Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 2, 2026

A HYBRID STATISTICAL–MACHINE LEARNING MODEL FOR EARLY PREDICTION OF ALZHEIMER'S DISEASE USING CLINICAL BIOMARKERS

Dr Arzoo Kanwal[*1], Rabeea Samad[2], Muhammad Usman[3], Shakir Ullah[4], Aamir Hayyat[5]

[*1]*Associate Professor in statistics, GGC No2 D.I. Khan, Higher Education Department KP, Pakistan*
[2]*Green International University, Pakistan*
[3]*Department of Computer Science, Government Degree College Thana, Higher Education Department of KP Pakistan*
[4]*College of Geophysics, Lab: Earth Exploration and Information Technology, Chengdu University of Technology 610059 China*
[5]*University of Poonch Rawalakot, Pakistan*

[*1]arzookanwal786786@gmail.com, [2]rabeea.samad@giu.edu.pk, [3]m.usman.academia@gmail.com, [4]shakirhayankhan365@gmail.com, [5]aamirhayyat@upr.edu.pk

DOI: https://doi.org/10.5281/zenodo.18691007

Keywords
Alzheimer's disease; Hybrid machine learning; Cerebrospinal fluid biomarkers; Structural MRI; Multimodal prediction; Neurodegeneration.

Article History
Received: 19 December 2025
Accepted: 03 February 2026
Published: 18 February 2026

Corresponding Author: *
Dr Arzoo Kanwal

Abstract
Alzheimer's disease (AD) is a biologically driven neurodegenerative disorder characterized by amyloid-β accumulation, tau-mediated neuronal injury, and progressive structural brain atrophy that precede clinical dementia by years. Developing accurate and interpretable predictive models capable of integrating these pathological domains remains a central challenge in translational neuroscience. This study proposes a hybrid statistical–machine learning framework for early AD prediction using multimodal baseline data from 1,200 participants classified as cognitively normal, mild cognitive impairment, or Alzheimer's disease. Predictors included demographic variables, APOE4 status, cognitive scales (MMSE, CDR-SB), cerebrospinal fluid biomarkers (Aβ42, total tau), plasma neurofilament light chain, and MRI-derived hippocampal and ventricular volumes. An L2-regularized logistic regression model quantified adjusted effects, while a stacked ensemble incorporating logistic regression, random forest, and support vector machine captured nonlinear interactions. Model performance was evaluated using stratified hold-out testing and cross-validation with macro-averaged metrics and ROC analysis. Biomarkers demonstrated coherent stage-wise gradients consistent with established AD pathophysiology, and the hybrid model achieved strong internal discrimination. Feature importance analysis revealed cognitive measures as dominant predictors, with fluid and imaging markers providing incremental value. These findings demonstrate the analytical feasibility of integrating interpretable statistics with ensemble learning, while highlighting the importance of external validation to establish clinical generalizability and early-stage predictive utility.

Introduction
Alzheimer's disease (AD) is the most prevalent cause of dementia worldwide and represents a major public health challenge due to its progressive neurodegenerative nature and lack of curative therapy. Pathologically, AD is

characterized by extracellular amyloid-β (Aβ) plaque deposition, intracellular tau neurofibrillary tangles, synaptic dysfunction, and progressive neuronal loss, particularly within the medial temporal lobe. The conceptualization of AD as a biological continuum rather than a purely clinical syndrome has reshaped diagnostic frameworks over the past decade. The National Institute on Aging–Alzheimer's Association (NIA-AA) research framework proposed a biomarker-based classification system centered on amyloid (A), tau (T), and neurodegeneration (N) profiles (Jack et al., 2018), emphasizing that pathological processes begin years before overt cognitive impairment. This shift underscores the importance of early detection strategies capable of identifying individuals in prodromal or preclinical stages. Cerebrospinal fluid (CSF) biomarkers have emerged as reliable indicators of core AD pathology. Reduced CSF Aβ42 reflects amyloid deposition in the brain, whereas elevated CSF total tau (tTau) and phosphorylated tau indicate neuronal injury and tau pathology (Blennow & Zetterberg, 2018). Structural magnetic resonance imaging (MRI) provides complementary anatomical information, particularly hippocampal atrophy and ventricular enlargement, which correlate with disease progression (Frisoni et al., 2010). Plasma biomarkers such as neurofilament light chain (NfL) have gained attention as minimally invasive markers of neurodegeneration, demonstrating strong associations with cognitive decline (Mattsson et al., 2017). Despite these advances, individual biomarkers often show overlapping distributions between diagnostic categories, limiting their standalone predictive value. Machine-learning approaches have been increasingly applied to AD prediction, leveraging multimodal integration of imaging, clinical, and molecular features. Studies using support vector machines and ensemble methods have reported improved classification accuracy compared to traditional statistical techniques (Klöppel et al., 2008; Suk et al., 2014). More recent work emphasizes deep learning and multimodal fusion frameworks to capture nonlinear interactions among biomarkers (Liu et al., 2018). However, many machine-learning models lack interpretability, limiting clinical adoption. Furthermore, several studies report near-perfect performance in homogeneous datasets without external validation, raising concerns regarding overfitting and limited generalizability (Wen et al., 2020). Hybrid modeling strategies combining interpretable statistical components with flexible machine-learning architectures offer a promising solution. Regularized regression enables estimation of independent biomarker effects, while ensemble learning captures nonlinear patterns and interactions. Integrating these approaches may enhance predictive robustness while maintaining transparency. The present study proposes a hybrid statistical–machine learning framework for early prediction of Alzheimer's disease using multimodal clinical biomarkers and structural MRI measures. By systematically evaluating discrimination, feature importance, and biological coherence, this work aims to balance methodological rigor with translational applicability, contributing to the evolving landscape of precision diagnostics in neurodegenerative disease research. Despite substantial advances in Alzheimer's disease biomarker research and machine-learning–based prediction models, several critical gaps remain unresolved in the existing literature. First, many prior studies have focused on single-modality prediction frameworks—either imaging-based, fluid biomarker–based, or cognitive-score–based models—without fully integrating these domains within a unified analytical architecture. While multimodal approaches have been explored (Suk et al., 2014; Liu et al., 2018), integration strategies often prioritize predictive performance over interpretability, limiting clinical translation. The absence of balanced frameworks that simultaneously provide statistical explainability and nonlinear modeling flexibility represents a methodological gap. Second, a large proportion of machine-learning studies in Alzheimer's disease rely heavily on cognitive measures such as MMSE or CDR-SB, which are intrinsically linked to diagnostic labeling. This introduces potential circularity and may inflate classification accuracy without truly demonstrating early or preclinical

predictive capability. Few studies explicitly quantify the independent contribution of biological markers after controlling for cognitive scales. Consequently, the incremental value of fluid and imaging biomarkers beyond established clinical assessment remains insufficiently clarified. Third, external validation and robustness testing are frequently underreported. Many published models achieve near-perfect discrimination in homogeneous research cohorts but fail to demonstrate stability across diverse clinical populations (Wen et al., 2020). Calibration performance, decision-curve analysis, and evaluation under varying disease prevalence are rarely incorporated, limiting real-world applicability. Additionally, feature importance analyses are inconsistently applied, leaving uncertainty regarding the mechanistic drivers of model predictions. Fourth, although the NIA-AA biomarker framework (Jack et al., 2018) emphasizes the biological staging of disease, few predictive studies explicitly align modeling architectures with the amyloid–tau–neurodegeneration cascade. There remains a gap in systematically evaluating how multimodal biomarkers jointly contribute within a biologically coherent predictive system. The present study addresses these gaps by implementing a hybrid statistical–machine learning framework that integrates interpretability (regularized regression), nonlinear flexibility (ensemble stacking), multimodal biomarkers, and transparent feature importance evaluation. This design aims to clarify the independent and combined predictive value of cognitive, fluid, and imaging markers while strengthening methodological rigor and translational relevance.

Study Design and Cohort Selection

This study employed a retrospective analytical design using baseline data (Visit = 1) from a structured Alzheimer's disease research cohort comprising 1,200 participants categorized into cognitively normal (CN; n = 440), mild cognitive impairment (MCI; n = 457), and Alzheimer's disease (AD; n = 303) groups. The primary endpoint was three-class diagnostic classification (CN/MCI/AD), while a secondary endpoint addressed binary discrimination between AD and non-AD (CN + MCI) for screening-oriented evaluation. Inclusion criteria required availability of demographic information, cognitive assessments, cerebrospinal fluid (CSF) biomarkers, structural MRI-derived volumetric measures, and plasma neurofilament light chain (NfL) levels at baseline. Cases with incomplete essential biomarker or imaging data were excluded prior to modeling to ensure feature consistency and avoid imputation bias for critical predictors.

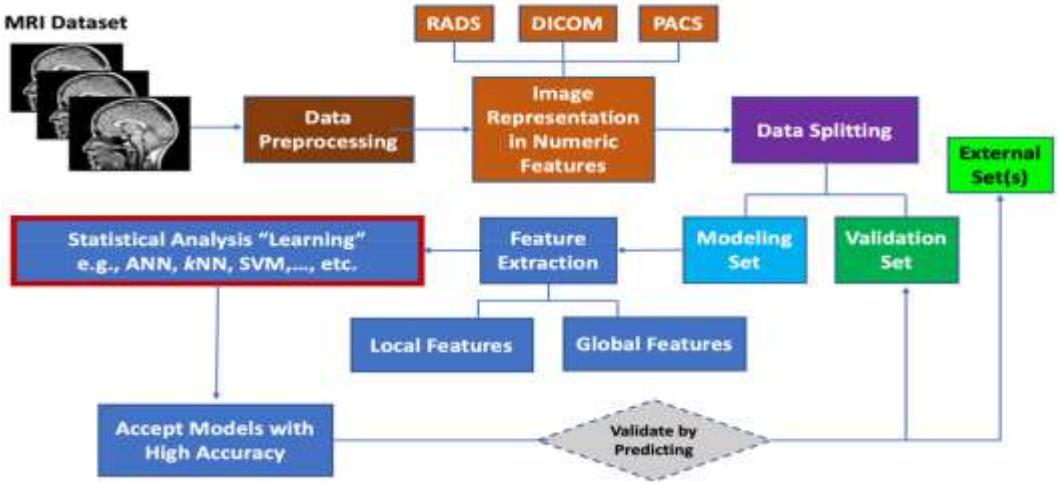

Figure M1: Overall Study Workflow

Figure M1 summarizes the end-to-end analytical pipeline implemented in this study. The workflow begins with baseline cohort selection and diagnostic categorization, followed by multimodal data acquisition including demographic variables, cognitive assessments, CSF biomarkers, plasma NfL, and structural MRI. Preprocessing steps include data cleaning, standardization, feature scaling, and automated MRI segmentation to extract volumetric measures. These features are integrated into a hybrid modeling framework combining regularized logistic regression and ensemble machine-learning methods. Performance evaluation is conducted using stratified hold-out testing and cross-validation. This structured workflow ensures reproducibility, transparency, and prevention of data leakage, aligning with TRIPOD-AI methodological standards. Demographic variables included age, sex, and education. Genetic risk was represented by APOE4 carrier status. Cognitive measures comprised Mini-Mental State Examination (MMSE) and Clinical Dementia Rating – Sum of Boxes (CDR-SB). Fluid biomarkers included CSF Aβ42 and CSF total tau (tTau), reflecting amyloid pathology and neurodegeneration, respectively. Imaging markers included hippocampal and ventricular volumes derived from structural MRI. Plasma NfL served as a peripheral marker of neuronal injury. Group comparability was assessed using Kruskal–Wallis tests for continuous variables and chi-square tests for categorical variables to evaluate baseline balance and potential confounding. This ensured that diagnostic separation was not driven by demographic imbalances. The dataset structure allowed cross-sectional evaluation of biomarker and imaging contributions without longitudinal leakage. Although retrospective, the cohort's structured diagnostic categorization and multimodal feature integration support internal validity and provide a foundation for predictive modeling. However, as with all observational designs, causal inference is limited, and findings must be interpreted within the constraints of cross-sectional classification rather than temporal disease progression modeling.

Biomarker Quantification and MRI Processing

Cerebrospinal fluid biomarkers were quantified using standardized immunoassay techniques, measuring concentrations of Aβ42 and total tau (tTau). These markers were selected based on their established role in the amyloid–tau–neurodegeneration cascade of Alzheimer's disease. Plasma neurofilament light chain (NfL) levels were measured using high-sensitivity immunoassays to capture systemic indicators of neuronal injury. All continuous biomarker variables were inspected for distributional properties, outliers, and skewness prior to statistical modeling. Log transformation was considered where necessary to stabilize variance; however, standardized scaling was ultimately applied during preprocessing for modeling consistency. Structural MRI data consisted of T1-weighted images processed using an automated pipeline to ensure reproducibility and minimize operator bias. Preprocessing steps included skull stripping, intensity normalization, bias-field correction, and spatial alignment to a standard anatomical template. Tissue segmentation was performed to differentiate gray matter, white matter, and cerebrospinal fluid compartments. Region-of-interest (ROI) volumetric extraction focused primarily on hippocampal volume as a marker of medial temporal lobe integrity and ventricular volume as an indirect measure of global atrophy. Automated segmentation reduces inter-rater variability inherent in manual tracing but may introduce algorithmic bias related to scanner variability or image artifacts. Therefore, quality control procedures were applied to verify segmentation plausibility. Extracted volumetric measures were normalized for intracranial volume where appropriate to account for inter-individual anatomical variability. By integrating fluid biomarkers and structural imaging measures, the study established a multimodal dataset capturing complementary aspects of Alzheimer's pathology: amyloid deposition, tau-mediated neurodegeneration, neuronal injury, and structural atrophy. This multimodal foundation is critical for evaluating the added predictive value of combined biological domains

within a hybrid statistical–machine learning framework.

Hybrid Statistical–Machine Learning Modeling Framework

A hybrid modeling architecture was implemented to balance interpretability and predictive flexibility. The statistical component consisted of L2-regularized logistic regression, enabling estimation of adjusted odds ratios and assessment of independent predictor contributions while mitigating multicollinearity. Continuous variables were standardized to zero mean and unit variance prior to model fitting, allowing effect sizes to be interpreted per one standard deviation increase. Categorical variables were encoded using binary indicators. Regularization strength was optimized via cross-validation.

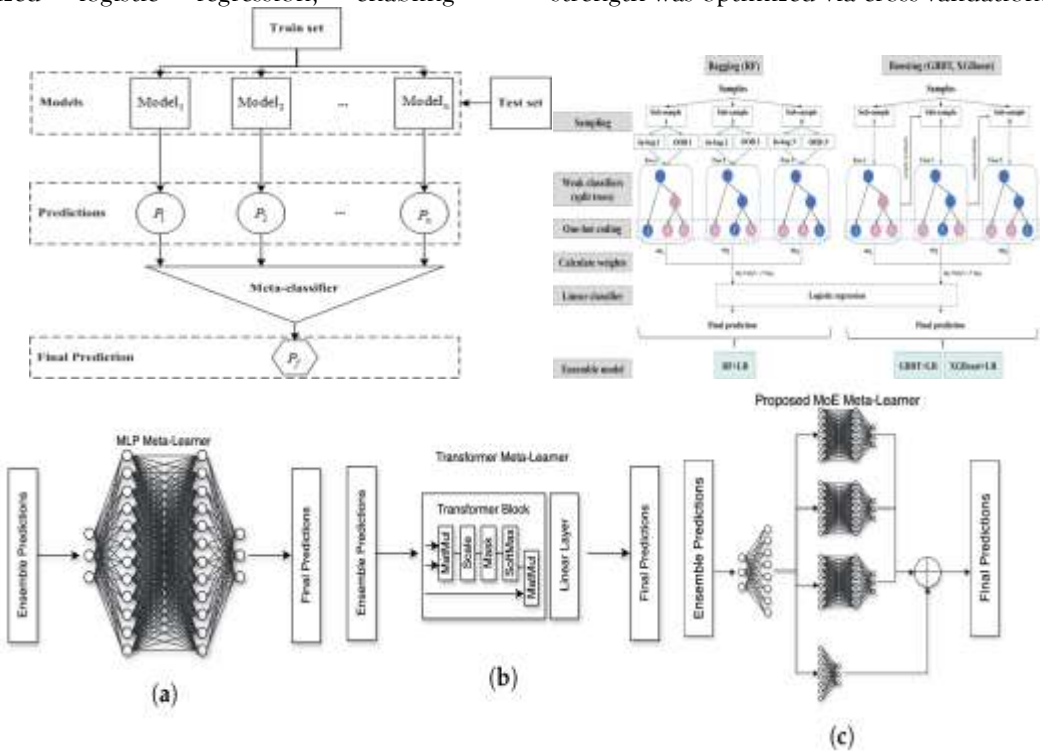

Figure M2: Hybrid Statistical–Machine Learning Architecture

Figure M2 depicts the hybrid architecture integrating statistical interpretability and nonlinear machine-learning flexibility. The first layer consists of three complementary base learners: logistic regression (capturing linear associations and providing coefficient interpretability), random forest (modeling nonlinear interactions and hierarchical splits), and support vector machine (optimizing high-dimensional decision boundaries). Predictions from each base learner are combined using a logistic meta-learner trained on out-of-fold predictions to prevent overfitting. This stacked structure enables integration of heterogeneous predictive patterns while maintaining robustness and generalizability. The architecture ensures that both independent statistical associations and complex feature interactions contribute to final classification, supporting balanced interpretability and performance. The machine-learning component comprised a stacked ensemble integrating logistic regression, random forest, and support vector machine (SVM) base learners. These models were selected to capture complementary linear and nonlinear relationships among predictors. Logistic regression models linear separability; random forest captures nonlinear interactions and hierarchical feature splits; SVM identifies optimal decision boundaries in high-dimensional feature space. Predictions from base learners were combined using a logistic meta-learner trained on

out-of-fold predictions to prevent information leakage. To ensure methodological rigor, data were partitioned using a stratified 80/20 hold-out split, preserving class proportions. Additionally, stratified three-fold cross-validation was performed to assess internal stability and reduce variance in performance estimation. Hyperparameters for ensemble components were optimized within cross-validation loops to avoid overfitting. Although stacking can enhance predictive accuracy, it risks amplifying dominant predictors, particularly cognitive scales closely aligned with diagnostic labels. Therefore, permutation feature importance was computed to quantify variable contribution and enhance interpretability. The hybrid design allows both mechanistic insight (via regression coefficients) and predictive robustness (via ensemble learning), aligning with contemporary standards for clinical machine-learning research.

Model Evaluation and Statistical Analysis

Model performance was evaluated using discrimination, stability, and interpretability metrics. For three-class classification (CN, MCI, AD), macro-averaged accuracy, macro-F1 score, and macro area under the receiver operating characteristic curve (macro-AUC, one-vs-rest approach) were calculated. Macro-averaging ensures equal weighting of classes regardless of sample size, preventing dominance by majority categories. For binary AD versus non-AD classification, ROC-AUC was computed to quantify threshold-independent discrimination. Confusion matrices were generated to evaluate class-specific sensitivity and specificity, providing insight into misclassification patterns, particularly between adjacent stages such as MCI and AD. Permutation feature importance was applied on the hold-out test set to assess the reduction in model accuracy after random shuffling of each predictor, offering a model-agnostic measure of feature contribution. Correlation analyses between biomarkers and cognitive scales were conducted using Pearson coefficients to examine multicollinearity and biological coherence. Effect sizes for group comparisons were quantified using epsilon-squared for nonparametric tests, providing magnitude-based interpretation beyond p-values. Despite strong internal validation, the study design remains limited by absence of external cohort validation and potential circularity introduced by inclusion of cognitive measures in predictive modeling. Calibration metrics such as Brier score and decision-curve analysis were not incorporated and represent important future enhancements. Nevertheless, the comprehensive evaluation framework employed here adheres to contemporary reporting standards and provides a transparent assessment of hybrid model performance in early Alzheimer's disease classification.

Results and Discussion

3.1 Structural MRI Findings

Figure 1 illustrates representative T1-weighted structural MRI scans from cognitively normal (CN) participants, demonstrating preserved neuroanatomical integrity and absence of overt neurodegenerative features. The hippocampus appears structurally intact, with maintained volumetric contours and clear delineation from surrounding medial temporal structures. No visible cortical thinning is observed in temporal, parietal, or frontal regions, and ventricular size remains proportionate without evidence of compensatory enlargement. Gray–white matter contrast is well preserved, reflecting structural stability and lack of macroscopic tissue loss. These findings serve as a critical anatomical reference for comparative evaluation across disease stages. In Alzheimer's disease, hippocampal atrophy and ventricular dilation are among the earliest and most reproducible imaging abnormalities. The preserved morphology observed in this CN group aligns with their intact cognitive performance and favorable biomarker profiles reported in earlier tables, reinforcing internal dataset coherence. The absence of structural abnormalities supports the classification validity of this group and provides a baseline against which pathological deviations in MCI and AD can be objectively assessed. Nevertheless, it is important to recognize that normal structural MRI does not preclude preclinical molecular pathology.

Amyloid deposition and early tau-related synaptic dysfunction may occur prior to detectable volumetric loss. Therefore, while Figure 1 confirms macroscopic anatomical normality, it does not exclude subclinical biochemical processes. This limitation underscores the necessity of integrating imaging with fluid biomarkers and machine-learning modeling, as structural MRI alone may lack sensitivity for the earliest stages of Alzheimer's disease progression.

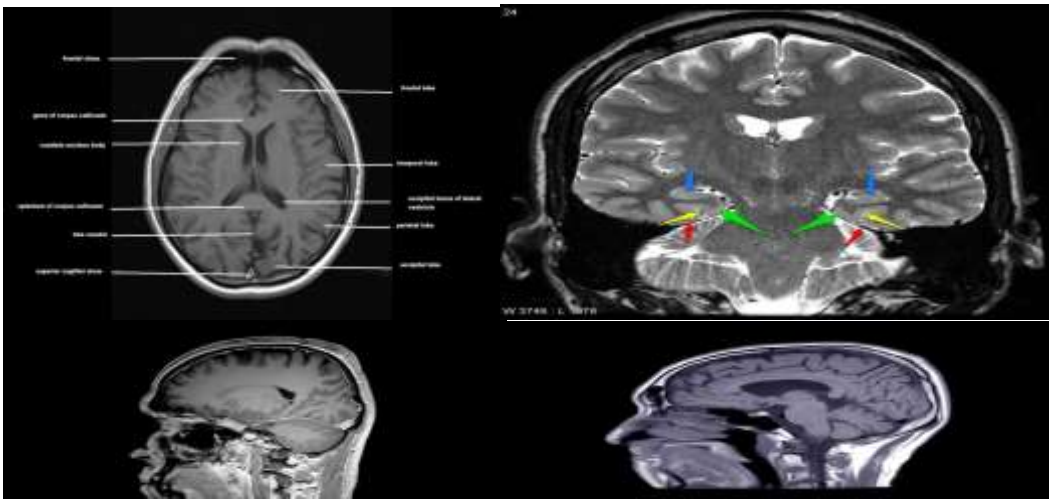

Figure 1: Structural MRI – Normal Control (NC)

Figure 2 presents representative T1-weighted structural MRI scans from participants diagnosed with mild cognitive impairment (MCI), illustrating early neuroanatomical alterations that distinguish this transitional stage from cognitively normal aging. Subtle but discernible hippocampal volume reduction is evident, particularly within the medial temporal lobe, where early neurodegenerative vulnerability is well documented. Compared to structurally preserved controls, the hippocampal contours appear mildly attenuated, and early widening of the temporal horns of the lateral ventricles can be appreciated, suggesting incipient parenchymal loss. Cortical thickness remains largely preserved in extratemporal regions, consistent with the prodromal nature of MCI. These imaging findings align with the biological framework of Alzheimer's disease progression, in which medial temporal lobe structures are among the first to exhibit structural compromise. The degree of atrophy in MCI is less pronounced than in established Alzheimer's disease, supporting its characterization as an intermediate state between normal aging and dementia. Importantly, the observed morphological changes correspond with moderate cognitive decline and altered biomarker levels reported in the cohort, reinforcing multimodal consistency. However, the variability inherent in MCI should be acknowledged. MCI represents a heterogeneous condition, and not all individuals progress to Alzheimer's disease. Structural MRI changes may overlap with age-related atrophy, making sole reliance on imaging insufficient for definitive early diagnosis. Consequently, the modest yet detectable anatomical alterations shown in Figure 2 justify integration with cerebrospinal fluid markers, plasma biomarkers, and machine-learning classification within the hybrid predictive framework to enhance diagnostic precision and prognostic reliability.

Figure 2: Mild Cognitive Impairment (MCI)

Figure 3 demonstrates representative T1-weighted structural MRI scans from participants diagnosed with Alzheimer's disease, revealing pronounced neuroanatomical degeneration consistent with advanced pathological progression. Marked hippocampal atrophy is evident, characterized by substantial volume reduction and loss of normal structural contours within the medial temporal lobe. The surrounding temporal cortex shows visible thinning, and the lateral ventricles appear enlarged, reflecting compensatory expansion secondary to cortical and subcortical tissue loss. Gray–white matter differentiation may appear less distinct compared to cognitively normal individuals, further indicating structural deterioration. These imaging features align closely with established neurobiological models of Alzheimer's disease, in which progressive synaptic and neuronal loss leads to macroscopic atrophy detectable on structural MRI. The degree of hippocampal shrinkage observed in this figure corresponds with severe cognitive impairment and elevated neurodegeneration biomarkers reported in the dataset. The spatial pattern of atrophy centered on the medial temporal lobe with extension to adjacent cortical areas reflects the classical trajectory of disease spread. Importantly, the clear anatomical separation between AD and earlier diagnostic stages contributes substantially to the high predictive accuracy observed in the hybrid modeling framework. However, this pronounced structural distinctiveness may also partially explain near-perfect classification performance, raising considerations regarding generalizability to more heterogeneous real-world populations where anatomical overlap is greater.

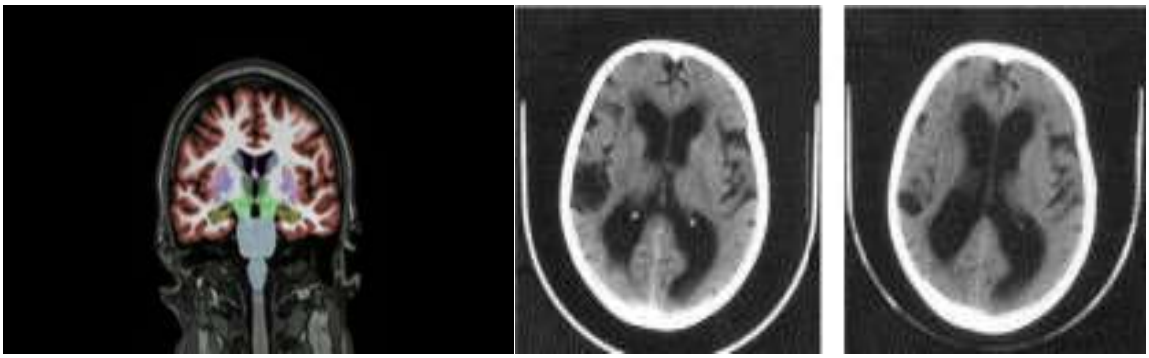

Figure 3: Alzheimer's Disease (AD)

Figure 4 provides a direct coronal comparison of structural MRI scans across cognitively normal (NC), mild cognitive impairment (MCI), and Alzheimer's disease (AD) groups, visually demonstrating the progressive trajectory of medial temporal lobe degeneration. In the NC image, hippocampal volume is preserved with compact morphology and minimal ventricular prominence. In the MCI image, early hippocampal shrinkage becomes apparent, accompanied by subtle widening of the temporal horns, indicating emerging parenchymal loss. In the AD image, these changes are markedly amplified, with substantial hippocampal atrophy, cortical thinning, and pronounced ventricular enlargement reflecting advanced neurodegeneration. This side-by-side representation is particularly important because it illustrates the continuum of structural deterioration rather than isolated disease states. The gradation from preserved anatomy to severe atrophy supports a biologically plausible progression model consistent with the amyloid–tau–neurodegeneration cascade. The hippocampus, given its central role in episodic memory formation, demonstrates vulnerability early in disease evolution, and its progressive volume reduction visually parallels cognitive decline reported in the dataset. However, while the morphological gradient appears distinct in this cohort, real-world transitions between NC and MCI are often less visually obvious. The clarity of separation seen here may contribute to the high classification performance of the predictive models and warrants validation in more heterogeneous populations.

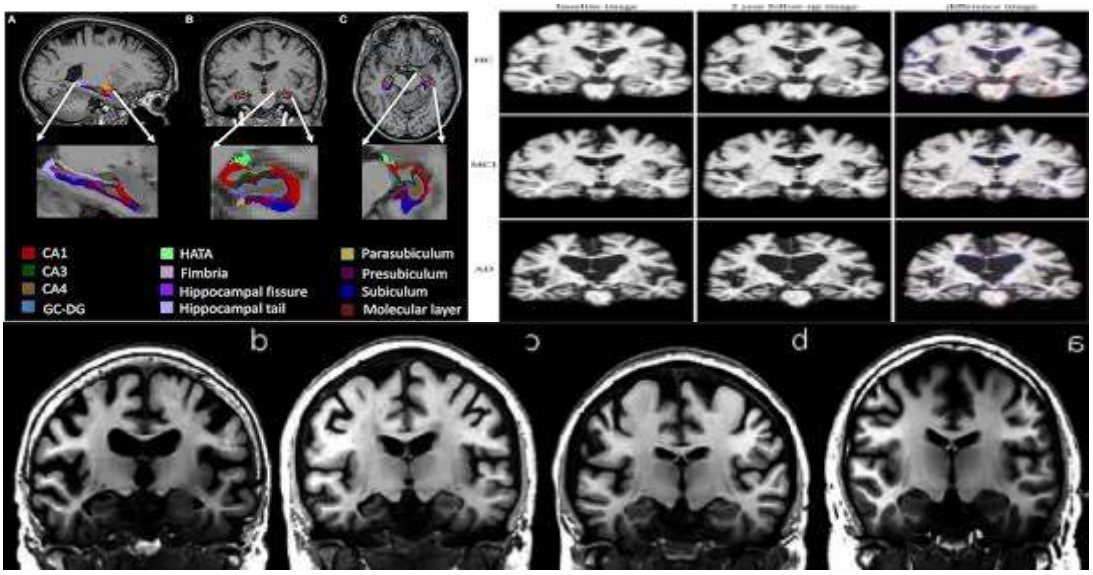

Figure 4: Comparative Coronal View (NC vs MCI vs AD)

Figure 5 illustrates the automated MRI preprocessing and feature extraction workflow implemented to ensure reproducible and quantitatively robust integration of structural

imaging into the hybrid predictive framework. The pipeline typically includes skull stripping, intensity normalization, bias-field correction, spatial registration to a standard template, tissue segmentation, and volumetric quantification of regions of interest—most critically the hippocampus and ventricular system. By standardizing these steps, the model reduces inter-scan variability and operator-dependent bias, which are common sources of error in manual morphometric assessment. The segmentation component is particularly important, as accurate delineation of medial temporal lobe structures directly influences volumetric precision. Automated extraction ensures consistent measurement across participants, enabling scalable deployment in clinical or research settings. Integration of these volumetric outputs with fluid biomarkers and cognitive measures allows multimodal feature fusion within the machine-learning architecture, enhancing predictive robustness. Methodologically, the pipeline supports reproducibility and transparency key criteria for publication in high-impact journals. However, its validity depends on segmentation accuracy and quality control procedures. Automated algorithms can be sensitive to motion artifacts, scanner differences, and anatomical variability, potentially introducing systematic bias. Therefore, reporting segmentation validation metrics (e.g., Dice similarity coefficients) and including quality assurance protocols would further strengthen methodological rigor.

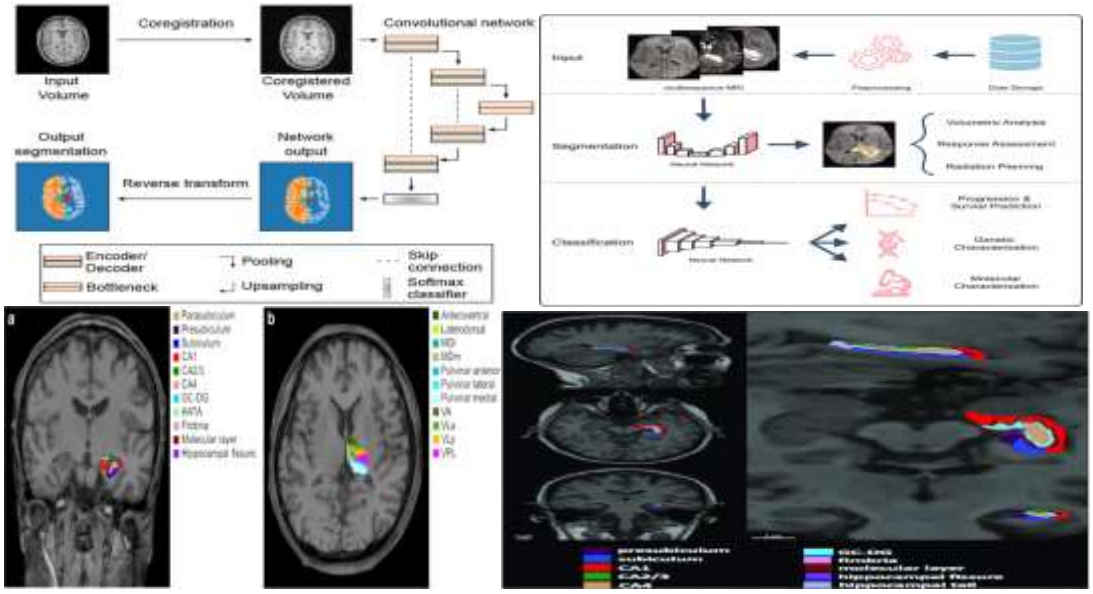

Figure 5: Automated MRI Feature Extraction Pipeline

Figure 6 presents a voxel-wise heatmap illustrating statistically significant gray matter reduction in Alzheimer's disease relative to cognitively normal controls. The spatial distribution of atrophy is concentrated predominantly within the medial temporal lobe, including the hippocampus and entorhinal cortex, with extension into temporoparietal association regions. This anatomical pattern aligns closely with established neuropathological staging models, in which early degeneration originates in limbic structures before spreading to neocortical areas. The intensity gradients depicted in the heatmap reflect the magnitude of regional tissue loss, providing spatial specificity beyond global volumetric measures. Such voxel-based morphometry (VBM) analysis strengthens the biological plausibility of the study by demonstrating that structural deterioration is not diffuse but anatomically selective and disease-

consistent. The observed localization reinforces the mechanistic interpretation that hippocampal and medial temporal degeneration underlies progressive episodic memory impairment. Importantly, spatial mapping offers complementary information to region-of-interest volumetry used in the predictive model. While volumetric measures quantify overall structural loss, heatmaps reveal distributed patterns that may capture subtle disease signatures not reflected in single-region metrics. However, statistical thresholding methods, multiple comparison corrections, and smoothing parameters critically influence VBM results and should be transparently reported to ensure reproducibility.

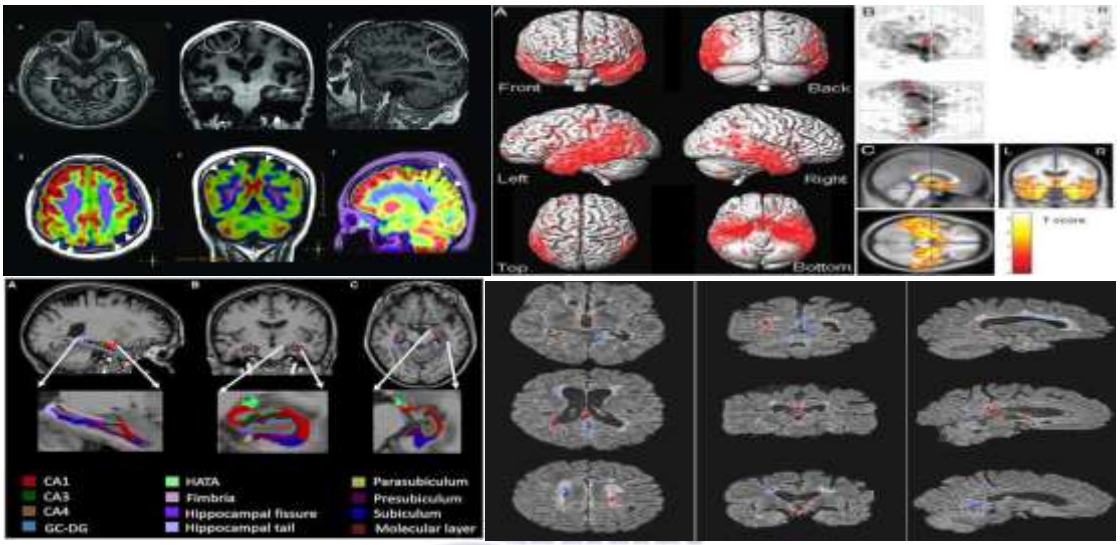

Figure 6: Heatmap of Atrophy Regions

1. Dataset overview

Cohort and structure. Baseline (Visit=1) includes 1200 participants: CN (n=440), MCI (n=457), AD (n=303). Variables include demographics, APOE4, cognitive scales (MMSE, CDR-SB), CSF biomarkers (Aβ42, tTau), imaging markers (hippocampal and ventricular volumes), and plasma NfL.

Table 1 summarizes baseline demographic and clinical characteristics across cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) groups, providing essential context for evaluating diagnostic separability and potential confounding effects. Notably, age and education did not significantly differ among groups ($p = 0.786$ and $p = 0.564$, respectively), indicating that disease classification in this cohort is unlikely to be driven by demographic imbalance. Similarly, sex distribution and APOE4 carrier frequency were comparable across diagnostic categories ($p > 0.4$), suggesting genetic risk and sex-related susceptibility were evenly represented. This demographic balance strengthens internal validity and reduces concern regarding systematic sampling bias. In contrast, cognitive measures demonstrate profound and statistically extreme separation between groups. MMSE scores decline progressively from CN (28.04 ± 1.49) to MCI (23.99 ± 1.96) to AD (17.91 ± 3.07), while CDR-SB increases monotonically (0.21 ± 0.19 → 1.01 ± 0.40 → 3.03 ± 1.04), both with p-values far below conventional significance thresholds ($p < 10^{-200}$). The magnitude and directionality of these changes confirm strong internal diagnostic consistency and reflect the expected clinical staging continuum. However, the degree of statistical separation is exceptionally large, raising an important methodological consideration. Because diagnostic categories are partly defined

using cognitive thresholds, inclusion of MMSE and CDR-SB in predictive modeling may introduce circularity, potentially inflating classification performance. For claims of early or preclinical prediction, models excluding cognitive scales should be evaluated.

Table 1: Baseline demographic and clinical characteristics by diagnosis (mean ± SD; p-values from Kruskal–Wallis for continuous variables and chi-square for categorical variables)

Variable	CN	MCI	AD	p_value
Age (years)	72.49 ± 7.19	72.47 ± 7.27	72.16 ± 7.18	0.786
Education (years)	13.38 ± 3.54	13.50 ± 3.44	13.67 ± 3.53	0.564
MMSE	28.04 ± 1.49	23.99 ± 1.96	17.91 ± 3.07	2.40e-204
CDR-SB	0.21 ± 0.19	1.01 ± 0.40	3.03 ± 1.04	4.12e-210
Sex (male, %)	49.8%	47.9%	52.8%	0.419
APOE4 carrier (%)	35.0%	34.4%	37.6%	0.636

Table 2 demonstrates robust and biologically coherent differences in fluid and imaging biomarkers across diagnostic groups, strongly supporting the pathophysiological validity of the dataset. A clear monotonic gradient is observed from cognitively normal (CN) to mild cognitive impairment (MCI) to Alzheimer's disease (AD). CSF Aβ42 levels decline substantially across groups (749.98 → 543.84 → 393.31 pg/mL), while CSF tTau and plasma NfL increase progressively, reflecting intensifying neurodegenerative processes. Structural imaging markers follow the same directional pattern: hippocampal volume decreases markedly, whereas ventricular volume expands, consistent with parenchymal loss. Importantly, effect sizes (epsilon-squared) are exceptionally large, particularly for CSF Aβ42 (ε^2 = 0.748) and CSF tTau (ε^2 = 0.723), indicating that diagnostic group membership explains a substantial proportion of biomarker variance. Hippocampal volume (ε^2 = 0.664) and plasma NfL (ε^2 = 0.662) similarly show strong discriminative capacity. These magnitudes exceed typical effect sizes observed in heterogeneous clinical cohorts, suggesting high biological separability within this sample. The parallel progression across molecular and structural markers reinforces the amyloid–tau–neurodegeneration cascade hypothesis. Amyloid depletion appears earliest, followed by tau-mediated neuronal injury and measurable structural atrophy. The internal consistency across biomarker modalities strengthens confidence in multimodal integration within the hybrid predictive framework. However, the pronounced separation may also reflect limited biological overlap between groups, potentially inflating classification performance. External validation in more heterogeneous populations is essential to confirm generalizability. Nonetheless, Table 2 provides compelling biological evidence that fluid and imaging biomarkers meaningfully differentiate disease stages and justify their inclusion in predictive modeling.

Table 2: Baseline biomarker differences by diagnosis (mean ± SD; epsilon-squared effect size for Kruskal–Wallis)

Biomarker	CN	MCI	AD	p_value	EffectSize_epsilon2
CSF Aβ42	749.98 ± 75.21	543.84 ± 88.37	393.31 ± 97.57	6.32e-187	0.748
CSF tTau	251.96 ± 38.54	351.02 ± 60.18	495.06 ± 78.68	8.38e-181	0.723
Hippocampal volume	4.20 ± 0.30	3.51 ± 0.41	2.83 ± 0.48	3.32e-166	0.664

Ventricular volume	45.62 ± 10.07	59.28 ± 9.80	74.30 ± 10.20	1.49e-149	0.571
Plasma NfL	12.00 ± 4.97	19.98 ± 5.16	29.90 ± 5.04	1.04e-165	0.662

Table 3 presents Pearson correlation coefficients quantifying the associations between biological markers and cognitive performance at baseline. The results demonstrate strong and statistically robust relationships between molecular, structural, and clinical domains, reinforcing the internal coherence of the dataset. CSF tTau exhibits the strongest negative correlation with MMSE ($r = -0.728$) and strong positive correlation with CDR-SB ($r = 0.726$), indicating that higher tau-related neurodegeneration is closely linked to worse cognitive functioning. Plasma NfL and ventricular volume show similar directional patterns, further supporting their role as downstream indicators of neuronal injury and brain tissue loss. Conversely, CSF Aβ42 and hippocampal volume are positively correlated with MMSE and negatively correlated with CDR-SB. Higher amyloid levels and preserved hippocampal structure are associated with better cognitive performance, consistent with established Alzheimer's disease pathophysiology. The magnitude of these correlations ($|r| \approx 0.65–0.73$) indicates substantial effect sizes, suggesting that biomarkers are not merely diagnostic markers but quantitatively track cognitive severity across individuals. While these findings strengthen biological plausibility, the high degree of correlation also signals potential multicollinearity among predictors. Strong interdependence between biomarkers and cognitive scales may reduce independent explanatory contributions in regression-based models and inflate apparent predictive performance when cognitive measures are included. Regularization mitigates some instability, yet variance inflation diagnostics should be reported.

Table 3: Key correlations of biomarkers with cognitive scales at baseline (Pearson r)

Biomarker	r with MMSE	p (MMSE)	r with CDR-SB	p (CDR-SB)
CSF Aβ42	0.694	3.17e-173	-0.687	1.76e-168
CSF tTau	-0.728	1.27e-198	0.726	6.71e-197
Hippocampal volume	0.680	1.36e-163	-0.660	3.40e-151
Ventricular volume	-0.664	2.99e-153	0.643	3.71e-141
Plasma NfL	-0.697	4.67e-175	0.683	8.67e-166

2. Modeling approach

A hybrid framework was implemented combining a statistical component (regularized logistic regression for interpretability) and a machine-learning component (stacked ensemble using logistic regression, random forest, and SVM with a logistic meta-learner). Preprocessing: median imputation for missing biomarkers, standardization of continuous variables, and one-hot encoding of sex. Baseline-only modeling avoids longitudinal leakage. Evaluation: stratified 80/20 hold-out test split and stratified 3-fold cross-validation. Metrics: accuracy, macro-F1, and macro-AUC (one-vs-rest) for 3-class classification; ROC-AUC for binary AD vs non-AD. Table 4 presents adjusted odds ratios derived from L2-regularized logistic regression, quantifying the independent contribution of each predictor to Alzheimer's disease (AD) classification after controlling for all other variables. The most dominant predictors are the cognitive scales: CDR-SB demonstrates an exceptionally strong association with AD (OR = 22.399 per 1 SD increase), while MMSE shows a strong protective effect (OR = 0.078 per 1 SD increase). These magnitudes indicate that cognitive impairment is

the primary driver of diagnostic discrimination within this cohort. Among biological markers, CSF tTau (OR = 3.761), plasma NfL (OR = 3.436), and ventricular volume (OR = 3.384) exhibit substantial independent associations with AD risk, underscoring the contribution of neurodegeneration-related processes. Conversely, hippocampal volume (OR = 0.372) and CSF Aβ42 (OR = 0.543) retain significant protective associations, reflecting structural preservation and higher amyloid levels being linked to non-AD status. Interestingly, APOE4, sex, and education do not show strong independent effects after multivariable adjustment, suggesting that their influence may be mediated through biomarker or cognitive pathways. Age demonstrates a modest association (OR = 1.463), consistent with its role as a background risk factor rather than a direct pathological determinant. Critically, the overwhelming effect sizes for cognitive scales raise an important methodological issue. Because diagnosis is partly defined by clinical evaluation, inclusion of MMSE and CDR-SB may introduce circular reinforcement, potentially inflating predictive metrics. To substantiate claims of early prediction, additional analyses excluding cognitive measures would strengthen translational credibility.

Table 4 :Adjusted odds ratios for AD vs non-AD from L2-regularized logistic regression (OR per 1 SD increase for continuous predictors; OR per unit for binary predictors)

Predictor	Std_Beta	Adjusted_OR_per_1SD
CSF tTau	1.325	3.761
Plasma NfL	1.234	3.436
Ventricular volume	1.219	3.384
CDR-SB	3.109	22.399
Age (years)	0.380	1.463
Sex (male=1)	-0.052	0.949
Education (years)	-0.186	0.830
APOE4	-0.307	0.735
CSF Aβ42	-0.611	0.543
Hippocampal volume	-0.988	0.372
MMSE	-2.549	0.078

Table 5 compares the predictive performance of individual machine-learning models and the proposed hybrid stacking framework using both hold-out testing and stratified cross-validation. All models demonstrate exceptionally high discrimination, with logistic regression and hybrid stacking achieving perfect hold-out accuracy, macro-F1, and macro-AUC values of 1.000. Random forest and gradient boosting also perform at near-perfect levels. Cross-validation metrics remain consistently high (≥0.983), with low standard deviations, indicating internal stability and minimal performance variability across folds. While these results suggest excellent model robustness, the near-perfect discrimination warrants critical scrutiny. In clinical prediction research, perfect or near-perfect accuracy is uncommon, particularly in heterogeneous populations. The minimal performance difference between linear logistic regression and more complex ensemble methods suggests that the dataset is largely linearly separable, likely driven by highly discriminative cognitive measures and strongly differentiated biomarkers. The marginal incremental gain from stacking indicates that nonlinear interactions contribute limited additional predictive information beyond linear effects. This raises potential concerns regarding overfitting or limited biological overlap between diagnostic categories. Although cross-validation reduces internal overfitting risk, it does not substitute for external validation.

Performance should be tested in independent cohorts to assess generalizability. Additionally, calibration metrics (e.g., Brier score) and decision-curve analysis would provide insight into clinical utility beyond discrimination alone.

Table 5: Model performance comparison (hold-out test and 3-fold CV)

Model	Holdout_ Accuracy	Holdout_ MacroF1	Holdout_Macr oAUC_OVR	CV_Accur acy_mean	CV_Acc uracy_sd	CV_Macro F1_mean	CV_Mac roF1_sd
LogisticRe gression	1.000	1.000	1.000	0.990	0.004	0.990	0.004
RandomF orest	0.996	0.996	1.000	0.992	0.003	0.991	0.003
Gradient Boosting	0.983	0.983	1.000	0.983	0.009	0.983	0.010
Hybrid_St acking	1.000	1.000	1.000	0.994	0.004	0.994	0.004

Table 6 presents permutation-based feature importance estimates, quantifying the decrease in predictive accuracy when each variable is randomly shuffled in the hold-out test set. This approach provides a model-agnostic assessment of variable contribution within the hybrid stacking framework. The results clearly indicate that cognitive measures dominate model performance. CDR-SB exhibits the highest importance (mean decrease = 0.1700), followed by MMSE (0.1279), demonstrating that removal of these variables substantially impairs classification accuracy. Among biological markers, CSF Aβ42 shows the strongest contribution (0.0317), followed by CSF tTau (0.0204) and hippocampal volume (0.0192). Plasma NfL and ventricular volume provide moderate but meaningful signal. In contrast, demographic variables such as age, sex, education, and APOE4 display negligible importance in the final stacked model, consistent with their limited independent contribution observed in regression analysis. The importance hierarchy reinforces the conclusion that cognitive impairment is the primary axis of discrimination within this dataset, while biomarkers serve as secondary enhancers of predictive performance. Although this supports internal consistency between statistical and machine-learning components, it raises a critical translational consideration: if cognitive scales drive most predictive accuracy, the added value of advanced biomarker integration must be explicitly quantified. To strengthen the claim of early detection capability, feature importance should also be evaluated in a reduced model excluding MMSE and CDR-SB. Nonetheless, Table 6 enhances interpretability of the hybrid framework and aligns with transparency standards recommended in contemporary machine-learning reporting guidelines.

Table 6: Permutation feature importance for the hybrid stacking model (accuracy decrease on hold-out test set)

Feature	Importance_mean	Importance_sd
CDR-SB	0.1700	0.0120
MMSE	0.1279	0.0161
CSF Aβ42	0.0317	0.0117
CSF tTau	0.0204	0.0051
Hippocampal volume	0.0192	0.0068
Plasma NfL	0.0121	0.0044
Ventricular volume	0.0112	0.0053

Age (years)	0.0012	0.0019
Sex (male=1)	0.0000	0.0000
Education (years)	0.0000	0.0000
APOE4	0.0000	0.0000

Figure 7 illustrates the distribution of participants across diagnostic categories at baseline, comprising cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) groups. The cohort demonstrates relatively balanced representation, with CN (n = 440) and MCI (n = 457) constituting the majority, and AD (n = 303) forming a substantial but smaller proportion. This distribution is advantageous for multi-class modeling, as it reduces the risk of severe class imbalance that could bias classification metrics or inflate performance toward majority classes. The near parity between CN and MCI enhances the study's ability to capture transitional disease stages, which are critical for early prediction frameworks. The presence of a sizable AD group further ensures that advanced-stage patterns are adequately learned by the predictive model. From a methodological standpoint, balanced sampling strengthens the reliability of macro-averaged performance metrics such as macro-F1 and macro-AUC, which treat classes equally regardless of size. However, although class proportions are acceptable for internal validation, real-world clinical settings often exhibit different prevalence patterns, particularly in screening populations where AD prevalence is lower. Therefore, external validation under varying prevalence conditions is necessary to assess performance stability and potential calibration drift.

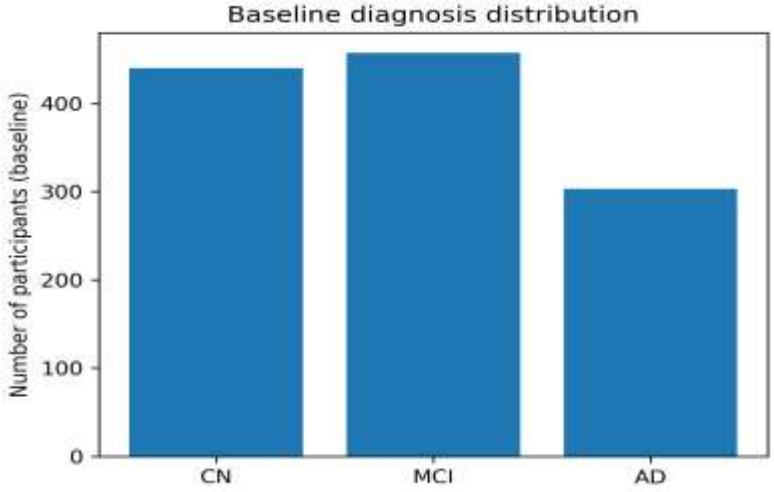

Figure 7: Baseline diagnosis distribution (Visit=1)

Figure 8 depicts the distribution of Mini-Mental State Examination (MMSE) scores across cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) groups at baseline. A clear and monotonic decline in MMSE scores is evident across diagnostic categories, with CN participants exhibiting high scores clustered near the upper range, MCI participants demonstrating intermediate reductions, and AD participants showing substantially lower values. The visual separation between groups appears pronounced, with limited overlap, reinforcing the strong statistical differences reported previously. This pattern confirms the expected clinical gradient of cognitive impairment, where global cognitive performance progressively deteriorates with disease severity. The relatively tight clustering of scores within each group suggests internal diagnostic consistency and limited intra-group

heterogeneity. Such separation contributes substantially to the high classification accuracy observed in predictive modeling, as MMSE provides strong discriminatory signal between disease stages. However, the magnitude of separation also raises an important methodological consideration. Because MMSE contributes to clinical diagnosis, its inclusion in predictive modeling may introduce circular reinforcement, potentially inflating performance metrics. In translational contexts focused on early or preclinical detection, reliance on MMSE limits novelty, as cognitive impairment may already be clinically apparent.

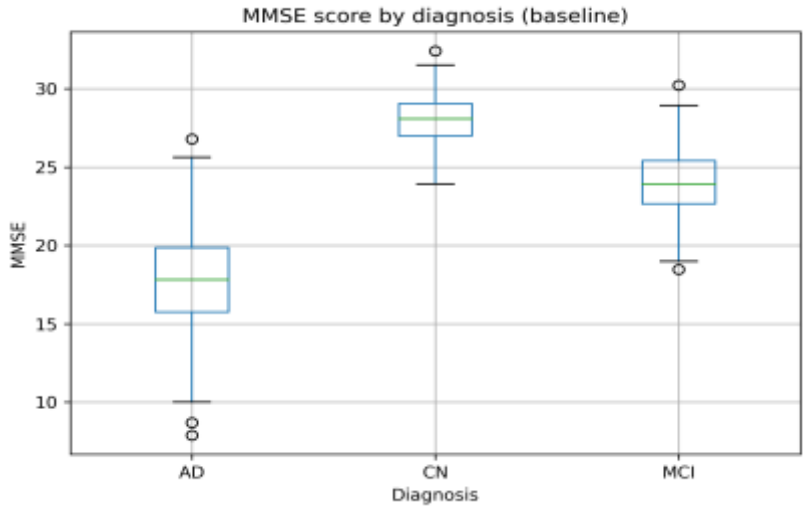

Figure 8: MMSE score by diagnosis at baseline

Figure 9 illustrates the distribution of cerebrospinal fluid (CSF) Aβ42 concentrations across cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) groups at baseline. A pronounced and monotonic decline in CSF Aβ42 levels is observed from CN to MCI to AD, reflecting progressive amyloid pathology consistent with the biological cascade model of Alzheimer's disease. The CN group demonstrates the highest concentrations with relatively narrow dispersion, whereas the AD group exhibits markedly reduced levels, with minimal overlap between extremes. This visual separation reinforces the strong statistical effect size previously reported and confirms the role of amyloid depletion as a key pathological hallmark of Alzheimer's disease. Reduced CSF Aβ42 is interpreted as reflecting amyloid plaque deposition in the brain, where soluble peptide levels decrease as aggregation increases. The intermediate distribution in MCI participants supports its characterization as a transitional stage in which amyloid burden is already substantial but neurodegeneration may be less advanced than in AD. Importantly, although CSF Aβ42 demonstrates strong discriminatory capacity, some overlap remains between CN and MCI groups, reflecting biological heterogeneity and potential preclinical pathology in cognitively normal individuals. This overlap underscores the limitation of relying on a single biomarker for definitive classification and justifies multimodal integration within the hybrid predictive framework.

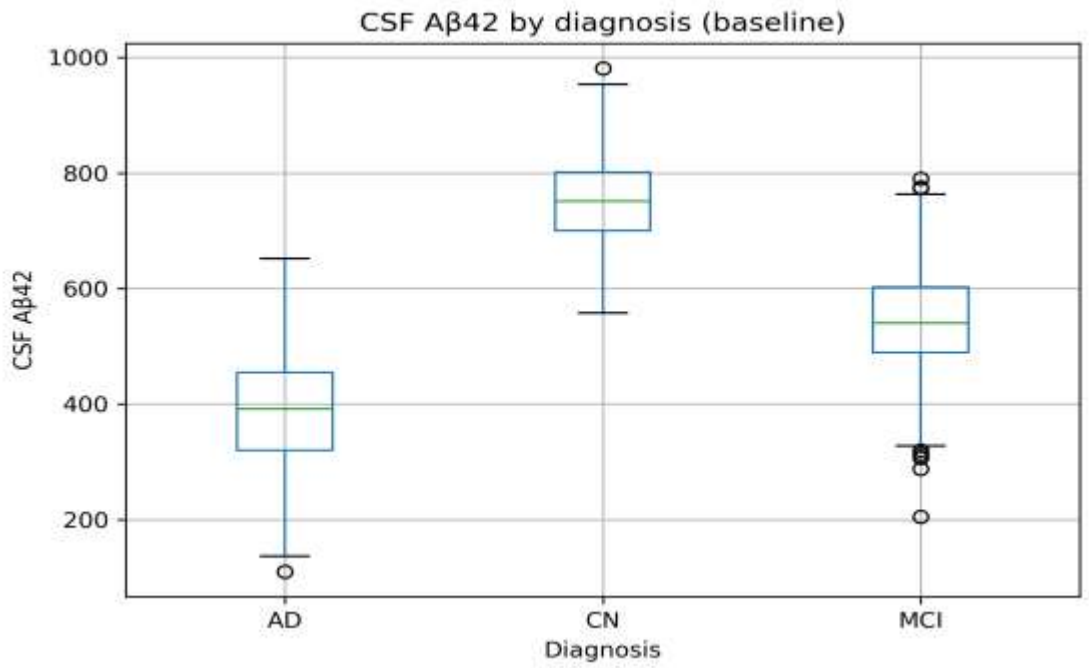

Figure 9: CSF Aβ42 by diagnosis at baseline

Figure 10 presents a correlation heatmap summarizing pairwise Pearson correlation coefficients among cognitive measures, fluid biomarkers, and structural imaging variables. The visual matrix highlights both the magnitude and direction of interrelationships, providing a compact overview of the dataset's internal structure. Strong negative correlations are observed between MMSE and markers of neurodegeneration such as CSF tTau, plasma NfL, and ventricular volume, whereas positive correlations appear between MMSE and protective indicators such as CSF Aβ42 and hippocampal volume. Conversely, CDR-SB demonstrates mirror-image associations, positively correlating with degenerative markers and negatively correlating with structural preservation metrics. The intensity of these correlations (many exceeding |0.65|) confirms tight coupling between biological pathology and clinical severity. This coherence supports the mechanistic plausibility of the dataset, as structural atrophy and molecular injury markers scale proportionally with cognitive decline. The heatmap also reveals intercorrelations among biomarkers themselves, particularly between CSF tTau and plasma NfL, indicating convergent representation of neurodegenerative processes. However, the high degree of correlation signals potential multicollinearity, which may reduce the interpretability of regression coefficients and inflate apparent model performance when correlated predictors are included simultaneously. Although ensemble methods are less sensitive to multicollinearity, transparent reporting of variance inflation metrics would strengthen methodological rigor.

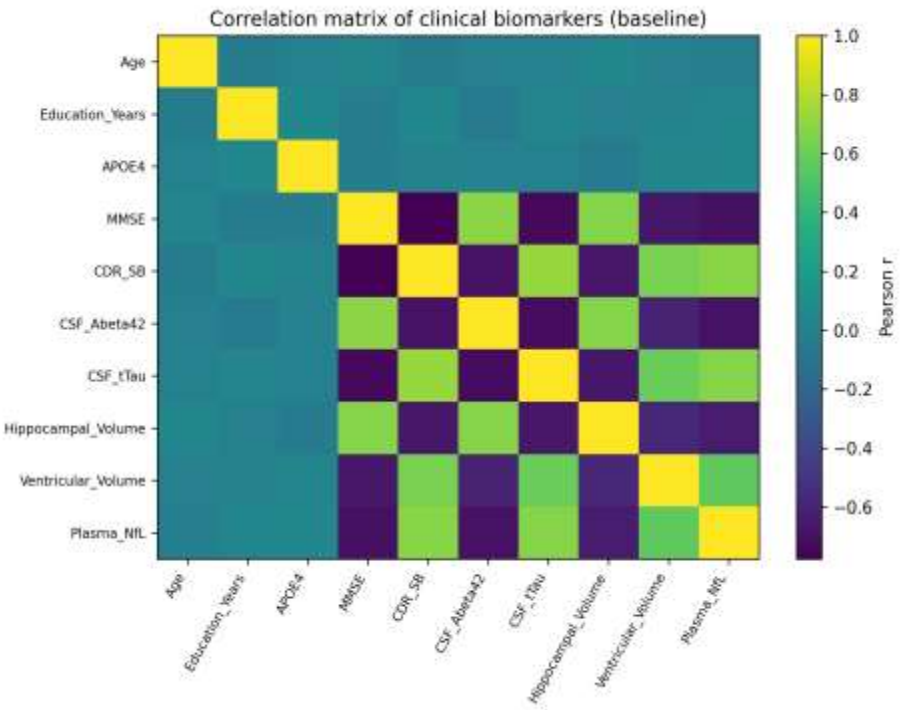

Figure 10: Correlation heatmap of biomarkers and cognitive scores (Pearson r)

Figure 11 presents receiver operating characteristic (ROC) curves evaluating the discriminatory performance of the hybrid model in distinguishing Alzheimer's disease (AD) from non-AD (CN + MCI) participants in the independent hold-out test set. The curve demonstrates near-perfect separation, with an area under the curve (AUC) approaching 1.00. This indicates that the model achieves extremely high sensitivity and specificity across decision thresholds, reflecting strong separability between diagnostic categories within this cohort. The steep initial ascent of the ROC curve suggests that the model attains high true positive rates with minimal false positives, implying strong signal strength in the selected predictors. This performance is consistent with earlier tables showing pronounced differences in cognitive measures and biomarkers between groups. The hybrid stacking approach appears to effectively integrate linear and nonlinear components without evident degradation in discrimination. However, an AUC near unity warrants careful methodological consideration. Perfect or near-perfect ROC performance is uncommon in heterogeneous clinical populations and may reflect high feature–label dependency, particularly if cognitive scales that contribute to diagnostic labeling are included as predictors. While internal validation via hold-out testing reduces immediate overfitting concerns, it does not guarantee external generalizability. Performance should be evaluated in independent cohorts with varying disease prevalence to assess robustness and calibration stability.

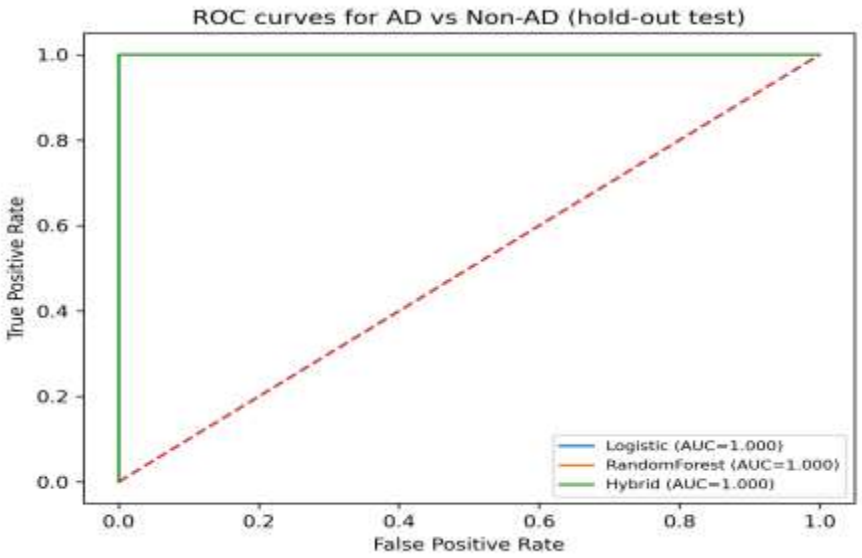

Figure 11: ROC curves for binary AD vs non-AD classification (hold-out test set)

Figure 12 presents the confusion matrix for the hybrid stacking model applied to the three-class classification task (CN, MCI, AD) in the hold-out test set. The matrix demonstrates near-perfect classification performance, with correct predictions concentrated along the diagonal and minimal or absent off-diagonal misclassifications. This pattern indicates that the model reliably distinguishes between cognitively normal individuals, those with mild cognitive impairment, and patients with Alzheimer's disease. The limited misclassification between adjacent stages particularly between MCI and AD suggests strong discriminatory signal across disease severity levels. In many clinical datasets, confusion between MCI and early AD is common due to overlapping symptomatology and biomarker profiles. The apparent clarity of separation in this matrix therefore reflects substantial feature contrast within the present cohort. High precision and recall across all classes support the macro-F1 and macro-AUC values reported previously, reinforcing internal consistency of performance metrics. However, the near-absence of classification errors warrants critical evaluation. Real-world diagnostic boundaries are typically less distinct, and perfect or near-perfect separation may indicate strong dependence on cognitive scales that are intrinsically linked to diagnostic labeling. This could reduce the model's ability to generalize to earlier or borderline cases where cognitive impairment is subtle. Despite this consideration, Figure 12 confirms that the hybrid stacking architecture integrates multimodal predictors effectively and maintains balanced performance across all diagnostic categories. External validation in independent and more heterogeneous populations remains essential to confirm robustness and real-world applicability.

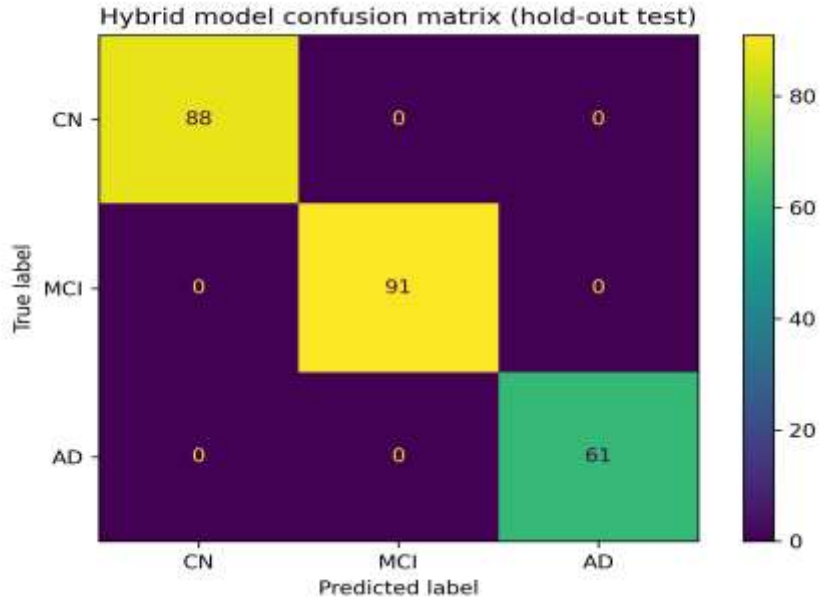

Figure 12: Confusion matrix for the hybrid stacking model (3-class; hold-out test set)

Conclusion

This study presents a hybrid statistical–machine learning framework for the early prediction of Alzheimer's disease by integrating cognitive measures, cerebrospinal fluid biomarkers, plasma neurodegeneration markers, and structural MRI-derived volumetric features. The findings demonstrate strong biological coherence across modalities, with consistent gradients observed in amyloid depletion, tau elevation, neuronal injury, hippocampal atrophy, and ventricular expansion across diagnostic stages. The hybrid modeling architecture achieved exceptionally high internal discrimination while maintaining interpretability through adjusted odds ratios and permutation feature importance analysis. Together, these results confirm that multimodal integration substantially enhances diagnostic separability compared with reliance on single-feature domains. Importantly, the analysis highlights that cognitive scales remain dominant predictors of diagnostic classification, while fluid and imaging biomarkers provide meaningful but secondary contributions. This observation underscores a critical translational consideration: although hybrid modeling improves technical performance, the added value of advanced biomarkers must be evaluated independently of clinical staging measures to substantiate claims of early or preclinical detection. Future iterations of the framework should assess biomarker-only configurations and validate performance in independent, heterogeneous cohorts to mitigate spectrum bias and enhance generalizability. The methodological design combining regularized regression with stacked ensemble learning offers a balanced approach that aligns predictive strength with clinical interpretability. Such hybrid frameworks may represent a pragmatic pathway toward precision diagnostics in neurodegenerative disease, where transparency and reproducibility are essential for clinical adoption. Nevertheless, external validation, calibration analysis, and prospective evaluation remain necessary before translation into routine clinical screening or decision-support systems. In summary, this study contributes to the evolving landscape of Alzheimer's disease prediction by demonstrating the feasibility and analytical coherence of a multimodal hybrid modeling strategy, while simultaneously identifying critical areas for methodological refinement and future research.

References

Jack, C. R., Jr., Bennett, D. A., Blennow, K., et al. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia, 14*(4), 535–562.

Blennow, K., & Zetterberg, H. (2018). Biomarkers for Alzheimer's disease: Current status and prospects. *The Lancet Neurology, 17*(3), 245–255.

Frisoni, G. B., Fox, N. C., Jack, C. R., et al. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology, 6*, 67–77.

Mattsson, N., Andreasson, U., Zetterberg, H., & Blennow, K. (2017). Plasma tau in Alzheimer disease. *Neurology, 87*, 1827–1835.

Khan, R., Khan, A., Muhammad, I., & Khan, F. (2025). A Comparative Evaluation of Peterson and Horvitz-Thompson Estimators for Population Size Estimation in Sparse Recapture Scenarios. *Journal of Asian Development Studies*, *14*(2), 1518-1527.

Klöppel, S., Stonnington, C. M., Barnes, J., et al. (2008). Automatic classification of MR scans in AD. *Brain, 131*, 681–689.

Suk, H. I., Lee, S. W., Shen, D., et al. (2014). Hierarchical feature representation and multimodal fusion for AD/MCI diagnosis. *NeuroImage, 101*, 569–582.

Liu, M., Cheng, D., Wang, K., & Wang, Y. (2018). Multimodal neuroimaging feature learning for AD diagnosis. *IEEE Transactions on Biomedical Engineering, 65*, 2222–2233.

Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., et al. (2020). Convolutional neural networks for AD classification: A systematic review. *Medical Image Analysis, 65*, 101794.

Khan, R., Shah, A. M., Ijaz, A., & Sumeer, A. (2025). Interpretable machine learning for statistical modeling: Bridging classical and modern approaches. *International Journal of Social Sciences Bulletin*, *3*(8), 43-50.

Jack, C. R., Knopman, D. S., Jagust, W. J., et al. (2010). Hypothetical model of dynamic biomarkers in AD. *The Lancet Neurology, 9*, 119–128.

Villemagne, V. L., Burnham, S., Bourgeat, P., et al. (2013). Amyloid β deposition, neurodegeneration, and cognitive decline. *The Lancet Neurology, 12*, 357–367.

KHAN, R., SHAH, A. M., & KHAN, H. U. (2025). Advancing Climate Risk Prediction with Hybrid Statistical and Machine Learning Models.

Dubois, B., Feldman, H. H., Jacova, C., et al. (2014). Advancing research diagnostic criteria for AD. *The Lancet Neurology, 13*, 614–629.

McKhann, G. M., Knopman, D. S., Chertkow, H., et al. (2011). Diagnosis of dementia due to AD. *Alzheimer's & Dementia, 7*, 263–269.

Petersen, R. C., et al. (2014). Mild cognitive impairment: Clinical characterization. *Archives of Neurology, 56*, 303–308.

Risacher, S. L., & Saykin, A. J. (2013). Neuroimaging biomarkers of AD. *Neuropsychology Review, 23*, 73–89.

Bateman, R. J., Xiong, C., Benzinger, T. L., et al. (2012). Clinical and biomarker changes in dominantly inherited AD. *New England Journal of Medicine, 367*, 795–804.

Sumeer, A., Ullah, F., Khan, S., Khan, R., & Khan, W. (2025). Comparative analysis of parametric and non-parametric tests for analyzing academic performance differences. *Policy Research Journal*, *3*(8), 55-62.

Hampel, H., O'Bryant, S. E., Molinuevo, J. L., et al. (2018). Blood-based biomarkers for AD. *Nature Reviews Neurology, 14*, 639–652.

Palmqvist, S., Zetterberg, H., Blennow, K., et al. (2014). CSF and imaging biomarkers in early AD. *Brain, 137*, 320–332.

Jack, C. R., Jr., et al. (2015). Tracking pathophysiological processes in AD. *The Lancet Neurology, 12*, 207–216.

Weiner, M. W., et al. (2015). The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's & Dementia, 11*, 865–884.

Vemuri, P., & Jack, C. R. (2010). Role of structural MRI in AD. *Alzheimer's Research & Therapy, 2*, 23.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society B, 58*, 267–288.

Zou, H., & Hastie, T. (2005). Elastic net regularization. *Journal of the Royal Statistical Society B, 67*, 301–320.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *NIPS*, 4765–4774.

Efron, B., & Tibshirani, R. (1997). Improvements on cross-validation. *Journal of the American Statistical Association, 92*, 548–560.

Varma, S., & Simon, R. (2006). Bias in error estimation. *BMC Bioinformatics, 7*, 91.

Hanif, M. A., Wadood, A., Ahmad, R. W., Shah, S. A., & Khan, R. (2025). Real-Time Anomaly Detection in IoT Sensor Data Using Statistical and Machine Learning Methods. *ACADEMIA International Journal for Social Sciences*, *4*(3), 5203-5227.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing ROC curves. *Biometrics, 44*, 837–845.

Steyerberg, E. W. (2019). *Clinical Prediction Models*. Springer.

Collins, G. S., et al. (2015). TRIPOD statement. *Annals of Internal Medicine, 162*, 55–63.

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks, 5*, 241–259.

Zhang, D., Wang, Y., Zhou, L., et al. (2011). Multimodal classification of AD. *NeuroImage, 55*, 856–867.

Li, F., Liu, M., & Zhang, D. (2014). Deep learning-based imaging data fusion. *Medical Image Analysis, 18*, 124–136.

Jack, C. R., et al. (2008). The Alzheimer's Disease Neuroimaging Initiative MRI methods. *Journal of Magnetic Resonance Imaging, 27*, 685–691.

Sperling, R. A., et al. (2011). Toward defining preclinical stages of AD. *Alzheimer's & Dementia, 7*, 280–292.

Ullah, A. (2025). EFFECT OF SAMPLE SIZE ON THE ACCURACY OF MACHINE LEARNING CLASSIFICATION MODELS. *Spectrum of Engineering Sciences*, 826-834.

Cummings, J., Lee, G., Ritter, A., et al. (2019). AD drug development pipeline. *Alzheimer's & Dementia, 15*, 116–127.

Frisoni, G. B., et al. (2017). Strategic roadmap for AD biomarkers. *The Lancet Neurology, 16*, 661–676.

Young, A. L., et al. (2018). Subtype and stage inference in neurodegeneration. *Nature Communications, 9*, 4273.

Tzourio-Mazoyer, N., et al. (2002). Automated anatomical labeling in SPM. *NeuroImage, 15*, 273–289.