



RESEARCH ARTICLE

OPEN ACCESS

AN INTERPRETABLE MULTI-OMICS ELASTICNET PIPELINE FOR ALZHEIMER'S DISEASE CLASSIFICATION

Aya Abdelaziz¹, Amira Gherboudj², Tunahan Çakir³, Salah Eddine Aliouane⁴,
Mohamed Abdelhafid Hamidechi⁵

^{1, 2, 4}Department of Applied Biology, Mentouri Brothers University, Constantine, Algeria

³Department of Bioengineering, Gebze Technical University, Kocaeli, Turkey

⁵Department of Cellular and Molecular Biology, Abbès Laghrour University of Khenchela, Khenchela, Algeria

¹<http://orcid.org/0009-0007-8283-2693>, ²<https://orcid.org/0000-0002-3972-9216>, ³<http://orcid.org/0000-0001-8262-4420>,

⁴<http://orcid.org/0000-0002-4910-2935>, ⁵<http://orcid.org/0009-0005-3403-6620>

E-mail: aya.abdelaziz@doc.umc.edu.dz, a.gherboudj@umc.edu.dz, tcakir@gtu.edu.tr, salah-eddine.aliouane@doc.umc.edu.dz,
hamidechi.abdelhafid@univ-khenchela.dz

ARTICLE INFO

Article History

Received on: April 02, 2026

Reviewed on: May 04, 2026

Accepted on: June 20, 2026

Published on: September 30, 2026

Keywords:

Alzheimer's Disease,
Multi-Omics Integration,
Machine Learning,
Blood Transcriptomics,
SNP Biomarkers.

ABSTRACT

Alzheimer's disease (AD) classification from heterogeneous biomedical data remains challenging because omics feature spaces are high-dimensional, sample sizes are limited, and data integration must remain interpretable. This study developed an interpretable machine learning framework that integrated biological interpretation tools to distinguish cognitively normal (CN) from AD participants using Alzheimer's Disease Neuroimaging Initiative (ADNI) data. Peripheral-blood gene expression, genome-wide SNPs, and clinical/MRI variables were analyzed in two cohorts: a complete-data common cohort (n = 290; 150 CN, 140 AD) and a broader union cohort (n = 578; 284 CN, 294 AD). After upstream quality control and preprocessing, transcriptomic and genetic matrices were reduced in a separate cross-validated stability-selection stage which yielded fixed consensus top-5000 features per modality for downstream benchmarking. Five predictive configurations were then evaluated under stratified 5-fold cross-validation: expression-only, SNP-only, clinical-only, omics-only, and multi-omics. In the common cohort, SNP-only achieved the highest discrimination (AUC = 0.936; AP = 0.923), whereas multi-omics provided the best overall balance of performance (AUC = 0.932; accuracy = 0.890; Brier score = 0.099). In the union cohort, multi-omics performed best overall (AUC = 0.960; AP = 0.966; accuracy = 0.913). Biological interpretation identified recurrent AD-related candidate signals, including IQCK, APOE/APOC1, and PTGS2.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the predominant cause of dementia, accounting for an estimated 60% to 70% of cases [1] and impacting more than 52 million individuals globally with approximately ten million new cases occurring each year [2]. There is no cure or preventive measure for AD. Considering the substantial population affected by AD and other dementias globally, research institutions worldwide prioritize the prevention, mitigation, effective management and potential cure of these conditions [3]. AD is caused by the accumulation of the protein beta-amyloid outside neurons and the twisted structure of the protein tau within neurons [4]. The neurons that are initially injured in Alzheimer's disease are those in areas of the brain responsible for memory, language and thinking. As a result, the first indications are usually memory, language and cognitive impairments. Although these symptoms are new to the individual, the brain changes that underlie them are thought to start 20 years or more before the symptoms occur [5]. The diagnosis of Alzheimer's Disease relies on brain imaging and cognitive assessments; however, these methods often do not detect the condition in its initial phases, when therapeutic interventions may be most effective.

Blood-derived molecular data are appealing in this setting due to their relative ease of collection compared to cerebral fluid or sophisticated imaging and they can enhance clinical assessment rather than replace it [6]. From a technical standpoint, the problem lies not only in achieving a high-performing classifier but also in establishing a pipeline that is interpretable, reproducible and robust to overfitting, particularly when the quantity of molecular variables significantly exceeds the number of participants. Multi-omics integration is ideal for this challenge since different modalities capture distinct features of AD biology. Genome-wide SNPs indicate inherited susceptibility, blood transcriptomics capture downstream and state-specific activity, and clinical/MRI variables summarize phenotypic presentation. ADNI is a unique resource for assessing such fusion techniques in a standardized cohort [7]. In recent years, artificial intelligence (AI) and machine learning (ML) have become central tools in healthcare because they can extract clinically useful patterns from complex, high-dimensional, and multimodal data. Their impact has been especially visible in diagnostic imaging, risk stratification, outcome prediction, and precision-medicine workflows where they have shown strong utility in handling nonlinear and heterogeneous relationships.

In biomedical research, these methods are particularly valuable when the objective is not only to classify disease states but also to identify robust predictors from large molecular datasets that are difficult to analyze with standard statistical approaches alone. However, the growing success of AI/ML in healthcare has also highlighted the need for models that remain interpretable, reproducible, and clinically trustworthy rather than merely accurate [8]. This study was therefore developed as an interpretable computational pipeline that compares five modality configurations, assesses their performance across two complementary cohorts and subsequently examines the biological relevance of the selected features through differential expression analysis, overlap with curated AD genes and pathway interpretation using g:Profiler2. Its primary contribution lies not in producing a black-box predictive score but in offering a modular, engineering-focused framework that connects predictive performance with biologically meaningful signals, making it possible to obtain results that are simultaneously interpretable, biologically validated and potentially clinically actionable.

II. RELATED STUDIES

Recent studies on Alzheimer's disease prediction generally fall into three broad categories: single-modality blood biomarker approaches, multimodal fusion methods and explainable machine-learning models centered on biomarkers. Early work in the first category demonstrated that peripheral blood transcriptomic data can capture relevant disease-related signals while also highlighting a key challenge: model performance depends strongly on cohort composition, preprocessing procedures and feature-selection methods. For instance, Lee and Lee [9] assessed blood gene-expression-based prediction in the ADNI and AddNeuroMed cohorts, reporting a highest internal AUC of 0.657 in the ADNI subset, whereas external validation across cohorts frequently remained below 0.80. These results were significant because they confirmed the potential informativeness of blood expression data but they also emphasized the limited robustness of transcriptomics-only classifiers when biological signal is modest relative to technical variation and cohort heterogeneity. Another line of research showed that bringing together multiple data types can enhance diagnostic discrimination by leveraging the complementary strengths of each modality. In a notable example, Venugopalan et al. [10] combined MRI, clinical/EHR and SNP data within a deep-learning framework and found that multimodal models surpassed single-source approaches in the ADNI cohort. Their work helped solidify the broader idea that data integration improves predictive performance; however, the proposed framework was built around CNNs and autoencoders, was strongly oriented toward imaging-heavy pipelines and primarily reported results using accuracy, precision, recall and F1-score rather than metrics more informative for calibration.

Moreover, the top internal cross-validation accuracies for the principal multimodal settings were only about 0.78–0.79 and the authors also observed that performance declined when the available sample size was limited or when overlap across modalities was sparse. More recent studies have increasingly focused on explainable approaches to multimodal AD prediction. For example, AlMansoori et al. [11] examined 623 participants from ADNI using blood gene-expression data, SNPs and clinical variables and reported a headline AUC of 0.94 with an accuracy of 0.95 for the model incorporating all modalities. Yet this strongest performance was achieved only when cognitive scores were included in the clinical feature set and when the prediction task was simplified to a binary distinction between cognitively normal individuals and those with MCI or AD. By contrast, the more biomarker-centered models that excluded cognitive-score enrichment performed much less strongly, yielding AUCs of 0.52 for gene expression alone, 0.55 for SNPs alone, 0.65 for gene expression combined with clinical variables without cognitive scores and 0.63 for SNPs combined with the same clinical variables. These findings confirmed the benefit of multimodal learning, while at the same time highlighting the practical importance of achieving effective prediction from blood and genotype data within a sparse, interpretable framework that does not depend heavily on cognitive-score augmentation. A separate research direction has focused on translational blood-based multi-omics panels assessed in clinically diverse populations. In this context, Souchet et al. [12] designed a targeted blood panel comprising 19 biomarkers together with age and demonstrated strong specificity for distinguishing prodromal AD and AD dementia from non-AD brain disorders.

Their model achieved 93.0% specificity, 65.4% sensitivity and an AUROC of 81.9% during internal validation and retained 92.0% specificity with an AUROC of 71.8% in blind external validation. These findings are especially important because they show that blood multi-omics can deliver clinically meaningful specificity in a more challenging differential-diagnosis scenario than conventional CN-versus-AD classification. Nevertheless, the study reflects a targeted assay-validation approach rather than a high-dimensional framework centered on discovery and feature selection. Recent work by Lüleci et al. [13] further emphasized that multi-omics analyses in AD are also influenced by how disease status is defined. Using transcriptomic and metabolomic data from the ROSMAP cohort, they compared clinical and neuropathological diagnostic groupings and showed that neuropathological diagnosis more accurately reflected the molecular disease state. They also identified consensus biomarker candidates and pathway-level signals shared across diagnostic strategies, highlighting the importance of linking cohort definition to downstream biological interpretation. This perspective is especially relevant to the present study which likewise evaluates two cohort constructions and interprets biomarker candidates through multiple complementary analytical layers rather than through predictive performance alone. Collectively, these studies highlight both the promise of blood biomarkers and the advantages of multimodal integration, while also revealing a clear methodological shortcoming in the literature. What remain needed are pipelines that are interpretable rather than opaque, that distinguish the individual contribution of each modality, that can accommodate missing-modality conditions, that enforce fold-specific preprocessing and principled threshold selection and that validate biological relevance beyond predictive performance alone.

The present study was designed to meet this need by comparing five modality configurations, assessing performance in both a strict complete-case cohort and a broader union cohort and confirming biomarker relevance through cross-fold stability, differential expression, overlap with curated AD genes and pathway enrichment analysis with g:Profiler. This perspective is consistent with our findings: although expression-only, SNP-only and integrated models all captured meaningful signal, the integrated approaches provided the most balanced and best-calibrated performance overall, particularly in the broader union cohort [14].

III. MATERIALS AND METHODS

III.1 DATA SOURCE AND ANALYTIC COHORTS

This study used data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [15], a longitudinal public-private resource that integrates clinical, neuroimaging, genetic and molecular data for Alzheimer's disease research. Only participants classified as cognitively normal (CN) or Alzheimer's disease (AD) were included in the present analysis. Diagnostic labels were harmonized across ADNI tables and encoded as a binary outcome for supervised modeling. Following cross-modality alignment of participant identifiers and subsequent analytic filtering, two study cohorts were established. The common cohort included individuals with complete gene-expression, SNP genotype and clinical data availability ($n = 290$; 150 CN and 140 AD), allowing all unimodal and integrated modeling scenarios to be compared directly within the same participant set. In parallel, the union cohort comprised the broader aligned sample that remained after filtering ($n = 578$; 284 CN and 294 AD) and served to test the same analytical framework under a less restrictive inclusion criterion. Three data modalities were included in the analysis. Peripheral blood gene-expression profiles from the Affymetrix Human Genome U219 platform, with 49,386 probes retained following preprocessing. Genome-wide SNP data were extracted from ADNI genotype files and, after quality-control filtering 378,023 variants were available for analysis. The clinical modality comprised demographic and imaging-derived variables, including age, sex, education, marital status, intracranial volume and structural MRI volumetric measures of the hippocampus, entorhinal cortex, ventricles, whole brain, fusiform gyrus and middle temporal region. The analytical workflow was designed to address two linked objectives: to compare the diagnostic utility of each modality alone and in combination and to derive biologically interpretable candidate markers from the same modeling framework. For this reason, predictive modeling and biological interpretation were treated as complementary but analytically distinct components. A schematic overview of the workflow is shown in Figure 1.

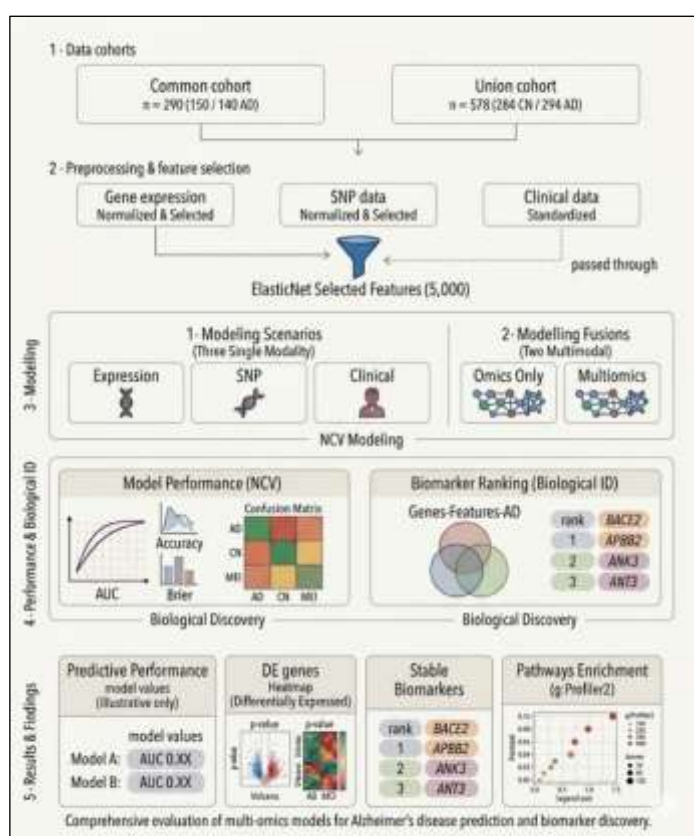


Figure 1: Study workflow and analysis design.

Source: Authors, (2026).

III.2 PREPROCESSING, FEATURE SELECTION AND MODELING

The key methodological barrier in this study was the omics data's high-dimensional complexity in comparison to the available sample size. Both the expression and SNP matrices included considerably more features than participants which increased the risk of overfitting, fold-specific instability and poor interpretability if modeled directly. To overcome this, the workflow was divided into two interlinked stages. To begin, an upstream stability-based feature reduction stage was applied individually to the expression and SNP matrices to provide reproducible reduced feature panels. Second, these reduced matrices were used in a subsequent comparative benchmarking stage to evaluate predefined modeling scenarios using leakage-free cross-validation. This structure was designed to reduce dimensionality while maintaining transparency, reproducibility and coefficient-based interpretability.

All modality tables were initially harmonized by participant identification to ensure that expression, genotype and clinical records matched the same individual and diagnosis. Preprocessing was performed in a strictly fold-restricted way across all analyses during cross-validation. Any data-dependent process, such as imputation, scaling, categorical encoding, model fitting and decision-threshold selection, was learned from the training partition and then applied unchanged to the matching held-out fold. This architecture was routinely employed to prevent information leakage and ensure objective out-of-fold evaluation. The upstream stability-selection stage was also cross-validated but it was used only to derive fixed consensus feature panels for downstream benchmarking rather than to report final performance directly.

Gene expression preprocessing: Expression preprocessing was performed at the probe level rather than after early compressing to genes because the predictive step was designed to maintain all quantifiable transcriptome variance prior to regularized feature selection. The probe identifiers were consistent across ADNI individuals and probes with more than 10% missing values were eliminated. The remaining missing expression values were imputed using the median computed from the training data for each cross-validation split. After imputation, probe intensities were standardized inside the training fold and the resulting scaling parameters were carried over to the held-out fold. This distinction is critical for interpretation: the classifier was trained on probe intensities, while gene symbols were assigned later for biological interpretation of specific traits. After preprocessing, the expression matrix had 49,386 probes.

Genome-wide SNP preprocessing: Genome-wide SNP data were preprocessed using a standard quality-control pipeline before model development. Beginning with PLINK-formatted genotype files, both samples and variants were filtered according to established thresholds for missingness and variant quality, including $\text{mind} = 0.2$, $\text{geno} = 0.1$, $\text{call rate} \geq 95\%$ and $\text{minor allele frequency (MAF)} \geq 0.01$. To limit redundancy among highly correlated loci, linkage-disequilibrium pruning was subsequently applied using the independent $50 \times 50 \times 0.2$ procedure. In the upstream stability-selection stage, any remaining missing genotype values were imputed using the most-frequent value within each training split, after which the imputed dosages were clipped to the valid 0–2 range and rounded to integer dosage values before ElasticNet feature selection. In the downstream benchmarking pipelines, the reduced SNP matrices were instead handled with training-fold median imputation followed by fold-specific standardization. This workflow produced a filtered and pruned SNP matrix comprising 378,023 variants, designed to preserve broad inherited-risk information while minimizing technical noise and excessive convergence.

Clinical data preprocessing: Clinical data preprocessing was deliberately kept straightforward and transparent so that the clinical branch functioned as a practical low-dimensional baseline rather than as a separate radiological modeling pipeline. Demographic and MRI-derived variables were matched by participant identifier and combined into a unified tabular dataset. Categorical features were transformed into numerical inputs through one-hot encoding, whereas continuous variables were imputed using the median when needed and subsequently standardized within the training folds and then applied unchanged to the held-out fold.

Omics Integration and Modeling: In the multimodal analyses, each data modality was preprocessed independently and the resulting feature sets were merged only after those modality-specific steps through an early-fusion strategy. More specifically, the expression and SNP matrices were combined for the omics-only condition, whereas the expression, SNP and clinical feature blocks were integrated for the multi-omics condition. This sequence ensured that preprocessing remained appropriate to each modality while preserving strict separation between training data and held-out samples. Given the extreme dimensionality of the raw omics spaces relative to cohort size, the upstream feature-reduction stage was conducted separately for expression and SNP data. This stage used 5-fold stratified cross-validation with shuffling and $\text{random_state} = 42$. Within each training split, the modality matrix was first imputed, standardized and variance-prefiltered to the top 20,000 features when necessary, then passed to an ElasticNet LogisticRegressionCV model configured with $\text{penalty} = \text{"elasticnet"}$, $\text{solver} = \text{"saga"}$, $\text{cv} = 3$, $\text{Cs} = 3$, $\text{l1_ratios} = [0.5]$, $\text{max_iter} = 5000$, $\text{scoring} = \text{"roc_auc"}$ and $\text{n_jobs} = -1$. From each fold, the top 5,000 features per modality were retained according to absolute coefficient magnitude. The final top-5000 expression and SNP panels were then assembled by cross-fold frequency of the retained feature IDs, yielding fixed consensus feature spaces for downstream benchmarking. Separately, top stable biomarkers were ranked from downstream fold-wise coefficient matrices using selection_freq , mean_coef and mean_abs_coef . The final benchmarking stage applied these fixed reduced matrices across five predefined scenarios: expression-only, SNP-only, clinical-only, omics-only (expression plus SNP) and multi-omics (expression plus SNP plus clinical).

Performance was evaluated with an outer StratifiedKFold configured as $\text{n_splits} = 5$, $\text{shuffle} = \text{True}$ and $\text{random_state} = 42$. Omics-containing models used LogisticRegressionCV with $\text{penalty} = \text{"elasticnet"}$, $\text{solver} = \text{"saga"}$, $\text{class_weight} = \text{"balanced"}$, $\text{inner_cv} = 5$, $\text{max_iter} = 5000$, $\text{Cs} = \text{np.logspace}(-3, 3, 13)$ and $\text{l1_ratios} = (0.1, 0.3, 0.5, 0.7, 0.9)$. Expression-only and SNP-only branches used Pipeline objects composed of SimpleImputer($\text{strategy} = \text{"median"}$), StandardScaler() and the ElasticNet classifier. The clinical branch used a ColumnTransformer with median imputation plus StandardScaler for numerical variables and most_frequent imputation plus OneHotEncoder($\text{handle_unknown} = \text{"ignore"}$) for categorical variables, followed by LogisticRegression with $\text{penalty} = \text{"l2"}$, $\text{solver} = \text{"liblinear"}$, $\text{class_weight} = \text{"balanced"}$, $\text{max_iter} = 5000$ and $\text{random_state} = 42$. Using a common modeling family across all five scenarios enabled direct comparison between modalities while preserving a transparent link between retained features and fitted coefficients. Rather than using a fixed cutoff of 0.50, binary decision thresholds were determined separately within each outer-fold training split. Specifically, the optimal threshold was selected from the training predictions using the Youden index and then applied once to the corresponding held-out fold. For each scenario, out-of-fold probabilities were saved and used to compute global ROC AUC, average precision (AP), accuracy, F1-score, precision, recall, specificity and Brier score. Fold-level metric tables were also retained to quantify variability across outer splits. Calibration was assessed using the Brier score together with calibration curves generated from pooled out-of-fold probabilities.

III.3 BIOLOGICAL VALIDATION AND INTERPRETATION

Biological interpretation was designed to complement predictive modeling while remaining analytically separate from it. In the present workflow, biological validation comprised two linked components: (i) knowledge-guided annotation of the selected feature sets, used to determine whether the retained genes and loci mapped onto established AD-related biology and (ii) model-derived interpretation, used to identify stable candidate biomarkers and transcriptomic alterations associated with the classification framework.

This distinction is important because the biological analyses were intended to support plausibility and interpretability rather than to optimize predictive performance. The first biological-validation component consisted of a curated overlap analysis using genes previously associated with AD. Genes captured by the selected expression probes together with genes mapped from the prioritized SNPs were compared against a curated AD-related gene list compiled from four evidence sources: the KEGG database, the GeneCards database, the AD-related genes reported by Hu et al. [16] and genes supported by recent Alzheimer's disease genome-wide association studies (GWAS) [17], [18], [19]. This overlap analysis played no role in model training, hyperparameter optimization, or feature ranking. Instead, its purpose was deliberately narrower and methodologically more defensible to evaluate whether the selected feature sets recovered known AD-relevant biology more frequently than would be expected from an unrelated set of features. Accordingly, the findings from this step were interpreted as an assessment of biological plausibility rather than as evidence of causality for each overlapping gene. The second biological-validation component consisted of pathway enrichment analysis performed with g:Profiler2 [14] on the same selected gene sets. This step extended the overlap analysis from the level of individual genes to the level of coordinated biological processes. Whereas the curated overlap analysis assessed concordance with previously reported AD-associated genes, the enrichment analysis evaluated whether the selected genes clustered within biologically coherent pathways and functional themes. This distinction is especially important in the context of regularized models where features with modest individual effects may still collectively reflect a shared underlying mechanism. In the present study, enrichment findings were therefore interpreted as providing systems-level biological context for the selected feature panels rather than as a replacement for independent external validation. Model-derived interpretation was then performed through stability-based biomarker ranking. Selected features were ranked using their selection frequency, mean coefficient and mean absolute coefficient magnitude across resampled training subsets.

On this basis, candidate biomarkers were operationally defined as features that contributed consistently to the classifier across resampling iterations and were therefore less likely to represent fold-specific artifacts. This step provided a direct link between predictive modeling and biological interpretation by highlighting the features most reproducibly retained by the learning framework. The final interpretive component consisted of differential-expression analysis performed on the selected expression feature matrix used in the downstream modeling workflow. Comparisons between AD and CN participants were conducted using Welch's two-sample t-test and the resulting p-values were corrected for multiple testing with the Benjamini–Hochberg false-discovery-rate procedure. For each gene, the analysis reported mean expression in the AD and CN groups, the mean difference between groups, the raw p-value and the adjusted q-value. This step offered an orthogonal perspective on transcriptomic alteration which was essential because predictive relevance does not necessarily correspond to univariate group differences. A given feature may enhance classification through multivariable interactions even when its mean expression shift is limited, whereas a gene with strong differential expression may contribute little once correlated predictors are modeled jointly. Reporting both stability-based feature selection and differential expression therefore enabled a more balanced and biologically grounded interpretation. Overall, these validation layers were intended to strengthen biological interpretation while maintaining a strict distinction between predictive modeling and downstream inference. The curated overlap and pathway-enrichment analyses established whether the selected feature sets mapped onto credible AD-related biology, whereas stability ranking and differential expression highlighted the features most consistently associated with the predictive framework. Candidate biomarkers were therefore interpreted with caution, with strongest emphasis placed on signals supported by convergent evidence across multiple dimensions, including curated AD-gene overlap, pathway coherence, cross-fold stability and transcriptomic differential expression.

IV. RESULTS AND DISCUSSIONS

IV.1 PREDICTIVE PERFORMANCE ACROSS MODALITIES

Table 1 and Figure 2 summarize the out-of-fold performance of the five evaluated scenarios. In the common cohort, all models performed above chance but their error profiles differed by modality. SNP-only produced the highest discrimination ($AUC = 0.936$; $AP = 0.923$) and the highest specificity (0.913) indicating that inherited genetic signal was the strongest single-modality discriminator in the strict complete-case cohort. Expression-only also performed strongly ($AUC = 0.910$; $AP = 0.902$) and achieved the highest recall among the single-modality models (0.907) whereas clinical-only was clearly the weakest baseline ($AUC = 0.821$). Multi-omics did not surpass SNP-only on every rank-based metric in this cohort but it delivered the best overall balance with $AUC = 0.932$, accuracy = 0.890, $F1 = 0.887$, recall = 0.893, specificity = 0.887 and the lowest Brier score (0.099). This pattern suggests that integration mainly improved robustness, calibration and error balance in the common cohort. The larger union cohort sharpened the advantage of integration. Multi-omics became the best overall model across the main metrics, reaching $AUC = 0.960$, $AP = 0.966$, accuracy = 0.913, $F1 = 0.911$, precision = 0.959, specificity = 0.961 and the lowest Brier score (0.074). Omics-only was a close runner-up ($AUC = 0.953$; $AP = 0.961$), while SNP-only remained the best single modality by discrimination ($AUC = 0.948$; $AP = 0.953$). Expression-only preserved the highest sensitivity (recall = 0.891) but did so at the cost of substantially more false positives, reflected by its lower specificity (0.687). By contrast, the multi-omics model reduced false positives sharply while retaining strong sensitivity. From an engineering viewpoint, this is the most practically relevant result of the study: the fused model preserved the complementary strengths of the single modalities while reducing their isolated weaknesses. A practical way to interpret these results is through the complementary error profiles of the single-modality models. Expression-only tended to favor sensitivity, especially in the union cohort, where it produced the fewest false negatives but at the cost of many more false positives. SNP-only behaved in the opposite way, with markedly better specificity and precision, consistent with inherited-risk features acting as strong rank-ordering variables. Multi-omics combined these strengths: it retained strong sensitivity while sharply reducing false-positive calls which is why it dominated not only AUC and average precision in the union cohort but also accuracy, F1-score, specificity and Brier score. This indicates that integration was not merely increasing one favorable metric while worsening others; rather, it improved the overall operating point. Fold-level summaries from the outer cross-validation splits indicated that the strongest models were also reasonably stable across resamples. For the multi-omics model, the common cohort yielded mean $\pm$ SD values of $AUC = 0.928 \pm 0.040$, $AP = 0.911 \pm 0.053$ and Brier = 0.099 ± 0.030 across the five outer folds, whereas the union cohort yielded $AUC = 0.961 \pm 0.022$, $AP = 0.966 \pm 0.020$ and Brier = 0.074 ± 0.020 . These fold-wise results support the interpretation that the advantage of integration, especially in the union cohort, was not driven by a single favorable split but was reproduced across the evaluation folds.

Table 1: Out-of-fold classification performance in the common and union cohorts.

Cohort	Scenario	AUC	AP	ACC	F1	Brier
Common	Expression-only	0.910	0.902	0.848	0.852	0.120
	SNP-only	0.936	0.923	0.886	0.879	0.100
	Clinical-only	0.821	0.840	0.766	0.746	0.171
	Omics-only	0.919	0.894	0.872	0.869	0.107
	Multi-omics	0.932	0.913	0.890	0.887	0.099
Union	Expression-only	0.865	0.866	0.791	0.812	0.144
	SNP-only	0.948	0.953	0.870	0.862	0.090
	Clinical-only	0.838	0.857	0.773	0.781	0.163
	Omics-only	0.953	0.961	0.907	0.906	0.082
	Multi-omics	0.960	0.966	0.913	0.911	0.074

Source: Authors, (2026).

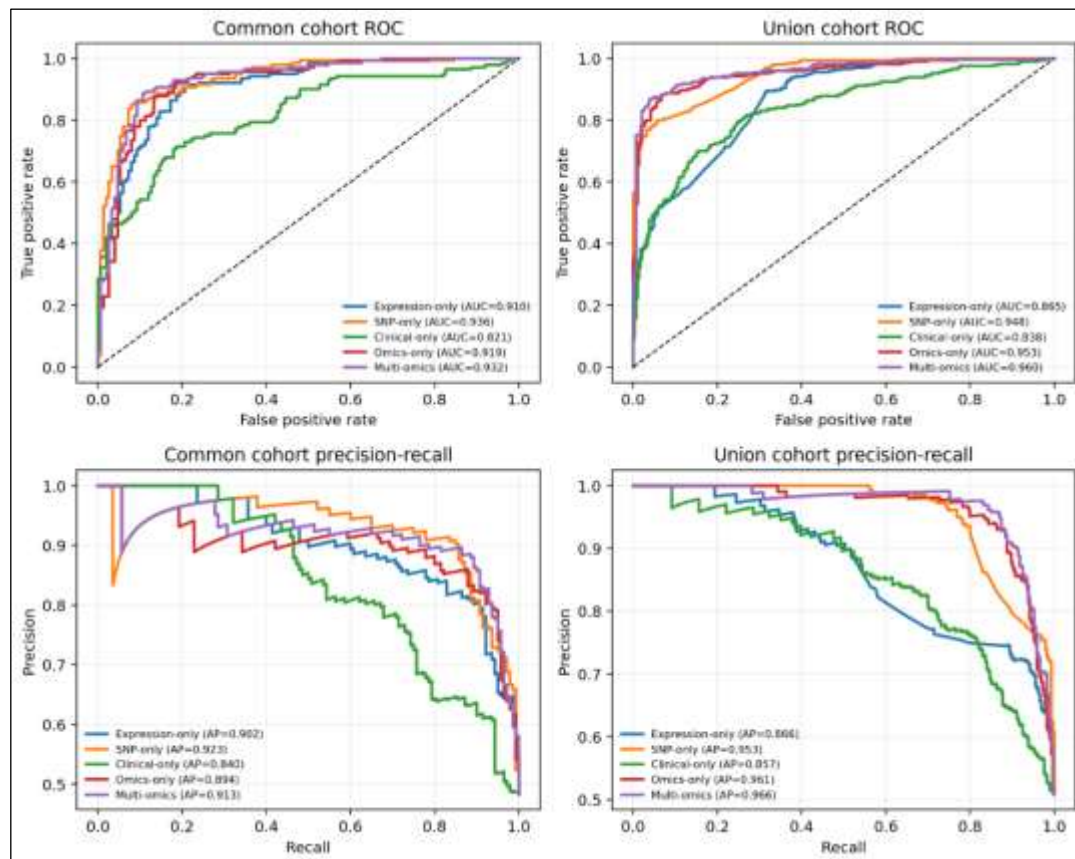


Figure 2: ROC and precision-recall curves in the common and union cohorts.

Source: Authors, (2026).

IV.2 COMPARATIVE POSITIONING RELATIVE TO PREVIOUS STUDIES

To position the present findings within the recent AD biomarker and multimodal-learning literature, Table 2 compares the current study with representative prior studies in terms of task design, modalities, use of cognitive variables, predictive performance and biological interpretability. The purpose of this comparison is not to claim strict superiority across non-identical datasets and endpoints but to clarify how the present framework differs methodologically and where it offers practical advantages.

Table 2: Comparison of the present study with representative AD classification studies.

Study	Data	Endpoint and design	Cognitive variables	Best performance	Biological interpretation	Main distinction
Lee and Lee, 2020 [9]	Blood gene expression	ADNI/AddNeuroMed; transcriptomics-only classification	No multimodal integration	Internal AUC 0.657 in ADNI	Pathway analysis of selected genes	Present study shows substantially stronger blood-expression-only performance and adds multimodal benchmarking
Venugopalan et al., 2021 [10]	MRI + clinical/EHR + SNP	Multimodal deep learning for AD/MCI/CN staging	Yes	Internal accuracies about 0.78–0.79	Limited coefficient-level interpretability	Present study uses a sparse and auditable model with stronger union-cohort discrimination and calibration

AlMansoori et al., 2024 [11]	Blood expression + SNP + clinical	ADNI; explainable ML	Best model used cognitive-score augmentation	AUC 0.94, ACC 0.95	Biomarker-oriented interpretation	Present study avoids cognitive-score dependence and retains stronger interpretability of modality contributions
Souchet et al., 2024 [12]	Targeted blood multi-omics + age	Prodromal AD / AD dementia vs non-AD disorders	No explicit multimodal feature benchmarking	Specificity 93.0%, sensitivity 65.4%, AUROC 81.9%	Targeted biomarker and pathway interpretation	Present study is a discovery-oriented high-dimensional framework rather than a targeted assay-validation design
Present study	Blood expression + SNP + non-cognitive clinical/MRI	CN vs AD; common and union cohorts; 5 scenarios	No cognitive scores used	Common: SNP-only AUC 0.936; Union: multi-omics AUC 0.960, AP 0.966, ACC 0.913	Stability ranking + DE + curated AD overlap + g:Profiler	Integrates strong predictive performance with explicit modality benchmarking and layered biological validation

Source: Authors, (2026).

Overall, Table 2 shows that the present study occupies a distinct position in the literature. Compared with earlier blood-expression studies, it recovers substantially stronger transcriptomic classification signal. Compared with multimodal deep-learning approaches, it achieves strong performance in a more transparent and auditable modeling framework. Compared with biomarker-centered multimodal studies that depend on cognitive-score augmentation, it addresses a stricter question by evaluating whether blood transcriptomics, inherited genetic variation and non-cognitive clinical/MRI context can still provide effective AD classification. Finally, compared with targeted assay-validation studies, it contributes a broader discovery-oriented framework that links performance with feature stability, differential expression, curated AD-gene overlap and pathway coherence.

IV.3 BIOLOGICAL RELEVANCE OF THE SELECTED FEATURES

The selected features demonstrated substantial consistency across cohorts and across analytical layers supporting the view that the retained panels captured biologically meaningful AD-related signal rather than arbitrary model-specific selection. This was evident in both the transcriptomic and SNP-derived branches, as well as in their agreement with differential-expression analysis, curated AD-gene overlap and pathway-level interpretation. Among the 50 most stable expression features, four genes in the common cohort overlapped the curated AD gene set (IQCK, ALPK2, SLC2A9 and PIK3R1), whereas two genes in the union cohort overlapped the same reference (IQCK and ICA1). In addition, 15 expression genes were shared between the common- and union-cohort top-50 lists, including IQCK, ADK, ATP9A, HES1, CCNA1, NKX3-1, PTGDS and NR4A3. This level of overlap suggests that the transcriptomic branch captured reproducible signal across cohort definitions rather than unstable background variation. At the transcript level, IQCK emerged as the strongest recurrent candidate because it appeared among the most stable expression features in both cohorts, overlapped the curated AD reference set and was also identified as differentially expressed in the common cohort. APOE was also notable, particularly because it encodes a component of the gamma-secretase complex central to amyloid processing. The common-cohort expression panel additionally highlighted NR4A3, ATP9A, HES1, PTGDS and ADK as recurrent features, whereas ICA1 was particularly notable in the union cohort given its reported link to APP processing and non-amyloidogenic signaling. The SNP branch showed a similarly coherent pattern.

In the common cohort, PTGS2 and CELF2 were among the leading mapped genes overlapping the curated AD list, whereas in the union cohort the highest-ranked signal was rs439401 at the APOE/APOC1 locus, followed by CYP19A1 and PTGS2. Across the top-50 SNP-derived mapped genes, seven were shared between the common and union cohorts, including PTGS2, CHST11, PARL and SLC9A9. Overall, these findings indicate that the sparse ElasticNet-derived SNP panels repeatedly retained biologically plausible AD-related loci rather than arbitrary mathematical selections. Differential-expression analysis reinforced this interpretation while also revealing a clear difference between cohort constructions. In the common cohort, the analysis identified a concise set of ten significant probes: TENM1, VNN3P, GOLGA6L22, PAN3, SPTLC1, APOE, IQCK, MTARC1, CRYBG1 and MS4A1. This narrow signal is consistent with the stricter cohort definition and may reflect a cleaner transcriptomic contrast between AD and controls. By contrast, the union cohort produced a substantially broader differential-expression profile, comprising 466 significant probes corresponding to 377 unique genes. This broader pattern likely reflects richer biological signal but also greater heterogeneity, an interpretation supported by the prominence of sex-associated transcripts such as XIST, RPS4Y1, KDM5D and DDX3Y. Even so, the union-cohort differential-expression results reinforced several biologically relevant genes, including APOE, IQCK, MTARC1, PTGDS, BLNK, FCRL1/FCRL2, ECHDC3 and LTB4R. These cohort-dependent differences in transcriptomic signal are illustrated in Figure 3, where the common cohort shows a narrower and cleaner profile, while the union cohort displays a broader but more heterogeneous differential-expression pattern.

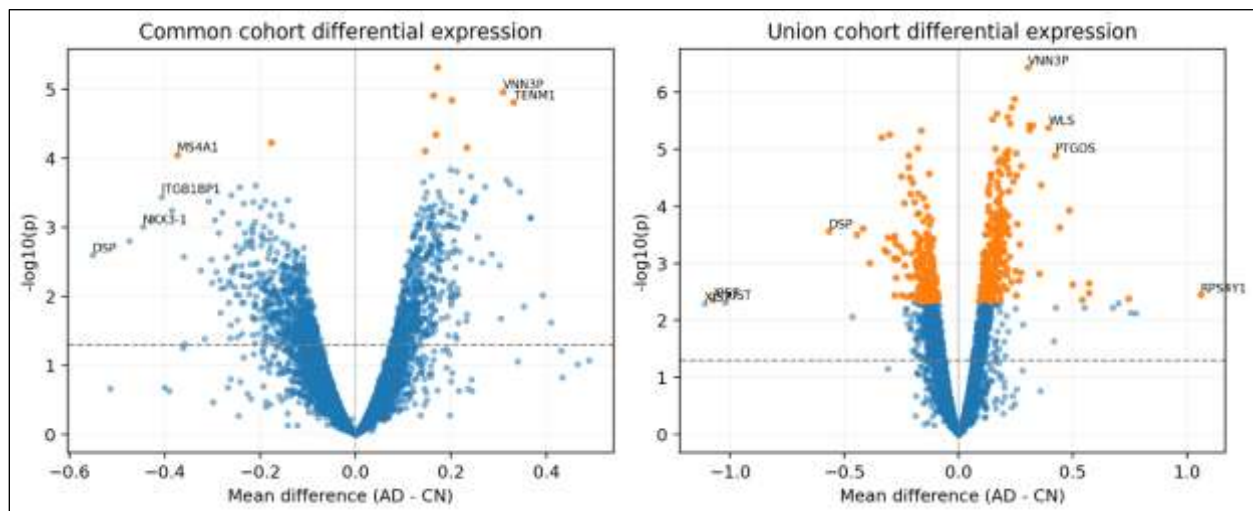


Figure 3: Differential-expression volcano plots for the common and union cohorts.

Source: Authors, (2026).

When considered jointly, these findings support organizing the candidate biomarkers into two evidence-based tiers. The first tier comprises the more strongly supported signals, including IQCK, APH1B, rs439401 at the APOE/APOC1 locus, PTGS2-linked loci, ADK and ICA1. These markers are supported by convergent evidence across multiple validation layers, including stability-based feature selection, overlap with curated AD-related genes, differential-expression results and prior biological literature. The second tier includes biologically plausible but more exploratory candidates such as CHST11, SLC9A9, PARL, GOLGA6L22, LRRC38 and UPP2. Presenting the biomarkers in this tiered way is preferable to a single undifferentiated list because it distinguishes findings supported by stronger convergent evidence from those that are more appropriately viewed as hypothesis-generating. At the broader feature-set level, overlap with the curated AD-related gene list was substantial in both cohorts. In the common cohort, 192 expression-derived genes and 108 SNP-mapped genes overlapped the curated AD list, with 28 genes shared between modalities; in the union cohort, the corresponding values were 196, 123 and 36. The g:Profiler analysis complemented these overlap counts by shifting the interpretation from individual markers to pathway-level coherence. Rather than asking only whether single genes overlapped known AD genes, this step assessed whether the selected panels converged on coherent biological functions. In line with both the selected gene content and the differential-expression findings, the resulting pathway profile supported themes related to immune and inflammatory regulation, synaptic and neuronal signaling, lipid and membrane biology and amyloid-related mechanisms. These orthogonal biological-validation layers are summarized in Table 3.

Table 3: Orthogonal biological validation summary for the selected feature sets.

Validation layer	Common cohort	Union cohort	Representative evidence	Interpretation
Top-50 stable expression features vs curated AD list	4 overlaps in top 50	2 overlaps in top 50	IQCK; ALPK2; ICA1	Targeted plausibility test for transcriptomic markers
Top-50 stable SNP features vs curated AD list	2 overlaps in top 50	3 overlaps in top 50	PTGS2; CELF2; APOE/APOC1; CYP19A1	Selected variants intersect known AD genetics
All selected mapped genes vs curated AD list	192 expression, 108 SNP, 28 shared	196 expression, 123 SNP, 36 shared	APP, APH1B, BACE2, CR1, ABCA7, AKT3	Strong overall AD signal across the consensus panels
Differential expression	10 significant probes / 10 genes	466 significant probes / 377 genes	TENM1, VNN3P, APH1B, IQCK, PTGDS	Orthogonal expression-level confirmation
g:Profiler2 pathway interpretation	Applied to stable feature-derived gene sets	Applied to stable feature-derived gene sets	Immune, synaptic, lipid and amyloid-related themes	Systems-level coherence beyond single-gene overlap

Source: Authors, (2026).

IV.4 PRACTICAL IMPLICATIONS, METHODOLOGICAL CONTRIBUTION AND LIMITATIONS

From an engineering and translational perspective, the main contribution of this study lies not only in the predictive performance achieved, but also in the development of a reproducible and interpretable framework for high-dimensional, heterogeneous biomedical data. The common cohort showed that SNPs were the strongest unimodal discriminators, whereas the union cohort demonstrated the clearest overall advantage of multi-omics integration. This indicates that the practical value of integration is not simply maximizing one metric, but achieving stronger overall behavior across discrimination, precision-recall performance, specificity and calibration. A particularly important design choice in the present study was the exclusion of clinical cognitive variables from the predictive feature set. Models that include cognitive test scores may partly recover the diagnostic construct itself rather than independent biomarker signal, because cognitive decline is already closely linked to clinical diagnosis [20]. By restricting the clinical branch to non-cognitive demographic and structural/MRI variables, the present framework asks a stricter and more biomarker-oriented question: whether blood transcriptomics, inherited genetic variation and non-cognitive phenotypic context can still classify AD effectively in an interpretable setting.

This makes the multimodal result more informative biologically, because the gain cannot be attributed simply to including variables that are already very close to the diagnostic label. Methodologically, the study contributes a transparent and modular workflow rather than a single black-box classifier. A separate stability-selection stage was first used to derive fixed reduced expression and SNP panels and these panels were then evaluated across five explicitly defined scenarios. This structure makes the contribution of each modality visible, strengthens reproducibility through fold-restricted preprocessing and training-only threshold selection and links predictive performance to downstream biological interpretation through stability ranking, differential expression, curated AD-gene overlap and pathway enrichment. From an engineering standpoint, this improves traceability, auditability and practical reuse in similar high-dimensional biomedical applications. The study also reinforces a broader principle in AD multi-omics research: upstream cohort-definition strategy can materially influence downstream biomarker and pathway interpretation. In the present work, the common cohort provided a cleaner complete-data benchmark for direct modality comparison, whereas the union cohort preserved more samples and revealed the clearest overall advantage of multi-omics integration. At the biological level, this difference was also reflected in the downstream omics interpretation: the common cohort yielded a narrower transcriptomic signal, while the union cohort produced a broader but more heterogeneous profile. This pattern is consistent with recent multi-omics work showing that how subjects are grouped can shape the molecular signals that emerge from AD data analysis. These findings should nevertheless be interpreted within the limits of internal validation. All analyses were conducted within ADNI and no independent external cohort was used to assess generalizability. In addition, the broader transcriptomic signal observed in the union cohort suggests that sex-related structure and other cohort-level effects may still contribute to part of the differential-expression pattern. The prioritized genes and loci should therefore be regarded as candidate biomarkers pending external replication and cross-platform validation. Overall, this study shows that a sparse ElasticNet-based multi-omics pipeline can provide a strong balance between predictive effectiveness, interpretability and biological coherence. Its practical significance lies in demonstrating that heterogeneous biomedical data can be integrated within a transparent and auditable framework while preserving clear links to plausible disease-related mechanisms.

V. CONCLUSIONS

This study proposed an interpretable ElasticNet-based multi-omics framework for distinguishing cognitively normal individuals from patients with Alzheimer's disease using ADNI data. Across the two analytical cohorts, SNP-based modeling provided the strongest unimodal performance, whereas the integrated multi-omics configuration achieved the best overall results in the broader union cohort, particularly in terms of discrimination, calibration and balanced classification performance. Beyond predictive accuracy, the main contribution of this work is the development of a transparent and reproducible analytical workflow that combines fold-restricted preprocessing, stability-based feature reduction, regularized benchmarking and downstream biological interpretation. This framework not only enabled direct comparison of five modality configurations, but also supported the identification of recurrent AD-related candidate signals, including IQCK, APH1B, PTGS2 and the APOE/APOC1 locus. Although the analysis remains internally validated within ADNI and therefore requires confirmation in independent cohorts, the results demonstrate that sparse multi-omics integration can provide a useful balance between predictive performance, interpretability and biological plausibility in Alzheimer's disease classification. The proposed workflow may therefore serve as a practical foundation for future blood-based and multimodal biomarker studies in neurodegenerative disease.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Aya Abdelaziz, Amira Gherboudj, Tunahan Çakir, Tunahan Çakir, Mohamed Abdelhafid Hamidechi.

Methodology: Aya Abdelaziz, Amira Gherboudj, Tunahan Çakir, Tunahan Çakir, Mohamed Abdelhafid Hamidechi.

Investigation: Aya Abdelaziz, Amira Gherboudj, Tunahan Çakir, Tunahan Çakir, Mohamed Abdelhafid Hamidechi.

Discussion of results: Aya Abdelaziz, Amira Gherboudj, Tunahan Çakir, Tunahan Çakir, Mohamed Abdelhafid Hamidechi.

Writing – Original Draft: Aya Abdelaziz, Amira Gherboudj, Tunahan Çakir, Tunahan Çakir, Mohamed Abdelhafid Hamidechi.

Writing – Review and Editing: Aya Abdelaziz, Amira Gherboudj, Tunahan Çakir, Tunahan Çakir, Mohamed Abdelhafid Hamidechi.

Resources: Aya Abdelaziz, Amira Gherboudj, Tunahan Çakir, Tunahan Çakir, Mohamed Abdelhafid Hamidechi.

Supervision: Aya Abdelaziz, Amira Gherboudj, Tunahan Çakir, Tunahan Çakir, Mohamed Abdelhafid Hamidechi.

Approval of the final text: Aya Abdelaziz, Amira Gherboudj, Tunahan Çakir, Tunahan Çakir, Mohamed Abdelhafid Hamidechi.

VII. REFERENCES

- [1] J. Rayathala, K. K. C. and V. P., "Review on Alzheimer's disease: past, present and future," *J Inn App Pharma Sci*, pp. 28–31, Jan. 2022, doi: 10.37022/jiaps.v7i1.274.
- [2] Alzheimer's Association, "2025 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 21, no. 4, p. e70235, Apr. 2025, doi: 10.1002/alz.70235.
- [3] L. Xu, Z. Wang, M. Li, and Q. Li, "Global incidence trends and projections of Alzheimer disease and other dementias: an age-period-cohort analysis 2021," *J Glob Health*, vol. 15, p. 04156, May 2025, doi: 10.7189/jogh.15.04156.
- [4] P. d'Errico and M. Meyer-Luehmann, "Mechanisms of Pathogenic Tau and A β Protein Spreading in Alzheimer's Disease," *Front. Aging Neurosci.*, vol. 12, p. 265, Aug. 2020, doi: 10.3389/fnagi.2020.00265.
- [5] T. E. Golde, "Alzheimer's disease – the journey of a healthy brain into organ failure," *Mol Neurodegeneration*, vol. 17, no. 1, p. 18, Mar. 2022, doi: 10.1186/s13024-022-00523-1.
- [6] S. E. Ankeny, J. R. Bacci, B. Decourt, M. N. Sabbagh, and M. M. Mielke, "Navigating the Landscape of Plasma Biomarkers in Alzheimer's Disease: Focus on Past, Present, and Future Clinical Applications," *Neurol Ther*, vol. 13, no. 6, pp. 1541–1557, Dec. 2024, doi: 10.1007/s40120-024-00658-x.
- [7] D. P. Veitch et al., "The Alzheimer's Disease Neuroimaging Initiative in the era of Alzheimer's disease treatment: A review of ADNI studies from 2021 to 2022," *Alzheimer's & Dementia*, vol. 20, no. 1, pp. 652–694, Jan. 2024, doi: 10.1002/alz.13449.

- [8] S. Polevikov, “Advancing AI in healthcare: A comprehensive review of best practices,” *Clinica Chimica Acta*, vol. 548, p. 117519, Aug. 2023, doi: 10.1016/j.cca.2023.117519.
- [9] T. Lee and H. Lee, “Prediction of Alzheimer’s disease using blood gene expression data,” *Sci Rep*, vol. 10, no. 1, p. 3485, Feb. 2020, doi: 10.1038/s41598-020-60595-1.
- [10] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, “Multimodal deep learning models for early detection of Alzheimer’s disease stage,” *Sci Rep*, vol. 11, no. 1, p. 3254, Feb. 2021, doi: 10.1038/s41598-020-74399-w.
- [11] M. E. AlMansoori, S. Jemimah, F. Abuhantash, and A. AlShehhi, “Predicting early Alzheimer’s with blood biomarkers and clinical features,” *Sci Rep*, vol. 14, no. 1, p. 6039, Mar. 2024, doi: 10.1038/s41598-024-56489-1.
- [12] B. Souchet et al., “Multiomics Blood-Based Biomarkers Predict Alzheimer’s Predementia with High Specificity in a Multicentric Cohort Study,” *The Journal of Prevention of Alzheimer’s Disease*, vol. 11, no. 3, pp. 567–581, May 2024, doi: 10.14283/jpad.2024.34.
- [13] H. B. Lüleci, A. Jones, and T. Çakır, “Multi-omics analyses highlight molecular differences between clinical and neuropathological diagnoses in Alzheimer’s disease,” *European Journal of Neuroscience*, vol. 60, no. 5, pp. 4922–4936, 2024, doi: 10.1111/ejn.16482.
- [14] L. Kolberg, U. Raudvere, I. Kuzmin, P. Adler, J. Vilo, and H. Peterson, “g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update),” *Nucleic Acids Research*, vol. 51, no. W1, pp. W207–W212, Jul. 2023, doi: 10.1093/nar/gkad347.
- [15] “ADNI | Alzheimer’s Disease Neuroimaging Initiative.” Accessed: Mar. 27, 2026. [Online]. Available: <https://adni.loni.usc.edu/>
- [16] Y.-S. Hu, J. Xin, Y. Hu, L. Zhang, and J. Wang, “Analyzing the genes related to Alzheimer’s disease via a network and pathway-based approach,” *Alzheimer’s Research & Therapy*, vol. 9, no. 1, p. 29, Apr. 2017, doi: 10.1186/s13195-017-0252-z.
- [17] S. J. Andrews, B. Fulton-Howard, and A. Goate, “Interpretation of risk loci from genome-wide association studies of Alzheimer’s disease,” *The Lancet Neurology*, vol. 19, no. 4, pp. 326–335, Apr. 2020, doi: 10.1016/S1474-4422(19)30435-1.
- [18] C. Bellenguez et al., “New insights into the genetic etiology of Alzheimer’s disease and related dementias,” *Nat Genet*, vol. 54, no. 4, pp. 412–436, Apr. 2022, doi: 10.1038/s41588-022-01024-z.
- [19] D. P. Wightman et al., “A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer’s disease,” *Nature Genetics*, vol. 53, no. 9, pp. 1276–1282, Sep. 2021, doi: 10.1038/s41588-021-00921-z.
- [20] G. M. McKhann et al., “The diagnosis of dementia due to Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease,” *Alzheimer’s & Dementia*, vol. 7, no. 3, pp. 263–269, May 2011, doi: 10.1016/j.jalz.2011.03.005.