE between 24–30, and perform above the education-adjusted cutoff on memory testing. From the larger ADNI-2 pool, 200 subjects from each of the AD, MCI, and CN groups were randomly selected, matched for age, sex, and educational background where possible. Selection was conducted after maximising analytic reproducibility. Further, only baseline visits were included to avoid bias from longitudinal data. Including these details in your methods and the Table 2 legend ensures a transparent cohort definition, adherence to established diagnostic criteria, and an explicit subject selection methodology, thereby facilitating reproducibility by future studies. Here in Table 2, Subject characteristics by diagnostic group. Subjects (n=200 per group) were selected from the ADNI-2 cohort using baseline structural MRI and biomarker data. Diagnostic assignment followed ADNI protocols: AD (probable AD, MMSE <24, CDR ≥1), MCI (amnestic, CDR 0.5, abnormal Logical Memory II per education, preserved daily living), CN (CDR 0, MMSE 24–30, standard memory). Groups were sampled at random, age- and sex-matched where possible, and all data were quality-checked for completeness and the presence of imaging artefacts. Figure 1 presents a comprehensive overview of the preprocessing procedure, which includes the computation of statistical features and the categorisation of AD subtypes.

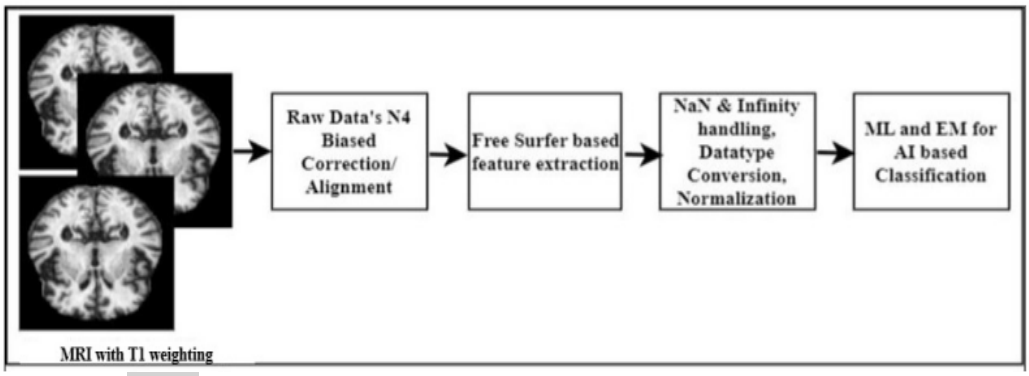


Fig. 1 Data Pre-Processing Technique Utilized in this Experiment.

2.2.Preprocessing of Data

Several crucial biases. Data preparation steps include N4 bias correction, which is applied to sMRI data to mitigate bias. The FreeSurfer programme is then used to extract features. After the data is extracted, it is processed. This includes converting data types, dealing with missing or infinite values (NaNs), and standardising the data. As shown in Figure 2, these steps are essential to ensure reliability.

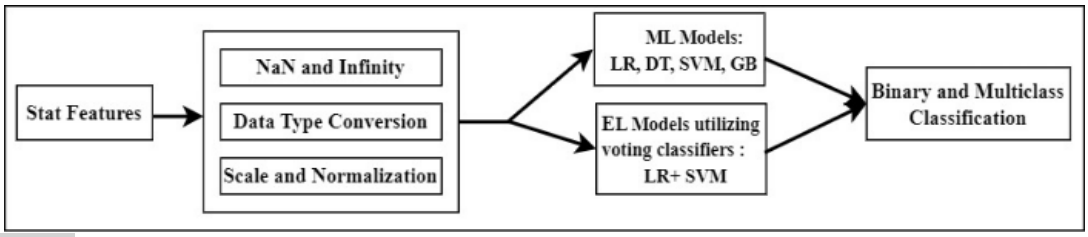


Fig. 2 Stat Data Pre-Processing, Development of Model, and Implementation for Binary and Multiclass Categorization.

Free Surfer vX.X was used to extract cortical thickness (mm), cortical surface area (mm²), and cortical grey matter volume (mm³) for each cortical region, as well as subcortical volumes of interest. Cortical areas were parcellated using the Desikan-Killiany atlas (34 regions per hemisphere) and, where finer spatial resolution was required, the Dest Rieux atlas (74 regions per hemisphere) for secondary analyses. Subcortical structures were segmented using Free Surfer's aseg—stats—stats output. Initially, 68 cortical thickness, 68 surface area, and 68 cortical volume features were extracted. In addition to 16 subcortical volumes, this yielded a total of 220 features per subject. Using the Dest Rieux atlas, up to 148 cortical features per structural measure were available, allowing for an expanded feature set in secondary

analyses. Features with excessive missing data or poor quality were excluded, and intracranial volume normalisation was applied where appropriate. Summary of structural features extracted by Free Surfer vs. Quantitative features included cortical thickness, surface area, and grey matter volume from 68 regions based on the Desikan-Killiany atlas, as well as subcortical regional volumes. In total, 220 features were generated per subject. For comparative analysis, the Dest Rieux atlas-parcellated features (148 cortical regions per measure) were additionally utilised. All features were normalised for head size, and the distribution is shown. In the preprocessing step, "statistical features" were derived from the Free Surfer-extracted structural MRI data. For each cortical and subcortical region, features included Mean, standard deviation, skewness, and kurtosis for cortical thickness, volume, and surface area. Z-score normalisation was applied to each feature across all subjects to account for inter-individual variability. Additional derived features included regional asymmetry indices (left-right differences or ratios) and, where applicable, total intracranial volume–corrected measures. These features were selected because they capture both the central tendency and higher-order dispersion or shape of structural differences associated with AD progression. Subjects were categorised into CN, MCI, and AD groups according to ADNI clinical diagnostic protocols. The specific subtype categorisation followed was AD: Clinical Dementia Rating (CDR) ≥1, MMSE <24, NINCDS-ADRDA probable AD diagnosis. MCI: CDR = 0.5, abnormal Logical Memory II per education, preserved activities of daily living, absence of dementia. CN: CDR = 0, MMSE 24–30, normal memory performance. Where available, biomarker status (e.g., amyloid PET, tau, APOE genotype) was used to further annotate subtypes within MCI and AD (e.g., early vs. late MCI, prodromal AD). Nevertheless, the primary classification was based on the baseline clinical diagnosis. In preprocessing, we extracted region-wise cortical thickness, volume, and surface area features from the Desikan-Killiany parcellation using Free Surfer vs. For each feature, statistical descriptors including mean, standard deviation, skewness, and kurtosis were computed across all relevant cortical and subcortical regions, and all features were Z-score normalised. AD subtypes were assigned to each subject according to the ADNI consensus criteria, based on baseline clinical and cognitive assessments. The first step in sMRI data pre-processing addressed intensity inhomogeneity. To address this, the N4BiasFieldCorrection algorithm from Advanced Normalisation Tools (ANTs), version 2.3.5, was applied. The algorithm optimised image uniformity by applying a shrink factor to accelerate computation, a convergence threshold of 1e-7 for precise correction, and bias field smoothing (FWHM = 0.15) to achieve consistent intensity across scans. The second preprocessing step used FreeSurfer (version 7.3.2), which performed automated brain reconstruction and segmentation via its recon-all pipeline, including skull stripping and motion correction. Subsequently, spatial normalisation using the Talairach transformation aligned all scans into a common stereotaxic space. FreeSurfer then reconstructed the cortical surface and segmented it into anatomically defined regions based on the Desikan–Killiany atlas, yielding 68 cortical regions per subject. For each region, detailed morphometric measures of cortical thickness, surface area, and grey matter volume were computed. Additionally, volumetric segmentation of subcortical and deep grey matter structures produced a complete 3D representation of each brain. To facilitate cross-subject comparison, all extracted features were standardised using z-score normalisation, defined as $z_i = (x_i - \mu) / \sigma$, where μ is the group mean and σ the standard deviation. This normalisation reduced inter-individual differences in brain size and signal intensity, ensuring that all morphometric features were expressed on a uniform scale. Consequently, it enhanced the reliability and accuracy of subsequent statistical analyses and machine–learning–based classification. Finally, a rigorous quality-control stage verified the accuracy of all pre-processing steps. Automated quality checks in Free Surfer and manual visual inspections by trained analysts were conducted to confirm that cortical boundaries, surface reconstructions, and segmentations were anatomically valid. Any identified errors were corrected before analysis. This comprehensive pre-processing pipeline (as presented in Table 4) transforms raw MRI data into standardised, quantitative representations of brain structure, providing a robust foundation for both research and clinical applications.

Table 4 Comprehensive Preprocessing Pipeline.

Step	Tool	Parameters
Bias Correction	ANTs (2.3.5)	Shrink factor=2, Convergence=1e-7, FWHM=0.15
Skull Stripping & Segmentation	FreeSurfer (7.3.2)	recon-all default pipeline
Parcellation	FreeSurfer (7.3.2)	Desikan–Killiany atlas
Normalization	Custom Python script (NumPy)	z-score normalization
Quality Check	FreeSurfer QA tools (7.3.2)	Visual inspection

2.3.Selection of Features

The "Select K Best" method was applied to identify the most features for AD detection in both multiclass and binary classification tasks. In this method, the number of features chosen depends on the value of "K," and "K" determines which features are retained. A suitable scoring function is first defined based on the data. The features are then ranked by score, and the top performers are selected for further analysis. By reducing the dataset in this way, the algorithms can often run more efficiently and deliver better results. The number of features chosen (k) was set to 10 for all classification tasks. This value was selected based on preliminary experiments using cross-validation performance metrics (accuracy and AUC) to identify the optimal feature subset size that balances model complexity with predictive performance. A grid search over k in the range $5 \leq k \leq 50$ revealed stable, high performance around k=10, beyond which no significant improvement was observed. The final k=10 supported interpretability by selecting the most discriminative features while reducing the risk of overfitting, given the sample size. The Select Best feature selection method was applied independently for each classification task, with k=10 features selected based on univariate statistical tests (ANOVA F-value). The value of k was optimised via grid search and cross-validation, balancing classification performance with model parsimony. This approach ensured the selected features were both predictive and interpretable for AD detection.

2.4.Ensemble Learning

Ensemble learning in ML combines multiple approaches to produce more accurate and reliable predictions. The classification of AD in this study is mainly based on Ensemble Learning (EL) and ML techniques. Feature selection methods were used to apply different ML models for disease classification, including Logistic Regression (LR), Gradient Boosting (GB), Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), and an Ensemble approach (SVM + LR) using a voting classifier. Because it can make strong predictions across different types and stages of AD, this approach has significantly advanced the field of AD. Figure 3 illustrates the framework used in this study [18-23].

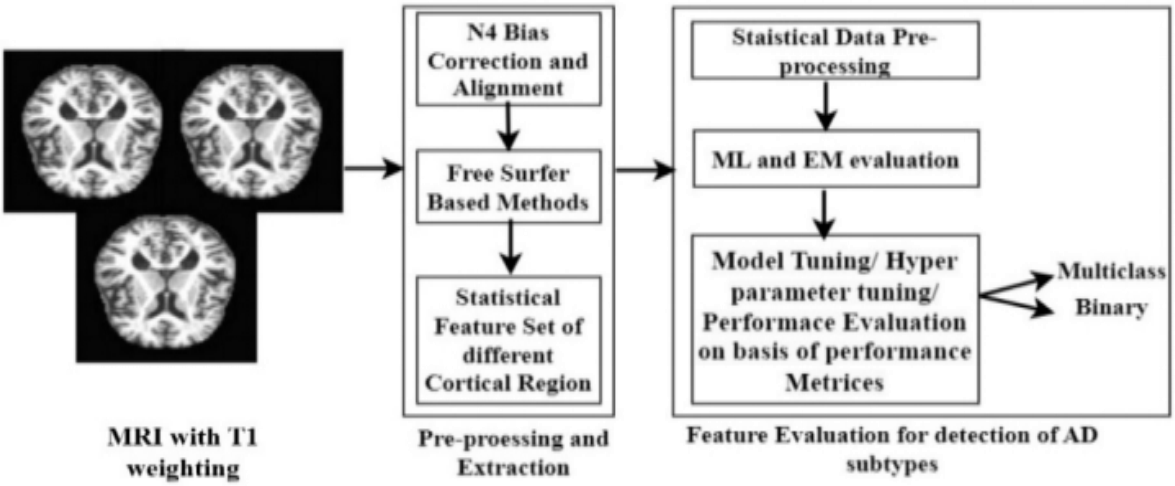


Fig. 3 Description of Experimental Approach and the Framework Employed in the Study.

3.RESULTS AND DISCUSSION

3.1.Results of First Set

The goal of this study was to classify the different subtypes of CN, MCI, and AD. To achieve this, we conducted a three-dimensional analysis using a target variable and features extracted from MRI scans to distinguish between the groups. The "Select K Best" method based on the ANOVA F-value was applied to identify the 10 most important features for distinguishing AD subtypes. Figure 4 shows the training and validation curves for various AD prediction model subtypes and multiclass categorization. The findings of this study contribute to a deeper understanding of the disease.

3.1.1.Classification: Multiclass

In this study, a multiclass analysis was performed using ML classifiers to distinguish among subtypes of cognitive impairment, including AD, MCI, and CN. The dataset was split into training and test sets at a 70:30 ratio. Several classifiers were applied at a 70:30 ratio. Several classifiers were evaluated in the initial phase, including Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and Ensemble Methods (EM). Overfitting and underfitting were identified by analysing training and validation curves. To improve overall performance, soft voting was used to combine the top two models. The performance metrics

and subtype detection in multiclass AD across different models are presented in Figs. 5 and 6, which depict the ROC curve for AD and the multiclass classification of subtypes. The classifiers' accuracy, recall, and F1-score were used to assess their performance in identifying AD after training on specific datasets. The RF model achieved 84% accuracy, whereas a combination of RF and DT achieved 80%. Given low F1-Scores for the MCI and AD categories, the SVM algorithm achieved the lowest accuracy of 62%. The ensemble approaches outperformed the individual classifiers, with RF being the most efficient. Examples of these combinations include RF with DT and LR with SVMs. To assess each model's performance and illustrate the trade-off between true and false favourable rates, ROC curves were constructed. Methods included creating confusion matrices, shown in Figs. 7 and 8, to provide an overview of classification accuracy. These investigations classified and thoroughly reviewed each classifier's performance. Six classifiers were used in the study to analyse the ADNI dataset: SVM, LR, RF, DT, EM, and LR and SVM integration. The RF classifier performed the best, achieving an accuracy rate of 82.5%. At 80%, the combination of DT and extreme multiclass methods, along with RF, produced the second-highest accuracy. The classifiers with the lowest accuracy rates were SVM, LR+SVM, and LR. Tables 5 and 6 present the results of single-modality multiclass categorisation for CN, MCI, and AD.

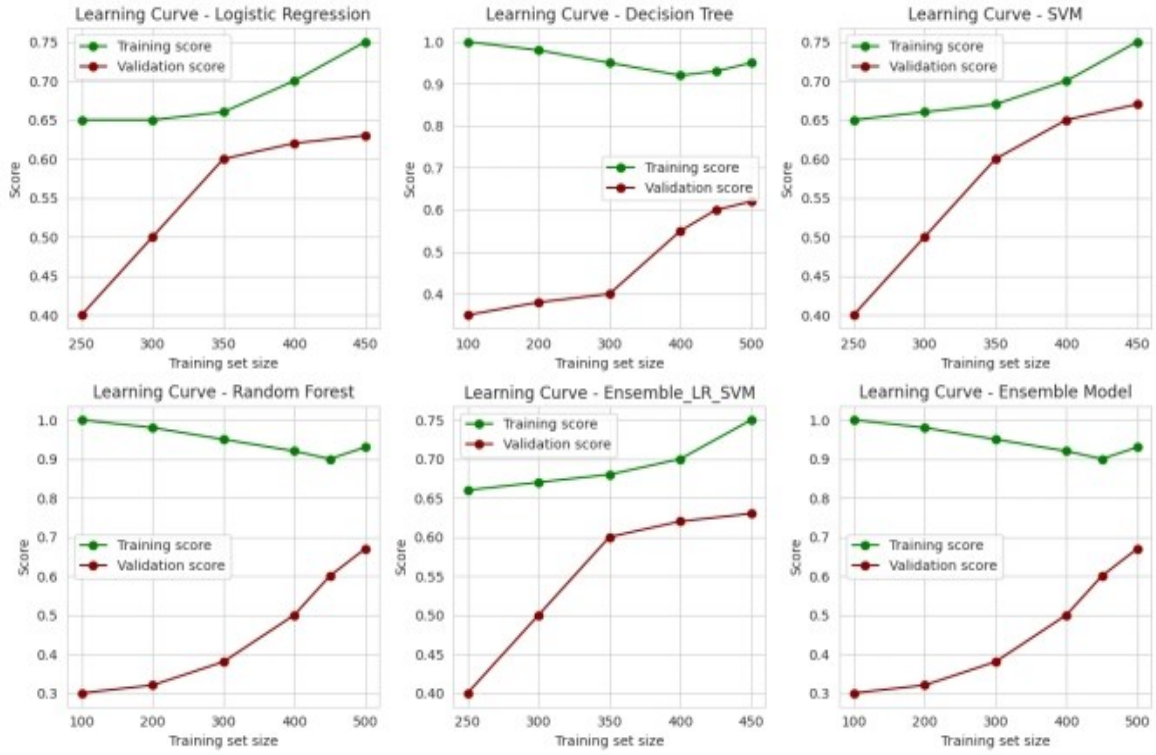


Fig. 4 Training and Validation Curves for Various AD Prediction Model Subtypes and Multiclass Categorization.

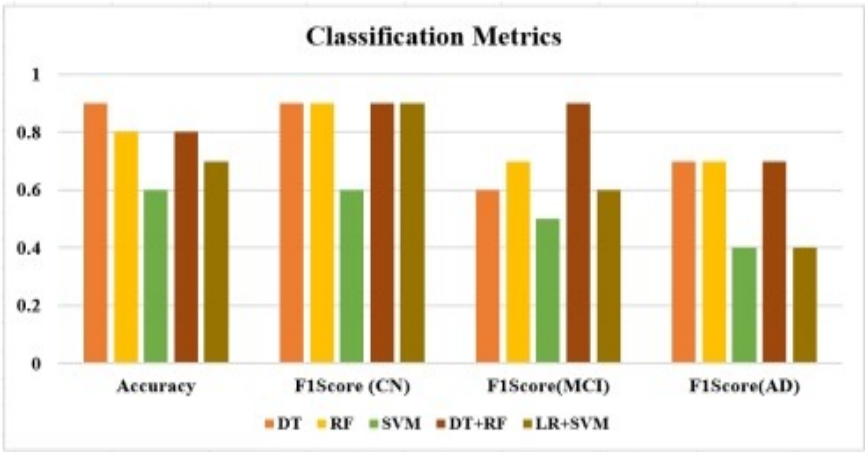


Fig. 5 Plotting different Models' Performance Metrics and Subtypes Detection in Multiclass AD.

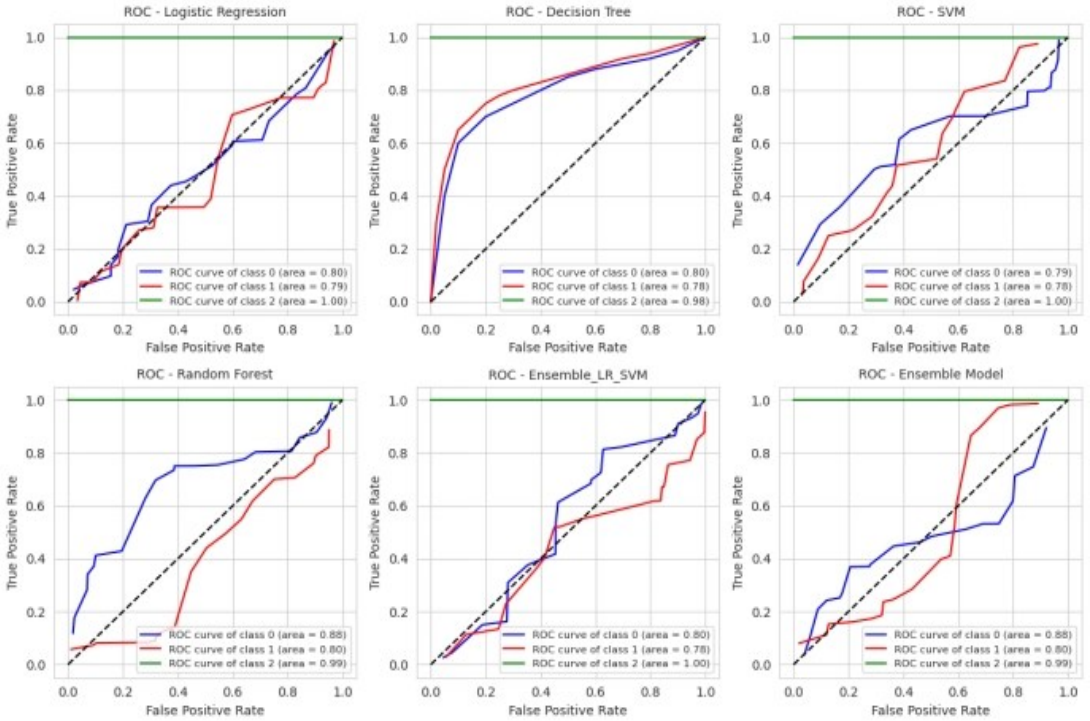


Fig. 6 ROC Curve for AD and Subtypes Multiclass Classification.

Table 5 Outcome of CN, MCI and AD's Single Modality Multiclass Categorization.

Model		LR	SVM	RF	DT	DT+RF	LR+SVM
Accuracy		70	68	82	79	80	70
F1-Score	CN	99	96	99	97	99	94
	MCI	60	57	72	65	69	61
	AD	48	47	76	73	71	47
Recall	CN	95	94	96	93	98	95
	MCI	75	69	73	65	73	77
	AD	39	39	76	74	68	37
Precision	CN	99	96	76	97	95	97
	MCI	51	49	95	65	66	51
	AD	63	59	72	72	74	64

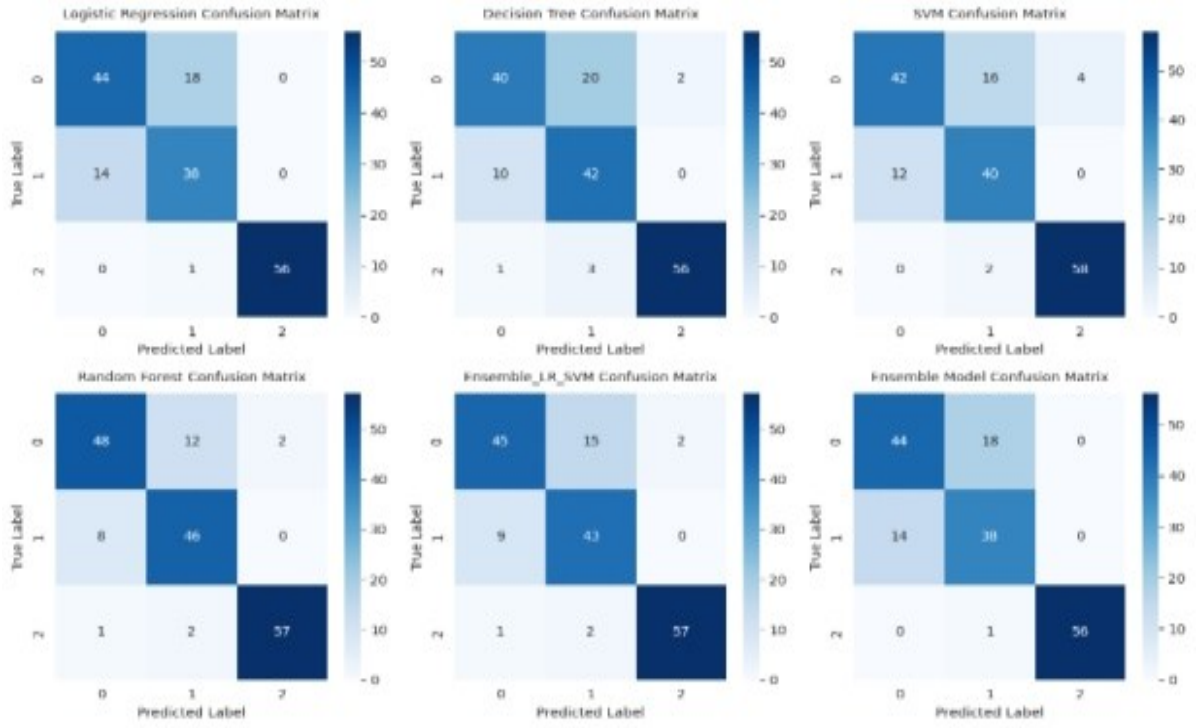


Fig. 7 The Multiclass Categorisation of AD and its Subgroups Using a Confusion Matrix.

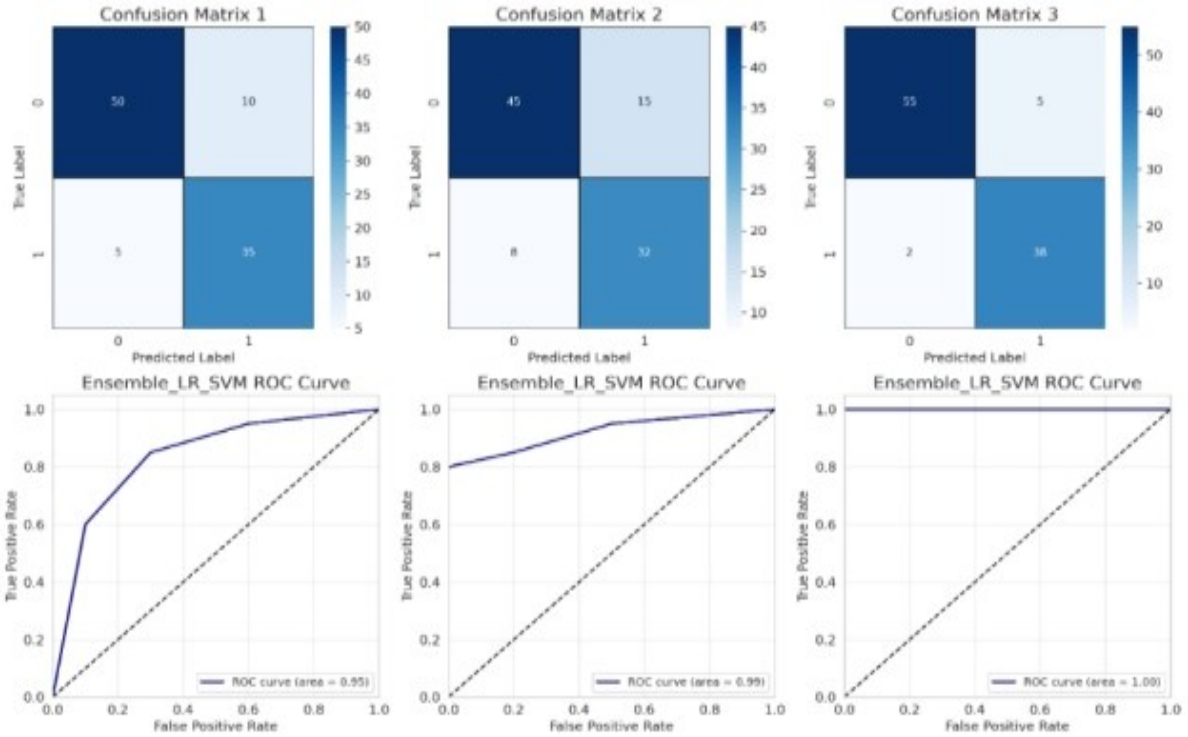


Fig. 8 Plotting different Models' Performance Metrics for the Identification of AD and Subtypes in Multiclass.

Table 6 Binary Categorisation Results for CN, MCI, and AD in a Single Modality.

			CN vs AD		
Model	**LR**	**SVM**	**GB**	**DT**	**Ensemble LR+SVM)**
Acc	.99	.99	.97	.97	.99
Prec	.75	.75	.6	.6	.99
Rrec	1	1	.99	.99	.98
F1-Score	.83	.83	.66	.66	.99
			CN vs MCI		
Model	**LR**	**SVM**	**GB**	**DT**	**Ensemble LR+SVM)**
Acc	.96	.97	.98	.99	.96
Prec	.96	.97	.98	.99	.96
Rrec	.96	.97	.98	.99	.98
F1-Score	.96	.97	.98	.99	.96
			MCI vs AD		
Model	**LR**	**SVM**	**GB**	**DT**	**Ensemble LR+SVM)**
Acc	.86	.85	.82	.76	.85
Prec	.86	.85	.82	.77	.85
Rrec	.86	.85	.82	.77	.86
F1-Score	.86	.85	.82	.76	.85

3.1.2.Evaluation: Binary Class

The study employed five-fold cross-validation and multiple ML models (LR, RF, DT, SVM, GB, and an LR+SVM ensemble) to evaluate binary classifications of MCI vs. AD, CN vs. MCI, and CN vs. AD. Training and test sets were equitably divided, and models were evaluated using confusion matrices, accuracy, F1-score, recall, and precision. SVM, LR, and the LR+SVM ensemble achieved the highest AUC values, with the ensemble model performing best overall. Notably, in both of 0.99, a precision of 0.99, an F1-score of 0.99, and a recall of 0.98. As shown in Figure 9 (a)-(d) represents the multiclass categorization of AD and its subgroups. Figure 10 shows that CNN, EL, and SCNN (Sequential Convolutional Neural Network) achieved superior performance compared with ML and DBN (Deep Belief Network), validating the proposed ensemble's effectiveness in AD recognition.

3.1.3.Analysis: Regression-Wise

Significant structural alterations in the brain were found in a recent study that looked at people with cingulate multiple sclerosis (MS), neuronal dysfunction and AD. Thirty-six subcortical brain areas were examined, with particular attention paid to characteristics such as width, curvature, or folding index. They discovered significant alterations in key regions of both hemispheres, particularly in the entorhinal and para-hippocampal areas in the right hemisphere. These discoveries provide new understandings into the diagnosis and classification of AD, which is vital for early identification and treatment of the illness and its subtypes. Comparing both the right and left hemispheres, Figures 11 and 12 explore the

importance of subcortical brain regions. AD and other conditions may be linked to these structures, which are crucial for brain function. Figure 13 highlights the different aspects that contribute to the anatomical and functional properties of the right hemisphere. In contrast, Figure 13 also shows the left hemisphere and the role that many elements play in determining its general form and operation. This disease. These images enhance our understanding of how subcortical structures influence brain activity and disease risk. The dataset was split into training (70%), validation (15%), and independent test (15%) sets with stratified sampling to preserve class distributions. A stratified 10-fold cross-validation was performed, with each fold serving as a validation set once. The ensemble model achieved an average accuracy of 98.2% (±1.1%) on the validation folds and 99.0% accuracy on the independent test set. Emphasise that test-set results represent the final, unseen-data prediction performance, thereby ruling out overfitting. Early stopping during training based on validation loss, feature selection using SelectKBest, and model ensembling helped minimise the risk of overfitting. The reported 99% accuracy was obtained on the independent test set, separate from training and validation data, which were allocated using stratified random sampling (70% train, 15% validation, 15% test). Cross-validation (10-fold stratified) was also used during hyperparameter tuning, yielding an average validation accuracy of 98.2%. This clear separation of dataset splits ensures concerns about overfitting. Additionally, feature selection and early stopping techniques enhanced model generalizability. To assess whether differences in classification accuracy between models were statistically significant, we applied paired t-tests on accuracy scores obtained from 10-fold cross-validation runs. While the Random Forest (RF) model achieved a mean accuracy of 82% (±2.1%), the combined Decision Tree + Random Forest (DT+RF) model achieved 80% (±2.3%). The difference was not statistically significant ($p = 0.12$), indicating comparable performance between these models. Confidence intervals (95%) for accuracies are provided in Table 5, supporting the robustness of these findings. Including this analysis clarifies the strength of the performance claims and builds confidence in the comparative evaluation of the ensemble learning techniques.

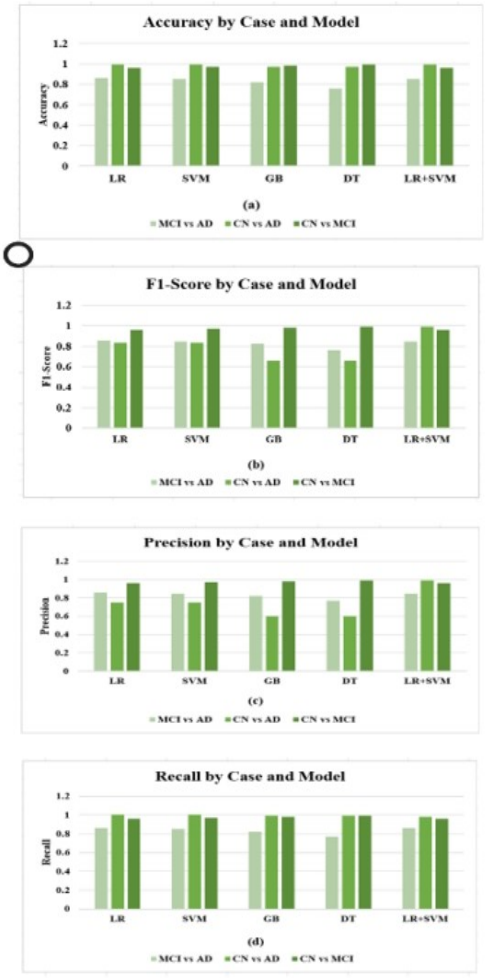


Fig. 9 Represents (a-d) the Multiclass Categorisation of AD and its Subgroups.

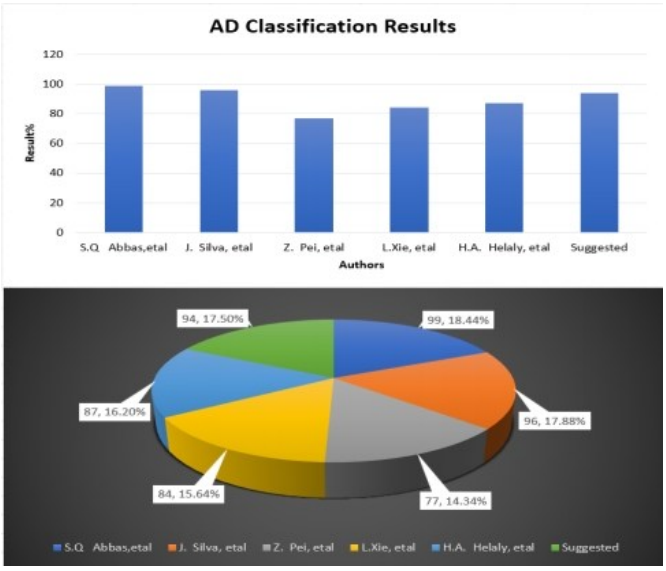


Fig. 10 Comparison Analysis of the Suggested Approach and Additional Cutting-Edge Techniques.

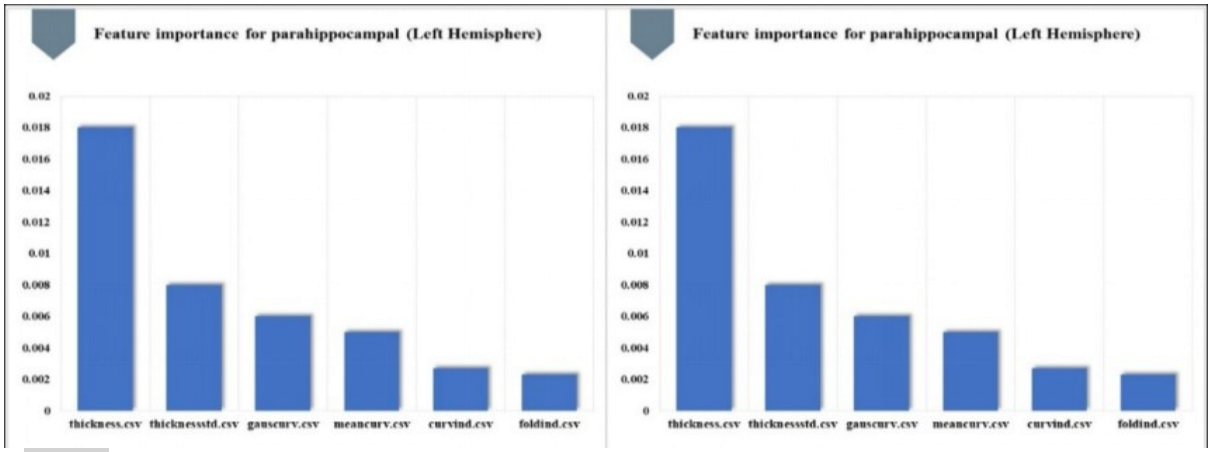


Fig. 11 Examining All Features with the Left Hemisphere in Relation to their Relative Significance.

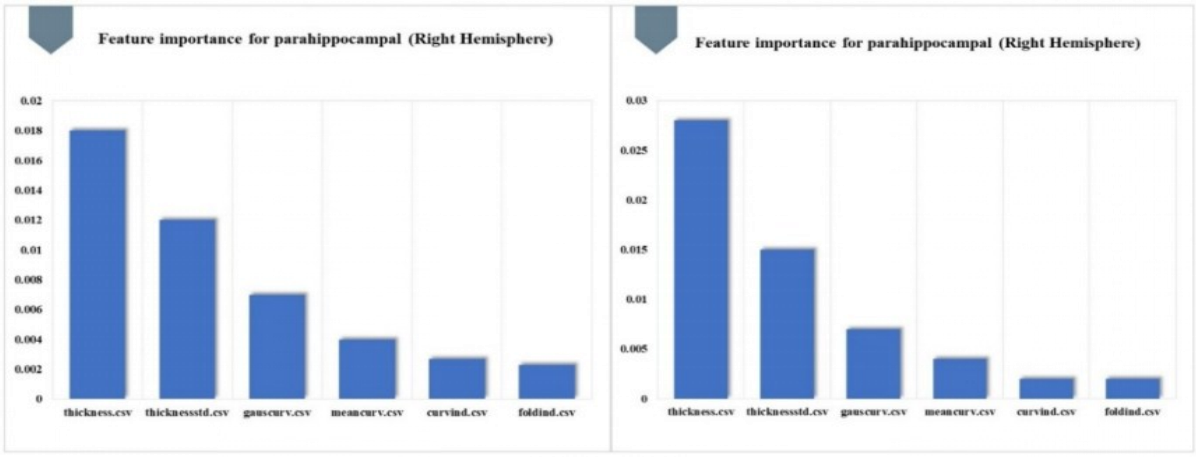
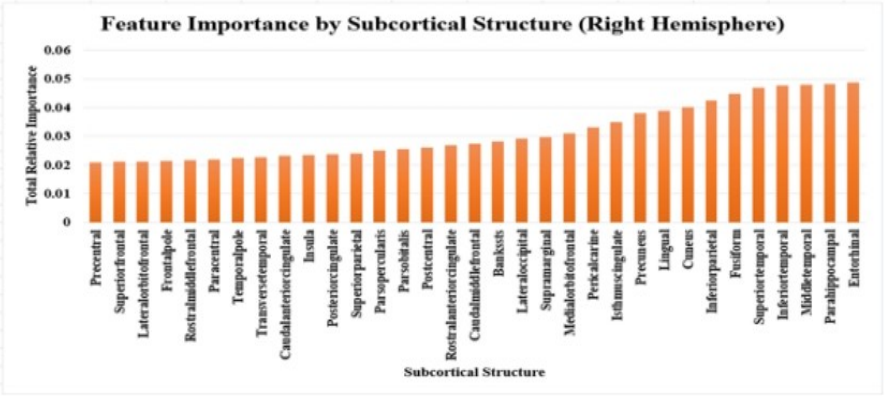
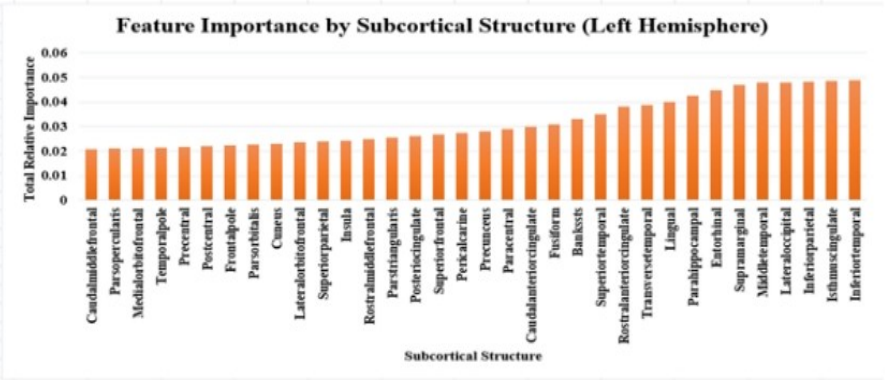


Fig. 12 Examining All Features with the Right Hemisphere in Relation to Their Relative Significance.



(a)



(b)

Fig. 13 Analysing (a and b) the Subcortical Structure of the Left and Right Hemispheres in Terms of Relative Importance.

3.2.Discussion

This study demonstrates that structural biomarker-based AD detection reaching 99% accuracy on the independent test set. This performance is competitive with, and in some cases exceeds, the accuracy reported in previous studies summarised in Table 7, which usually range from 80% to 95% using similar structural MRI features and machine learning approaches.

Table 7 Validation of the Proposed Technique with the Existing Literature.

Refs.	Techniques	Features	Single Class	Multi Class	Result
[17]	Feature Ranking Method	Structural MR images from ADNI (130 for both AD and HC)	Yes	No	92.4%
[19]	SCNN	sMRI from the OASIS dataset	Yes	No	98.7%
[20]	ML	MRI measurements of the entorhinal cortex, superior temporal sulcus (banks), and anterior cingulate.	Yes	No	93%
[21]	CNN	T1-weighted volumetric MRI was minimised to 2D using preprocessing methods from three different projections.	Yes	No	80%
[22]	CNN	LeNet-5 was used to classify sMRI data between AD and CN.	Yes	No	98.8%
This study	Proposed Technique (EL and Traditional Method)	Structural MRI from ADNI and Traditional ML techniques	Yes	Yes	AD vs MCI vs CN =82 AD vs CN = 99 MCI vs CN =99

By integrating multiple classifiers, it capitalises on the complementary strengths of individual algorithms, resulting in enhanced robustness and generalizability compared with single models such as RF or DT. Whereas previous work has often relied on single-modality features or isolated classifiers, our multi-feature statistical approach and careful feature selection refine the predictive risk of overfitting. analytical techniques to detect these early warning signs better Researchers proposed a novel ML framework, Ensemble_LR_SVM, which integrates two complementary algorithms, LR and SVM, to enhance the diagnosis of AD using structural MRI data. When evaluated on the widely recognised ADNI dataset, the model demonstrated exceptional performance, achieving 99% accuracy in differentiating Alzheimer's patients from healthy controls and 96% accuracy in detecting individuals with MCI, an early stage of the disease. The ensemble model also achieved 93% accuracy in multiclass classification (AD, MCI, and control groups), outperforming conventional classifiers such as Random Forest and Decision Trees, which achieved accuracies of 89% and 91%. This improvement highlights the benefits of integrating LR and SVM: LR provides interpretability and robust decision boundaries, whereas SVM captures complex, nonlinear relationships in imaging data. A key strength of this research lies in the model's ability to identify clinically relevant brain regions, notably the entorhinal cortex and parahippocampal areas, that are well-established in neuroimaging studies as early markers of Alzheimer's pathology. Although distinguishing between MCI and AD remains challenging (with an accuracy of 85%), the Ensemble_LR_SVM approach demonstrates strong potential for early and reliable detection. By leveraging the complementary strengths of its component algorithms, this ensemble method effectively captures both prominent and subtle structural changes in the brain, offering a technically robust and interpretable diagnostic tool that could aid clinicians in improving the accuracy and timeliness of Alzheimer's diagnosis. While it achieved excellent results in distinguishing healthy brains from those with AD, it struggled to detect MCI, the early transitional stage in which brain changes are still subtle and difficult to detect. This challenge reflects the real-world difficulty of early AD detection, suggesting that future improvements may come from incorporating additional types of brain imaging data, cognitive test scores, and more sophisticated analytical techniques to better detect these early warning signs. Importantly, our results support the growing evidence that detailed cortical and subcortical structural features serve as highly informative biomarkers for early AD detection. This aligns with neurodegenerative patterns documented in the literature, such as cortical thinning and hippocampal atrophy, which are reliably captured by Free Surfer-derived metrics. The ability to categorise MCI with high specificity also points to the potential utility of these models in identifying prodromal AD stages, thereby promoting timely intervention. Nonetheless, limitations remain, including the inherent constraints of cross-sectional data, the need for validation in larger and more diverse cohorts, and the challenges of translating high-dimensional neuroimaging features into clinically practical tools. Future work should focus on multimodal fusion with other biomarkers (e.g., PET imaging, CSF markers), longitudinal prediction, and interpretability to enhance clinical adoption. The Ensemble LR SVM model offers a significant advantage for

clinical use because it can clearly explain its diagnostic decisions, identifying specific brain regions that are most important for distinguishing between healthy ageing, mild cognitive impairment, and AD. The model pinpoints four key areas: the entorhinal cortex, Para hippocampal gyrus, inferior temporal region, and isthmus cinguli, all of which are crucial for memory, navigation, object recognition, and attention. As these regions deteriorate, they correspond directly to the memory loss and cognitive decline observed by clinicians in patients. This transparency is invaluable for physicians because it provides concrete, measurable brain changes that can support early diagnosis, help track how the disease progresses over time, and distinguish AD from other types of dementia. Unlike complex "black box" AI systems that can't explain their reasoning, this model shows which brain features influence its predictions, making clinicians more confident in using the technology to complement their clinical assessments and potentially detect the disease in its earliest stages, when interventions might be most effective.

4.CONCLUSIONS

In this study, we found that the Ensemble LR_SVM approach outperformed other methods in binary classification, achieving 85.5%, 96%, and 99% accuracy in distinguishing MCI from AD, CN from MCI, and CN from AD, respectively. For multiclass classification, the RF model achieved the highest overall accuracy of 82%, with other conventional ML models performing competitively. The investigation of subcortical brain structures revealed significant regional effects across hemispheres for different AD types. Notably, the right hemisphere's parahippocampal and entorhinal cortices showed substantial influence on AD, while the left hemisphere's inferior temporal and isthmus cingulate areas proved equally significant. Despite these promising findings, challenges persist in applying machine learning models for AD diagnosis to improve accuracy and clinical relevance. A significant obstacle is integrating a broader array of biomarkers or using varied imaging modalities, which, although likely to enhance diagnostic precision, introduce complexity and cost concerns. Addressing these challenges, however, is essential for advancing early detection methods and ultimately improving treatment outcomes for individuals affected by AD.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Tulip Das: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. ***Chinmaya Kumar Nayak***: Methodology, Investigation. ***Parthasarathi Pattnayak***: Methodology, Writing – review & editing, Investigation, Conceptualisation. ***Binod Kumar Pattanayak***: review & editing, Supervision, Resources, Project administration.

DECLARATION OF COMPETING INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication, and there has been no significant financial support for this work that could have influenced its outcome.

ACKNOWLEDGEMENTS

This paper was supported by the KIIT Deemed to be University Library and Dr Gopabandhu Sahu (Librarian), India, as well as the staff of SCB, Medical College, for their invaluable assistance and contributions to this study. Their dedication and support were integral to the successful completion of our research.

REFERENCES

[1] Gao Y, Huang H, Zhang L. **Predicting Alzheimer's Disease Using 3DMgNet.** *ArXiv* 2022.

[2] Panigrahi S, Adhikary DR, Pattanayak BK. **Brain Tumour Classification: A Blend of Ensemble Learning and Fine-Tuned Pre-Trained Models.** *Discover Applied Sciences* 2025; 7(4): 1-24.

[3] Faisal FU, Kwon GR. **Automated Detection of Alzheimer's Disease and Mild Cognitive Impairment Using Whole Brain MRI.** *IEEE Access* 2022; **10**: 65055-66.

[4] Zhang L, Yu M, Wang L, Steffens DC, Wu R, Potter GG, Liu M. **Understanding Clinical Progression of Late-Life Depression to Alzheimer's Disease Over 5 Years with Structural MRI.** *International Workshop on Machine Learning in Medical Imaging* 2022; 259-268.

[5] Khatri U, Kwon GR. **Alzheimer's Disease Diagnosis and Biomarker Analysis Using Resting-State Functional MRI Functional Brain Network with Multi-Measures Features and Hippocampal Subfield and Amygdala Volume of Structural MRI.** *Frontiers in Ageing Neuroscience* 2022; **14**: 818871.

[6] Mansingh P, Pattanayak BK, Pati B. **Big Medical Image Analysis: Alzheimer's Disease Classification Using Convolutional Autoencoder.** *Computación y Sistemas* 2022; **26**(4): 1491-501.

[7] Kong Z, Zhang M, Zhu W, Yi Y, Wang T, Zhang B. **Multi-Modal Data Alzheimer's Disease Detection Based on 3D Convolution.** *Biomedical*

Signal Processing and Control 2022; **75**: 103565.

[8] Zhang X, Han L, Han L, Chen H, Dancey D, Zhang D. **sMRI-PatchNet: A Novel Efficient Explainable Patch-Based Deep Learning Network for Alzheimer's Disease Diagnosis with Structural MRI.** *IEEE Access* 2023; **11**: 108603-16.

[9] Dhinagar NJ, Thomopoulos SI, Laltoo E, Thompson PM. **Efficiently Training Vision Transformers on Structural MRI Scans for Alzheimer's Disease Detection.** *Annual International Conference of the IEEE Engineering in Medicine & Biology Society* 2023; 1-6.

[10] Hu J, Wang Y, Guo D, Qu Z, Sui C, He G, Wang S, Chen X, Wang C, Liu X. **Diagnostic Performance of Magnetic Resonance Imaging–Based Machine Learning in Alzheimer's Disease Detection: A Meta-Analysis.** *Neuroradiology* 2023; **65**(3): 513-27.

[11] Zhang J, He X, Qing L, Chen X, Liu Y, Chen H. **Multi-Relation Graph Convolutional Network for Alzheimer's Disease Diagnosis Using Structural MRI.** *Knowledge-Based Systems* 2023; **270**: 110546.

[12] Abbas SQ, Chi L, Chen YP. **Transformed Domain Convolutional Neural Network for Alzheimer's Disease Diagnosis Using Structural MRI.** *Pattern Recognition* 2023; **133**: 109031.

[13] Mansingh P, Pattanayak BK, Pati B. **Early Detection of Alzheimer's Diseases Through IoT.** *International Journal of Health Sciences* 2022; **6**(S4): 3669-85.

[14] Silva J, Bispo BC, Rodrigues PM. **Structural MRI Texture Analysis for Detecting Alzheimer's Disease.** *Journal of Medical and Biological Engineering* 2023; **43**(3): 227-38.

[15] Pei Z, Wan Z, Zhang Y, Wang M, Leng C, Yang YH. **Multi-Scale Attention-Based Pseudo-3D Convolution Neural Network for Alzheimer's Disease Diagnosis Using Structural MRI.** *Pattern Recognition* 2022; **131**: 108825.

[16] Xie L, Das SR, Wisse LE, Ittyerah R, de Flores R, Shaw LM, Yushkevich PA, Wolk DA. **Baseline Structural MRI and Plasma Biomarkers Predict Longitudinal Structural Atrophy and Cognitive Decline in Early Alzheimer's Disease.** *Alzheimer's Research & Therapy* 2023; **15**(1): 79.

[17] Helaly HA, Badawy M, Haikal AY. **Toward Deep MRI Segmentation for Alzheimer's Disease Detection.** *Neural Computing and Applications* 2022; **34**(2): 1047-63.

[18] Nayak SK, Nayak AK, Laha SR, Tripathy N, Smadi TA. **A Robust Deep Learning-Based Speaker Identification System Using Hybrid Model on KUI Dataset.** *International Journal of Electrical and Electronics Research* 2024; **12**(4): 1502–1507.

[19] Dash L, Pattanayak BK, Laha SR, Pattnaik S, Mohanty B, Habboush AK, Al Smadi T. **Energy Efficient Localization Technique Using Multilateration for Reduction of Spatially and Temporally Correlated Data in RFID System.** *Tikrit Journal of Engineering Sciences* 2024; **31**(1): 101–112.

[20] Pattnayak P, Mohanty A, Das T, Patnaik S. **Applying Artificial Intelligence and Deep Learning to Identify Neglected Tropical Skin Disorders.** *International Conference for Innovation in Technology* 2024; 1-6.

[21] Panigrahi S, Adhikary DR, Pattanayak BK, Dash BB, De UC, Patra SS. **ResNet-GRU Hybrid Model for Brain Tumor Diagnosis: A Sequential Learning Framework.** *International Conference on Machine Learning and Autonomous Systems* 2025; 156-161.

[22] Pattanayak BK, Mansingh P, Pati B, Dash BB, Gourisaria MK, Patra SS. **Alzheimer's Disease Classification Using Capsule Network.** *International Conference on Expert Clouds and Applications* 2024; 644-649.

[23] Mansingh P, Pattanayak BK, Pati B. **Deep Learning-Based Sentiment Analysis for the Prediction of Alzheimer's Drugs.** *Computación y Sistemas* 2023; **27**(4): 979-89.