

1 **WIMOAD: Weighted Integration of Multi-Omics data for Alzheimer's Disease (AD)**

2 **Diagnosis**

3 Hanyu Xiao¹, Jieqiong Wang^{2*}, Shibiao Wan^{1*}

4

5 ¹ Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center,

6 Omaha, NE, United States, 68198

7 ² Department of Neurological Sciences, University of Nebraska Medical Center, Omaha, NE,

8 United States, 68198

9 *Corresponding contact:

10 Dr. Shibiao Wan, swan@unmc.edu

11 Dr. Jieqiong Wang, jiwang@unmc.edu

12 **Abstract**

13 As the most common subtype of dementia, Alzheimer's disease (AD) is characterized by a
14 progressive decline in cognitive functions, especially in memory, thinking, and reasoning ability.
15 Early diagnosis and interventions enable the implementation of measures to reduce or slow further
16 regression of the disease, preventing individuals from severe brain function decline. The current
17 framework of AD diagnosis depends on A/T/(N) biomarkers detection from cerebrospinal fluid or
18 brain imaging data, which is invasive and expensive during the data acquisition process. Moreover,
19 the pathophysiological changes of AD accumulate in amino acids, metabolism, neuroinflammation,
20 etc., resulting in heterogeneity in newly registered patients. Recently, next generation sequencing
21 (NGS) technologies have found to be a non-invasive, efficient and less-costly alternative on AD
22 screening. However, most of existing studies rely on single omics only. To address these concerns,
23 we introduce WIMOAD, a weighted integration of multi-omics data for AD diagnosis. WIMOAD
24 synergistically leverages specialized classifiers for patients' paired gene expression and
25 methylation data for multi-stage classification. The resulting scores were then stacked with MLP-
26 based meta-models for performance improvement. The prediction results of two distinct meta-
27 models were integrated with optimized weights for the final decision-making of the model,
28 providing higher performance than using single omics only. Remarkably, WIMOAD achieves
29 significantly higher performance than using single omics alone in the classification tasks. The
30 model's overall performance also outperformed most existing approaches, highlighting its ability
31 to effectively discern intricate patterns in multi-omics data and their correlations with clinical
32 diagnosis results. In addition, WIMOAD also stands out as a biologically interpretable model by
33 leveraging the SHapley Additive exPlanations (SHAP) to elucidate the contributions of each gene
34 from each omics to the model output. We believe WIMOAD is a very promising tool for accurate

35 AD diagnosis and effective biomarker discovery across different progression stages, which
36 eventually will have consequential impacts on early treatment intervention and personalized
37 therapy design on AD.

38 **Keywords:** Alzheimer's Disease, Multi-omics, Weighted Score Fusion, Early Diagnosis, DNA
39 Methylation

40

41 **Introduction**

42 Alzheimer's disease (AD) is the most common subtype of dementia, characterized by a
43 progressive decline in cognitive functions, notably in memory, thinking, and reasoning [1]. It is
44 closely associated with aging and exerts a persistent impact on cognitive functions [2]. With a
45 national care cost growth of \$24 billion from a year ago, reaching \$345 billion overall in 2023,
46 this neurodegenerative disease poses significant challenges for individuals and their families [3].
47 But according to previous study [4], AD is not an inevitable process of aging and there is the
48 possibility to prevent or delay the development of this demensia in certain proportion of people.
49 For primary healthcare and disease screening, the ability to achieve early and efficient diagnosis
50 of AD is crucial for effective intervention and treatment [5].

51 Typically, AD is characterized by the A/T/N framework [6]. The "A" component refers to
52 amyloidosis-beta peptide accumulation [7–9], and the "T" aspect, tauopathy, represents
53 hyperphosphorylated tau protein aggregation [10,11]. The "N" component, focusing on specific
54 aspects of neurodegeneration [12], gives an overall picture of neuronal and synaptic loss in the
55 patients' brains. So far, the majority of research relies on phenotypic data, particularly brain
56 imaging like Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and Positron
57 Emission Tomography (PET) [13,14]. With the advancements in artificial intelligence (AI)

58 algorithms [15], Chen et al. [16] have implanted U-Net, Multi-Layer Perceptron, and Graph Neural
59 Network for 3-class AD diagnosis, and Al-Otaibi et al. [17] demonstrate the deep transfer learning
60 on brain imaging with AutoEncoder structure, providing high classification performance. To
61 aggregate different information extracted from multiple types of images, MMTFN introduced by
62 Miao et al. [18] constructs a 3D multi-scale residual block layers and a Transformer network that
63 jointly learns the representations from MRI and PET images of 720 subjects and gets a 94.61%
64 accuracy between AD and Normal Control. Although the models are promising, utilizing the
65 imaging data as model inputs results in However, idealized brain imaging of patients remains
66 limited, and the neuropathological diagnosis is invasive and harmful to patients [19]. As
67 pathophysiological changes gradually accumulate in amino acids, metabolism, and
68 neuroinflammation, newly registered patients show considerable heterogeneity in the impaired
69 cognitive domains which will lead to increasing diagnostic costs [20,21], underscoring the need
70 for more precise and individualized diagnostic approaches [22–24].

71 With the progress in sequencing techniques, genetic data is increasingly being utilized as
72 external validation in AD studies as the less-expensive and less-invasive measurement [25]. For
73 example, researchers have identified many genetic risk factors for AD (e.g., *APOE* [26], *CRI* [27],
74 *ABCA7* [28], etc.) identified by Single Nucleotide Polymorphism (SNP) in Genome-Wide
75 Association Studies (GWAS) [29,30]. Transcriptomic analysis is also essential for biomarker
76 detection in complex diseases like AD. Guo and Yang [31] applied a transcriptome-wide
77 association study (TWAS) with reference transcriptomic data from brain and blood tissues and
78 detected 141 risk genes while Methys et al. [32] utilized advanced single-cell transcriptome
79 analysis and found cell-type specific disease-associated changes across various degrees of AD,
80 which can provide a molecular and cellular foundation for further investigation. As one of the main

81 components of the epigenetic data and highly correlated with aging [33], DNA methylation level
82 is found to be increased in peripheral cells of AD patients while correlating with worse cognitive
83 performances and *APOE* polymorphism [34,35]. However, considering the intricate nature of the
84 aging process and the progression of neurodegenerative disorders, relying on one data modality
85 only may underestimate other related risk factors in this complicated process, since one omics can
86 not convey all the information needed.

87 To enhance the effectiveness of current AD research, integrating genetic data could greatly
88 improve the accuracy, reliability, and interpretability of the computational model [36–38].
89 However, how to combine data from different omics layers to provide a holistic view of biological
90 systems remains the major challenge of this field. One general solution is to summarize all results
91 from transcriptomic, proteomics, metabolomics, etc., on brains and other tissues and form a
92 comprehensive understanding of the impact of one gene alterations in individual clinical
93 trajectories [39–42]. Factor analysis, which represents high-dimensional variables to a smaller
94 number of latent factors, is also brought up in multi-omics research (MOFA, multi-omics factor
95 analysis) [43]. iCluster [44], JIVE [45], and SLIDE [46] are all commonly used tools that jointly
96 model associations and the variance-covariance structure within each data type while reducing the
97 dimensionality for clustering. In AD studies, Bao et al. [47] proposed a structural Bayesian factor
98 analysis framework named SBFA that incorporates imaging and biological data for functional
99 assessment questionnaire (FAQ) score prediction. In addition, various integration or ‘fusion’
100 methodologies have been introduced through data concatenation with AI-based algorithms [48–
101 50], but models that focus on AD studies are rare [51]. Clinical information is also incorporated in
102 the integration process for better diagnosis performance [52].

103 Despite these advancements, significant gaps remain in integration studies. Firstly, most
104 genomic studies focus on SNPs or gene expression data, with less attention on methylation data,
105 which is highly related to aging and AD [53–55]. Secondly, widely used direct data concatenation
106 [56] for integration may lose some key information for each data modality, as each omics will have
107 different representations and data formats. To fill this gap, we proposed WIMOAD, which assigns
108 distinct weights for the prediction score of each omics classifier and integrate the results from
109 different data modalities to do the final decision-making, for different stages diagnosis of AD. Our
110 major contribution can be summarized as follows:

- 111 (1) We proposed a stacked weighted score-based multiomics (gene expression and methylation
112 data from ADNI) fusion model for Alzheimer’s disease diagnosis, which has surpassed the
113 performance of using single omics alone, as well as the existing integration methods.
- 114 (2) The stacking part of the ensemble model has dramatically improved the overall classification
115 outcome on both single omics and the integration of two omics
- 116 (3) The proposed model is accurate, easy to use, time-saving, and interpretable from a biological
117 view as we apply the Shapley Value [57] to quantify the contribution of individual genes for
118 model decision-making, which will help for new biomarker detection.

119

120 **Materials and Methods**

121 **Datasets**

122 The data used in this paper are from the genetic section of the Alzheimer’s Disease
123 Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI is a longitudinal multicenter
124 study that collected clinical, imaging, genetic, and biochemical biomarkers for early detection and
125 tracking of recruited cohorts across different time points. For our model, we collected the data of

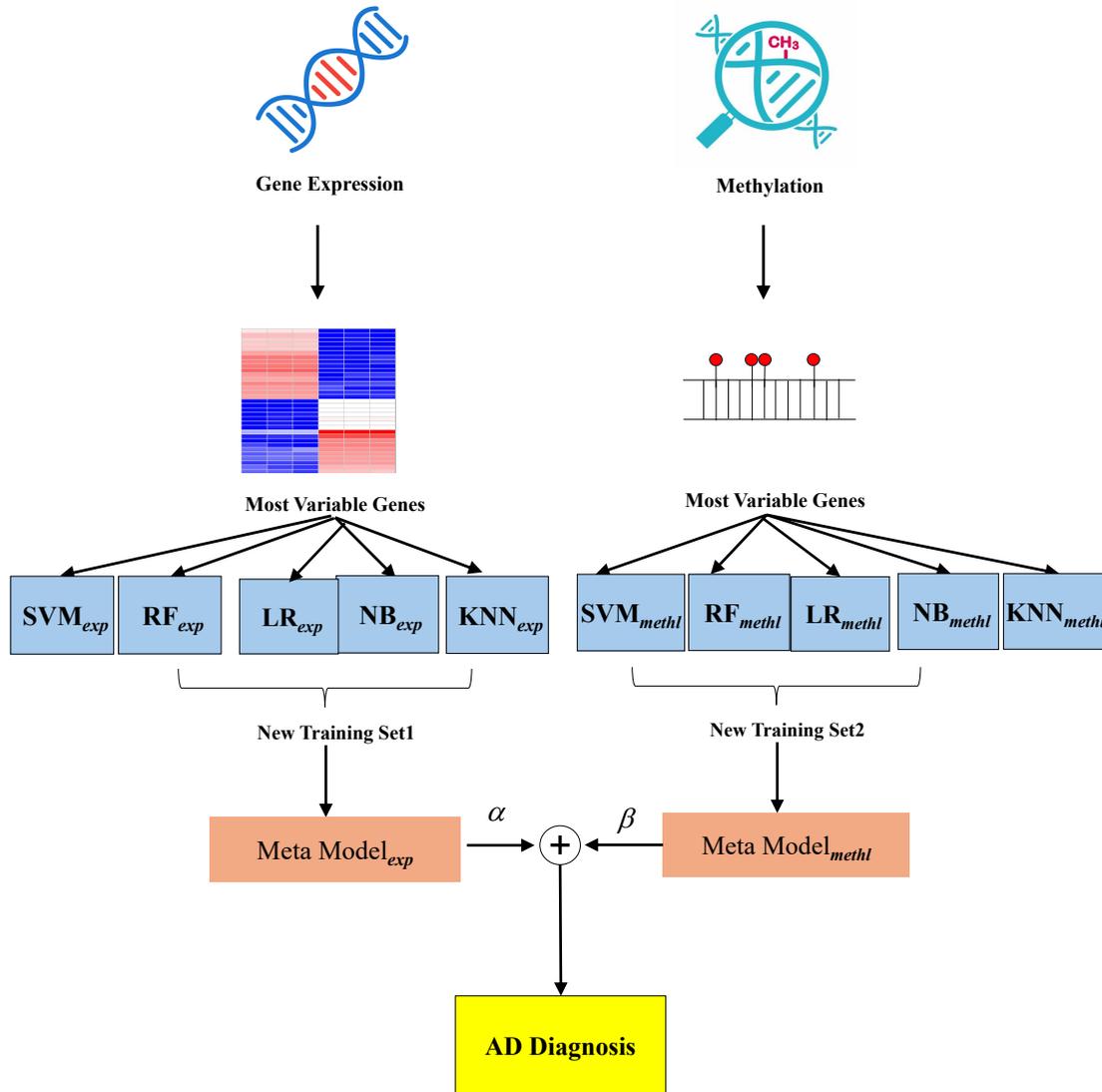
126 591 people's gene expression and methylation profiles as model input following the criteria that
127 the genetic profiles from different omics are paired for a certain sample (Originally we have 744
128 gene expression profiles and 649 methylation data records. The rest of the samples which only has
129 one omics data were eliminated). Among them, there are 203 Normal Controls (CN) subjects (age:
130 74.45 ± 5.78 , F/M: 101/102), 180 Early Mild Cognitive Impairments (EMCI) subjects (age: 71.44
131 ± 7.11 , F/M: 81/ 99), 113 Late Mild Cognitive Impairments (LMCI) subjects (age: 72.74 ± 7.67 ;
132 F/M: 45/68), which is 293 Mild Cognitive Impairments (MCI) and 95 Alzheimer's Diseases (AD)
133 (age: 74.28 ± 7.59 , F/M: 35/60). The demographic information of the data is shown in **Table.1**.
134 For subsequent binary group classification tasks, we have reprocessed the original categories as
135 follows: all samples, excluding the AD group, were categorized into a "patient" (PT) group to
136 facilitate 'PT-AD' binary classification. Furthermore, the EMCI and LMCI groups were combined
137 into a single MCI group, enabling the execution of other binary classification tasks related to MCI.
138 **Table. 1 The demographic information of the Selected Participants.** Data are mean \pm standard
139 deviation (std). CN: Normal Controls; EMCI: Early Mild Cognitive Impairments; LMCI: Late
140 Mild Cognitive Impairments; MCI: Mild Cognitive Impairments; AD: Alzheimer's Diseases; F:
141 Female; M: Male

Diagnosis	Samples	Age (mean \pm std)	Sex (F/M)
CN	N = 203	74.45 ± 5.78	101/102
EMCI	N = 180	71.44 ± 7.11	81/ 99
LMCI	N = 113	72.74 ± 7.67	45/68
AD	N = 95	74.28 ± 7.59	35/60

142

143 **Overview of WIMOAD Framework**

144 WIMOAD is a weighted score fusion model based on combining multiple base classifiers
145 [58]. The pipeline is shown in **Fig. 1**. After establishing the database, gene expression and
146 methylation data were extracted and paired according to patient ID to serve as model inputs. The
147 model processed these omics separately, extracting the most variable genes from both omics within
148 two categorized groups to use as features. For each data type, five commonly used machine
149 learning classifiers, Support Vector Machine (SVM) [59], Random Forest (RF) [60], Naïve Bayes
150 (NB) [61], Logistic Regression (LR) [62] and K-Nearest Neighbors (KNN) [63], were applied
151 independently to create new training sets with the prediction scores for meta-models, feedforward
152 Multi-Layer Perceptron (MLP) [64,65]. Finally, the meta-model prediction results from both gene
153 expression and methylation were combined using a weighted fusion mechanism [56]. The
154 ensemble result was used to make the final decision on AD diagnosis. Subsequent optimization
155 was performed for each classifier and the ensemble weight to enhance the integration model
156 performance. The model was validated under 10 times 10-fold cross-validation (CV). In each CV
157 round, the predicted score of each model was linearly combined by assigned weights for the final
158 decision of the whole model. Once trained, the models were interpreted using SHAP to explain
159 the results.



160

161 **Fig. 1. The Workflow of WIMOAD.** The process begins by identifying the most variable features
162 from paired gene expression and methylation data for classification. For each omics data, different
163 classifiers were trained. The outputs of the basic classifiers were considered as the new training
164 sets for two distinct meta-models, which used the predictions of base classifiers as inputs and
165 generated the overall prediction scores. For multi-omics integration, each meta-model is assigned
166 a weight for ensemble learning, which also controls the contributions of each meta-model to the
167 final decision. SVM_{exp}: Support Vector Machine for gene expression data. SVM_{methyl}: Support

168 *Vector Machine for gene methylation data. RF: Random Forest classifier. NB: Naïve Bayes*
169 *classifier. KNN: K-Nearest Neighbor classifier. LR: Logistic Regression.*

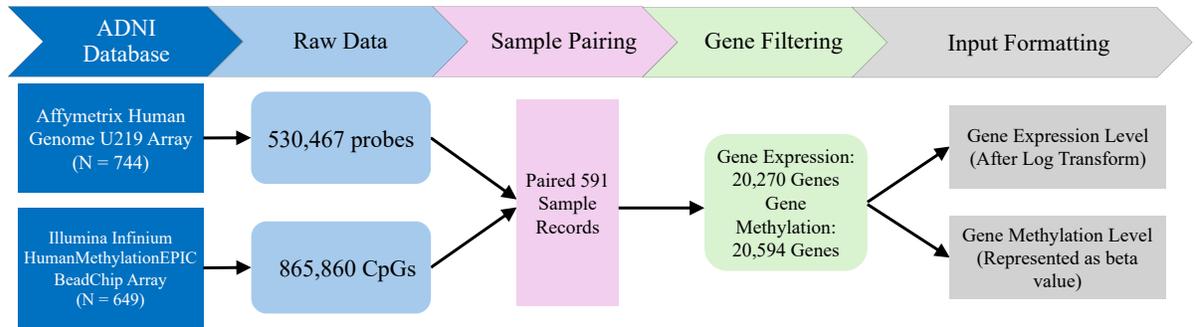
170

171 **Preprocessing of multi-omics data**

172 The gene expression profiling was provided with Affymetrix Human Genome U219 Array
173 from peripheral blood samples. The raw expression values generated by this platform were first
174 normalized using the Robust Multi-chip Average (RMA) method, resulting in 530,467 probes
175 corresponding to 49,293 transcripts from 744 samples. These probes were subsequently mapped
176 and annotated according to the human genome reference (hg19). Given that a single gene may be
177 associated with multiple probes, we selected the probe data corresponding to the first occurrence
178 of each gene in the processed matrix to represent the expression level of that gene for each
179 individual. Genes with missing information in the annotated data were excluded from further
180 analysis. Finally, the filtered data contains 20,270 annotated genes, and the expression matrix
181 underwent a log transformation for scaling, which aimed to improve the accuracy of classification
182 results.

183 Whole-genome DNA methylation profiling was conducted using the Illumina Infinium
184 HumanMethylationEPIC BeadChip Array. The original data samples were normalized with the
185 dasen method for downstream quality control (QC) including p-value criteria filtering, sex and
186 sample ID verification, with 649 samples remained. The database provided raw data for these 649
187 participants who had undergone the QC process for further analysis. We obtained beta values for
188 a total of 865,860 CpG sites by analyzing the channel signals. These CpG sites were subsequently
189 mapped to the human genome reference (hg19), resulting in methylation data for 20,594 genes.

190 The workflow of the multi-omics data preparation is summarized in **Fig. 2**.



191

192 **Fig. 2. The Preprocessing Steps for Multi-Omics Data.**

193 **Feature Selection**

194 For a supervised learning model, in the case of a high-dimensionality curse and to enhance
 195 prediction efficiency while simultaneously reducing the consumption of computational resources,
 196 feature selection is a key process for model prediction. We selected 1000 genes that show
 197 statistically significant within-group variance separately for different omics inputs based on the
 198 ANOVA F-value [66] calculated by the ‘SelectKBest’ package in scikit-learn with ‘f_classif’
 199 function. For comparison, we also employed median absolute deviation (MAD) and Fano factor
 200 for gene selection [67].

201

202 **Weighted Score Fusion**

203 In omics integration research, a common approach is to concatenate different types of data
 204 directly before classification. However, in this study, Exp and Methl data exhibit substantial
 205 differences in their representations and feature characteristics, which will result in suboptimal
 206 classification outcomes when directly concatenated or combined pairwise. Consequently, we
 207 employed a score fusion method to construct an integration model for multi-omics data. Initially,
 208 we assigned trained meta-model to each dataset separately for binary classification. Subsequently,

209 we performed a weighted linear aggregation of the obtained prediction scores to derive the final
210 prediction score of the model, which calculated as:

$$S_{Int} = \alpha * S_{Exp} + \beta * S_{Methl} \quad (1)$$

s.t. $\alpha + \beta = 1, \alpha \geq 0, \beta \geq 0$

211
212 Where S_{Int} is the integrated prediction score of two meta-models, which represents the
213 probability of a given sample belonging to a specific class. S_{Exp} as the score generated by the
214 gene expression meta-model and S_{Methl} as the score generated by the gene methylation meta-
215 model. The α and β are the weight coefficients to balance the scores. These coefficients are
216 determined by the validation data in the 10 times 10-fold CV through screening from $\alpha = 0$ to α
217 =1 in the linear combination. This approach ensures a more accurate and interpretable integration
218 of the diverse omics data types, accommodating the unique features of each dataset and enhancing
219 the overall classification performance.

220

221 **Evaluation of the Model Performance**

222 10 times 10-fold CV [68] was used to evaluate our WIMOAD. Specifically, we measured
223 accuracy (Acc), precision (Prec), Recall (Rec), F1-Score (F1), Matthews correlation coefficient
224 (MCC), Specificity (Sp), G-measure (G), Jaccard Index (Jacc) and Area Under Curve (AUC):

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$Prec = \frac{TP}{TP + FP} \quad (2)$$

$$Rec = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (4)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

$$Sp = \frac{TN}{TN + FP} \quad (6)$$

$$G = \sqrt{\left(\frac{TP}{TP + FP}\right) \times \left(\frac{TP}{TP + FN}\right)} \quad (7)$$

$$Jacc = \frac{TP}{TP + FP + FN} \quad (8)$$

225

226 **Model Interpretation with SHAP**

227 To develop an explainable model, we utilized the Kernel SHAP Explainer [57,69] for
 228 multi-kernel classifiers for different omics input. Given that different omics data modalities convey
 229 distinct types of information, interpreting each modality separately allows us to identify key genes
 230 contributing to the prediction results, providing a comprehensive understanding of the biological
 231 processes involved and highlighting critical genes that may be overlooked when considering a
 232 single data source. In addition, since we introduced the stacking strategy, multiple explainers were
 233 applied to different classifiers in each omics to see whether there are overlaps among the base
 234 models in contributing gene selection. We filtered the top 10 genes in this process for each kernel
 235 explainer based on the selected features and the running time.

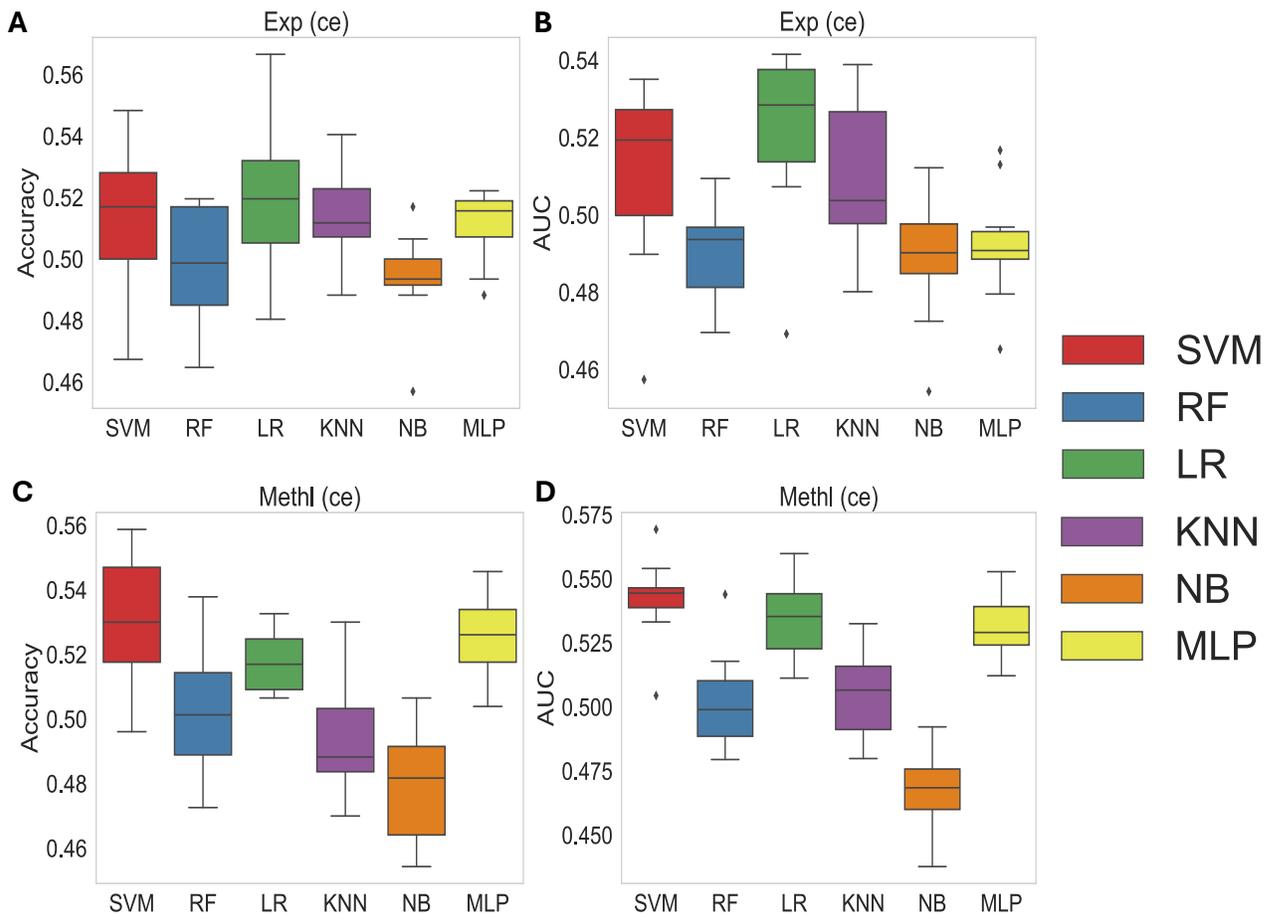
236

237 **Results**

238 **Machine Learning and Deep Learning Classifiers Comparison for Selected Samples**

239 WIMOAD In this paper, we initially selected SVM, Random Forest (RF), Logistic
 240 Regression (LR), K-Nearest Neighbors (KNN), Naïve Bayes (NB) and Multilayer Perceptron
 241 (MLP) as the classifiers. The Accuracy and AUC evaluation metrics of these classifiers's
 242 performance on gene expression (Exp) and methylation (Methl) data with group CN vs. EMCI are

243 shown in **Fig. 3**. With 10 times 10-fold CV, no classifier shew a performance higher than 60% in
 244 both accuracy and AUC. In addition, as a commonly used deep learning method, convolutional
 245 Multilayer Perceptron (MLP) did not exhibit higher AUC scores than conventional machine
 246 learning classifiers in the majority of groups for both omics. We finally applied some of the
 247 commonly used classifiers as the base models for further stacking study to achieve higher overall
 248 performance.



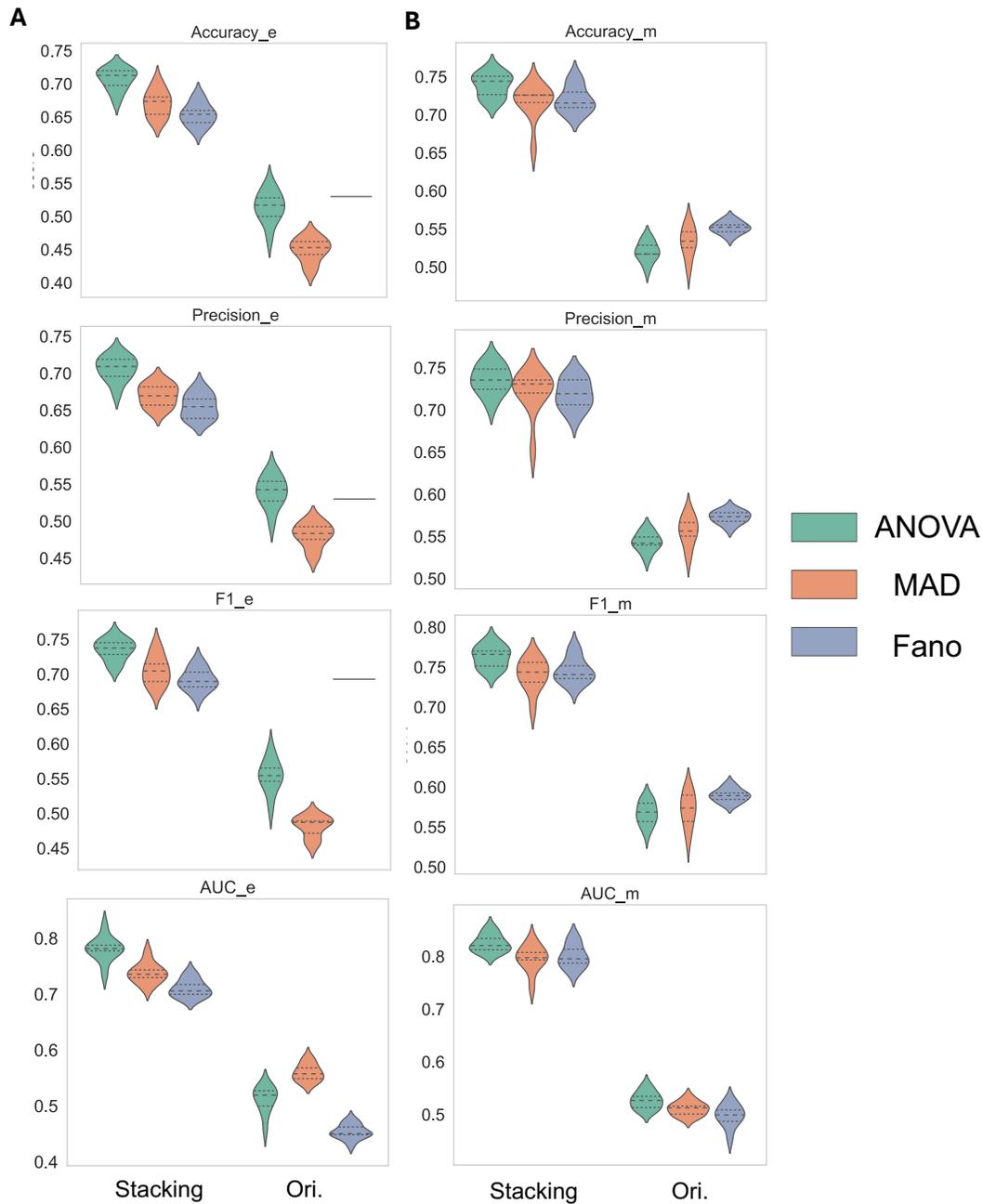
249
 250 **Figure 3. Comparing the Performance using one Classifier Directly on the Collected data for**
 251 **Binary Classification.** All classifiers are trained with the same feature dimensions and 10 times
 252 10-fold CV on CN vs. EMCI group. The performance was measured using the metric introduced
 253 previously. (A-B) Accuracy and AUC score comparison on gene expression data. (C-D) Accuracy

254 *and AUC score comparison on gene methylation data. SVM: Support Vector Machine LR: Logistic*
255 *Regression; MLP: Multilayer Perceptron; RF: Random Forest; NB: Naïve Bayes; KNN: K-*
256 *Nearest Neighbor.*

257

258 **The Stacking Ensemble Learning has Dramatically Improved the Overall Outcome**

259 Classifiers ensemble is due to the premise that ensembles can often achieve better
260 performance than individual classifiers. Except for general voting, stacking is also commonly used,
261 which combines the predictions of base-level classifiers together with the class label to establish
262 the meta-level dataset for decision-making, and is found to outperform voting [1][70]. We applied
263 the stacking technique using a three-layer (one hidden layer) MLP as the meta-model to enhance
264 the five base classifier outputs on single omics classification [71]. **Fig. 4** shows the CN vs. EMCI
265 group results in comparison before (SVM as the only classifier) and after introducing stacking,
266 including gene expression and methylation. Overall, there is about 20% improvement in the
267 performance matrix (Accuracy, Precision, Specificity, AUC) after applying stacking. Among the
268 three feature selection methods, ANOVA F-test selection achieved the highest performance after
269 stacking. We then select the ANOVA F-test for the feature selection block during the integration
270 model establishment.



271

272 **Figure 4. Model Improvement After Stacking.** The results are based on CN vs. EMCI Group. (A)

273 Classification performance improvement using gene expression data only before and after

274 stacking. (B) Classification performance improvement using gene methylation data only before

275 and after stacking. “_e”: gene expression; “_m”: gene methylation; ANOVA: ANOVA F-test for

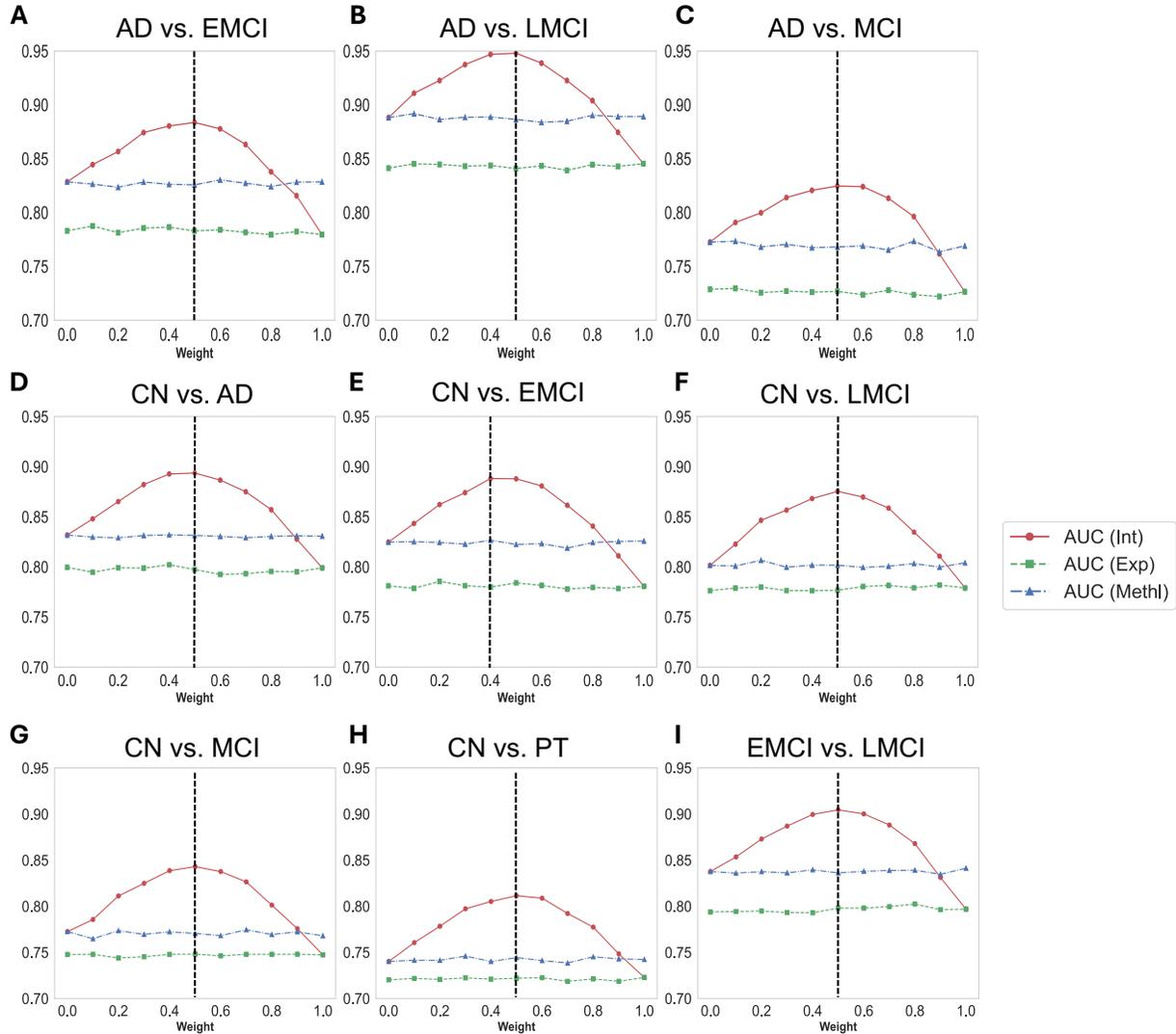
276 *feature selection; MAD: Median Absolute Deviation; Stacking: Results for stacking models; Ori.:*
277 *Results using one classifier (SVM) only.*

278

279 **Integration Model Achieved Higher Performance Than One Modality Only**

280 WIMOAD is a weighted score fusion model for binary group classification, with distinctly
281 assigned weight coefficients to balance the contribution of each data modality when reducing the
282 negative effect that results from the data collection to the minimum. **Fig. 5** show how the
283 coefficient of the Exp meta-model impacts the prediction accuracy of the final output. With
284 optimized weights, the value of feature integration and the potential for original sampling exceed
285 the performance of both Exp and Methl meta-model outputs. According to the AUC comparison,
286 the integration model can outperform both omics when assigning weight from 0.2 to 0.8, when
287 achieving the peaks around 0.5. Only the CN vs. EMCI group archives the peak when the weight
288 for the Exp meta-model is 0.4. For convenience of the test, we assigned the weight coefficient as
289 0.5 for each meta-model for further study.

290 Our constructed WIMOAD integration model demonstrated an improvement in
291 performance relative to single modality models, effectively mitigating the impact of poorly
292 performing data on the final classification results with pre-optimized weight coefficients for both
293 omics. As illustrated in **Fig. 6**, the integration model significantly enhanced the overall
294 performance compared to using one omics only.



295

296 **Figure. 5** Variation in AUC of the Integration Model with Changes in the Integration

297 **Coefficient.** the x-axis represents the increase of the integration coefficient α , which is the weight

298 assigned to the prediction results of Exp classifier. The y-axis represents the accuracy of the model.

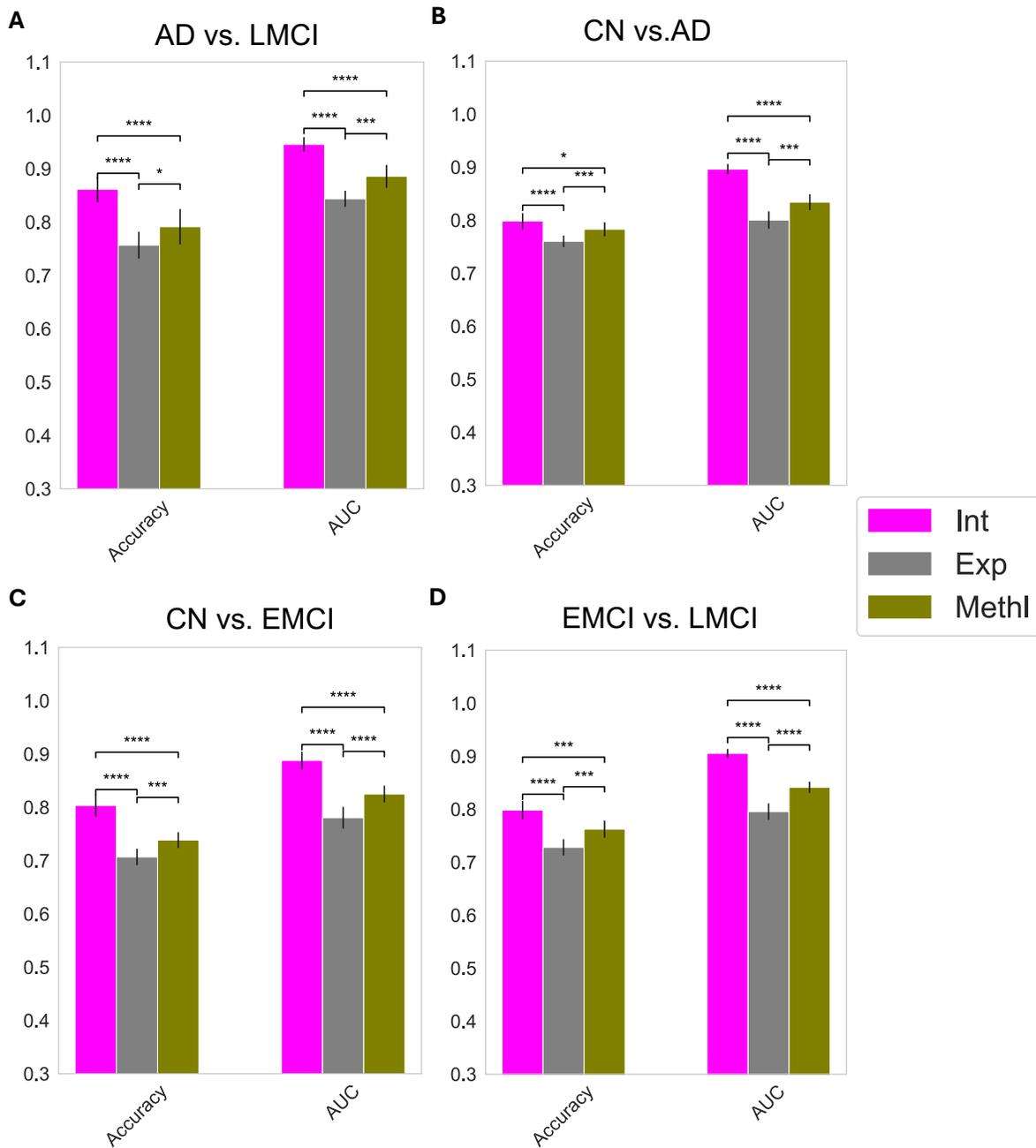
299 The vertical dashed black line represents the highest AUC with respect to the weight coefficient α .

300 In most tasks (8 out of 9), the integration has the best performance when $\alpha = 0.5$. (A) AD vs. EMCI

301 group. (B) AD vs. LMCI group. (C) AD vs. MCI group. (D) CN vs. AD group. (E) CN vs. EMCI

302 group. (F) CN vs. LMCI group. (G) CN vs. MCI group. (H) CN vs. PT group. (I) EMCI vs. LMCI

303 group.



305

306 **Figure. 6 Integration Performances of WIMOAD.** The x-axis represents the evaluation matrix,

307 and the y-axis represents the values. The results were generated under the best coefficient selected

308 ($\alpha = 0.5$) and cross-validated 10 times. (A) AD vs. LMCI group. (B) CN vs. AD group. (C) CN vs.

309 LMCI group. (D) EMCI vs. LMCI group. ‘*’: $p<0.05$; ‘**’: $p<0.01$; ‘***’: $p<0.001$; ‘****’:
 310 $p<0.0001$.

311

312 Comparison with State-of-the-art Predictors

313 **Table. 2** compares the performance of WIMOAD against the state-of-the-art predictors for
 314 AD diagnosis using the paired ADNI data in our case. Across all the nine groups, the WIMOAD
 315 demonstrates consistently higher accuracies (77.6% on average compared to 70.4% using
 316 IntegrationLearner [72] and 45.6% using MOGLAM [73]) and AUCs (86.9% on average
 317 compared to 69.4% with IntegrationLearner and 53.7% with MOGLAM) compared to the existing
 318 integration methods.

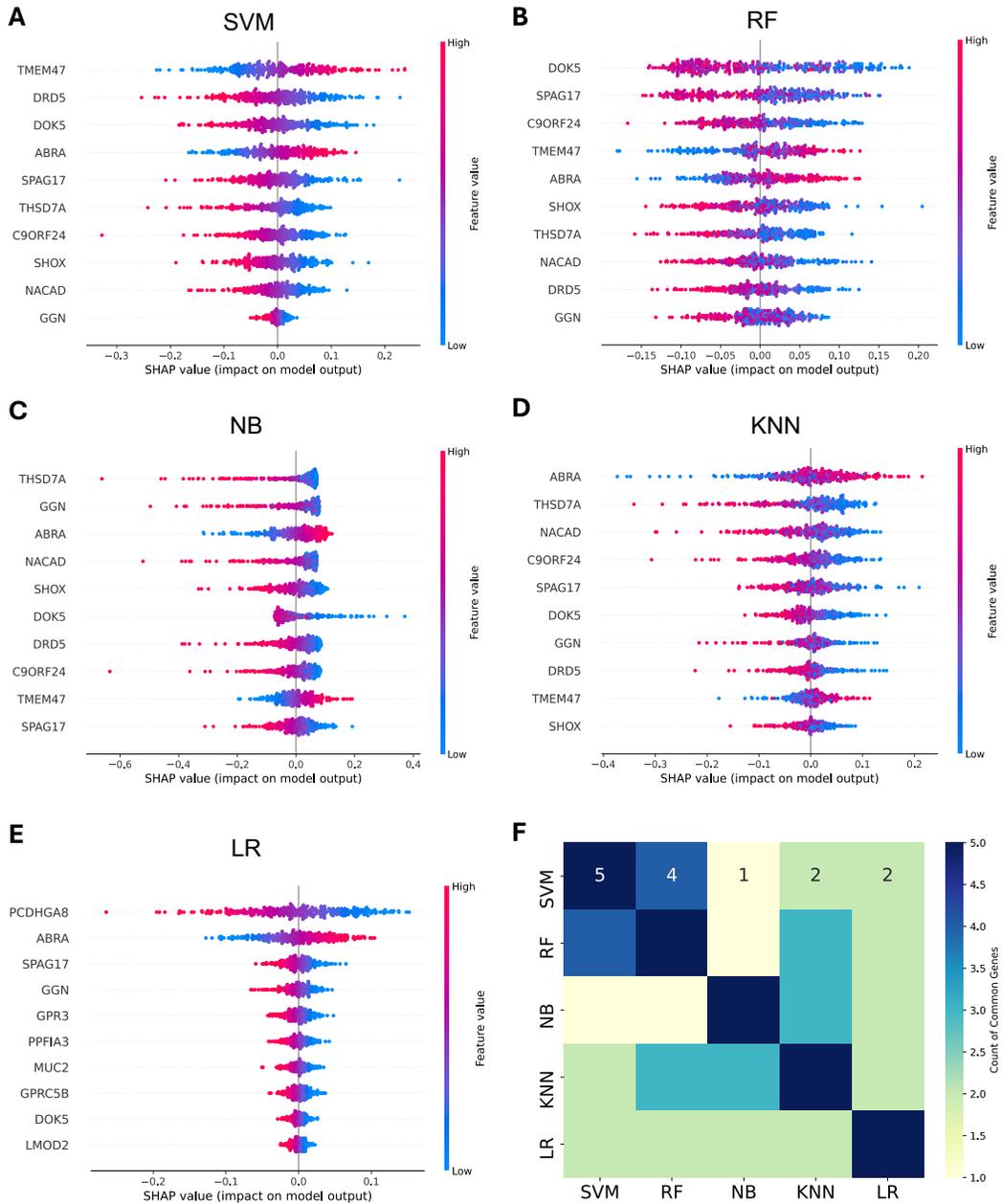
319 **Table. 2 Comparing state-of-the-art methods.** All model apply ADNI data as input source.

Groups	WIMOAD		IntegrationLearner [72]		MOGLAM [73]	
	Acc	AUC	Acc	AUC	Acc	AUC
AD vs. EMCI	0.776	0.882	0.712	0.686	0.333	0.531
AD vs. LMCI	0.862	0.946	0.698	0.743	0.450	0.495
AD vs. MCI	0.776	0.830	0.767	0.660	0.237	0.487
CN vs. AD	0.798	0.896	0.730	0.706	0.310	0.494
CN vs. EMCI	0.803	0.888	0.662	0.706	0.474	0.536
CN vs. LMCI	0.773	0.873	0.715	0.709	0.355	0.673
CN vs. MCI	0.743	0.845	0.671	0.678	0.592	0.574
CN vs. PT	0.709	0.810	0.685	0.671	0.733	0.489
EMCI vs. LMCI	0.740	0.847	0.695	0.685	0.621	0.556
Avg	0.776	0.869	0.704	0.694	0.456	0.537

320

321 **Contributing Genes Identification According to Shapley Values**

322 We leveraged SHAP explainer to enhance the interpretability of our approach by analyzing
323 the importance of each most variable genes selected for model output. As demonstrated in **Fig. 7**
324 **(A-E)**, the gene contributions represented by their respective SHAP values' magnitudes were
325 ranked for gene expression data of group CN vs. EMCI, elucidating the top 10 genes exerting the
326 most substantial influence on model predictions. Remarkably, discernible variations emerged
327 across different binary groups and omics data types. It becomes evident that the regulatory
328 dynamics, manifested through gene upregulation or downregulation, yield bidirectional effects on
329 the model's decision boundaries, influencing the classification outcome for individual samples.
330 After the intersection of top5 contributing genes among five classifiers, *ABRA* (*Actin-binding Rho-*
331 *activating protein*) is the gene present in the overlap. In the SHAP summary plot, if the ABRA
332 gene expression level is high, the model is more likely to predict the sample as EMCI.



333

334 **Fig. 7. SHAP Plots for Model Explanation and Contributing Genes Detection.** Top 10 most
 335 contributing genes and their influence on the model classification (sample being classified as
 336 EMCI) were exhibited. (A-E) SHAP summary plots for gene expression classifier of CN vs. EMCI
 337 group. the colors show the gene expression/methylation level of certain genes, and the SHAP

338 *values of the certain gene for each sample are denoted in the x-axis. Higher SHAP values for a*
339 *certain gene represent the higher possibility that with the expression/methylation value, the model*
340 *will classify the sample as AD. (F) The heatmap showing the overlapping genes of five gene sets*
341 *generated from the top5 contributing genes in each classifier.*

342

343 **Discussion**

344 In this study, we introduce WIMOAD (Weighted Integration for Superior Alzheimer's
345 Diagnosis), a supervised binary classification model that integrates the stacked classification
346 results from gene expression and methylation data through a weighted score fusion approach for
347 early diagnosis of AD. Additionally, the model applied SHAP to interpret the contributions of
348 different omics data and revealed distinct contributing genes across various data sources.
349 According to the 10 times 10-fold CV results, WIMOAD improves the overall performance by
350 integrating two omics in the binary classification task, especially in the classification case between
351 health control and early mild cognitive impairment.

352 WIMOAD is an integration model based on meta-learning. As the convolutional and MLP-
353 based classifiers and algorithms that applied deep learning did not provide better performances
354 with the datasets according to the classifiers comparison, we established meta-models that take the
355 predictions from different classifiers and the test label as a training dataset for model improvement
356 for each omics. By assigning weights for the score generated by each classifier to different omics
357 data profiles, there is a general increase in the model output, which results in one or more peaks
358 that the performance matrix of the model can surpass using single omics in the classification task.

359 After the establishment of the model, we tested other multimodal fusion models, such as
360 IntegrationLearner from Mallick et al. [72], a novel Bayesian ensemble method that combines

361 information across several longitudinal and cross-sectional omics data layers, and MoGLAM from
362 Ouyang et al. [73], which integrates a dynamic graph convolutional network, attention mechanism,
363 and omic integrated representation learning modules for fusing DNA methylation, miRNA, and
364 mRNA expression profiles for disease classification. Comparative analysis revealed that
365 WIMOAD consistently outperformed these methods across all classification groups. A likely
366 reason for WIMOAD's superior results is its use of weighted score fusion to aggregate predictions
367 from different classifiers, followed by decision-making, rather than directly concatenating data
368 from various sources as input for predictions.

369 For the interpretability of the model, WIMOAD applied SHAP for each data modality.
370 Instead of directly combining data, WIMOAD can extract specific representations from different
371 data modalities simultaneously and fully use all the information for the prediction. By quantifying
372 the contributions of the most variable genes separately, WIMOAD will contribute to the detection
373 of new biomarkers in multi-omics for early diagnosis, biomarker discovery, and precision therapy
374 design in AD studies. Given that the SVM model can currently only utilize KernelSHAP—an
375 algorithm within SHAP with relatively high computational complexity and longer runtime—we
376 have limited our presentation to the top ten genes (both expression and methylation) that most
377 significantly influence the model's predictions. Integrating SHAP into the decision-making process
378 allows for the visualization of how gene expression/methylation levels affect model predictions as
379 well. For instance, a higher expression level of a particular gene correlates with a higher
380 corresponding Shapley value, indicating that when the model detects high expression of this gene
381 in a sample, it is more likely to classify the sample into a specific category. This demonstrates that
382 the gene's expression level has a direct impact on the model's final prediction. Consequently,
383 incorporating the SHAP explainer makes it feasible to identify new biomarkers. Additionally, in

384 binary classification cases, the results obtained from different groups could potentially serve as
385 markers for identifying the various stages in the progression from healthy (CN) to MCI (EMCI
386 and LMCI) and AD.

387 The limitations of the WIMOAD model primarily center on the number of modalities it
388 deals with. WIMOAD currently integrates only gene expression and methylation data, whereas
389 most state-of-the-art integration models incorporate three or more data modalities. During the
390 development of WIMOAD, we attempted to include proteomics profiles [74] into consideration.
391 However, only 129 samples met the criteria of having gene expression, methylation, and
392 proteomics data after filtering, and these samples were only sufficient for the CN-LMCI binary
393 classification task. As a result, the model is limited to two types of omics data. Notably, since
394 our data all comes from the peripheral blood, the biomarker detection in the study needs further
395 investigation about how it links with the change in the brain, and how it will contribute to the
396 mechanism of the AD process.

397

398 **Conclusion**

399 In this paper, we proposed a weighted score fusion model named WIMOAD for multi-
400 omics integration in AD diagnosis. It is a meta-learning-based model that extracts information
401 from both gene expression and paired methylation profiles of samples for model decision-making.
402 Compared to the most recent models presented that incorporate statistical analysis and deep
403 learning algorithms, WIMOAD has surpassed most classification tasks with genetic data. By
404 adding the SHAP explainer in the workflow, top contributing genes or biomarkers from different
405 omics and how they affect the model classification results can be visualized. Additionally,
406 WIMOAD is also flexible in the number of data modalities included and straightforward to

407 implement. The future direction of our research will include incorporating commonly utilized
408 imaging data to develop a more comprehensive multi-modality-based diagnostic model that
409 enhances AD diagnostics' robustness and clinical applicability in disease pathology.

410

411 **Conflict of Interest**

412 The authors have declared that no competing interests exist.

413

414 **Funding**

415 Research reported in this publication was supported by the National Cancer Institute of the
416 National Institutes of Health under Award Number P30CA036727. This work was supported by
417 the American Cancer Society under award number IRG-22-146-07-IRG, and by the Buffett Cancer
418 Center, which is supported by the National Cancer Institute under award number CA036727. This
419 work was supported by the Buffet Cancer Center, which is supported by the National Cancer
420 Institute under award number CA036727, in collaboration with the UNMC/Children's Hospital &
421 Medical Center Child Health Research Institute Pediatric Cancer Research Group. This study was
422 supported, in part, by the National Institute on Alcohol Abuse and Alcoholism (P50AA030407-
423 5126, Pilot Core grant). This study was also supported by the Nebraska EPSCoR FIRST Award
424 (OIA-2044049). This work was also partially supported by the National Institute of General
425 Medical Sciences under Award Numbers P20GM103427 and P20GM130447. This study was in
426 part financially supported by the Child Health Research Institute at UNMC/Children's Nebraska.
427 This work was also partially supported by the University of Nebraska Collaboration Initiative
428 Grant from the Nebraska Research Initiative (NRI). The content is solely the responsibility of the
429 authors and does not necessarily represent the official views from the funding organizations.

430

431 **Authors' contributions**

432 SW conceived and designed the study. HX developed the algorithm, performed the experiments,
433 analyzed the data and implemented the WIMOAD package. All authors participated in writing the
434 paper. The manuscript was approved by all authors.

435

436 **Data availability**

437 All the data used in this manuscript are publicly available in the corresponding references.

438 WIMOAD is available at <https://github.com/wan-mlab/WIMOAD>.

- 440 [1] C.R. Jack, D.A. Bennett, K. Blennow, M.C. Carrillo, B. Dunn, S.B. Haeberlein, D.M.
441 Holtzman, W. Jagust, F. Jessen, J. Karlawish, E. Liu, J.L. Molinuevo, T. Montine, C.
442 Phelps, K.P. Rankin, C.C. Rowe, P. Scheltens, E. Siemers, H.M. Snyder, R. Sperling,
443 Contributors, C. Elliott, E. Masliah, L. Ryan, N. Silverberg, NIA-AA Research Framework:
444 Toward a biological definition of Alzheimer's disease, *Alzheimer's & Dementia* 14
445 (2018) 535–562. <https://doi.org/10.1016/j.jalz.2018.02.018>.
- 446 [2] L. Pini, M. Pievani, M. Bocchetta, D. Altomare, P. Bosco, E. Cavedo, S. Galluzzi, M.
447 Marizzoni, G.B. Frisoni, Brain atrophy in Alzheimer's Disease and aging, *Ageing Research*
448 *Reviews* 30 (2016) 25–48. <https://doi.org/10.1016/j.arr.2016.01.002>.
- 449 [3] *Alzheimers_report.pdf*, (n.d.).
- 450 [4] F.E. Matthews, B.C.M. Stephan, L. Robinson, C. Jagger, L.E. Barnes, A. Arthur, C.
451 Brayne, Cognitive Function and Ageing Studies (CFAS) Collaboration, A two decade
452 dementia incidence comparison from the Cognitive Function and Ageing Studies I and II,
453 *Nat Commun* 7 (2016) 11398. <https://doi.org/10.1038/ncomms11398>.
- 454 [5] J. Rasmussen, H. Langerman, Alzheimer's Disease – Why We Need Early Diagnosis,
455 *DNND Volume 9* (2019) 123–130. <https://doi.org/10.2147/DNND.S228939>.
- 456 [6] C.R. Jack, D.A. Bennett, K. Blennow, M.C. Carrillo, H.H. Feldman, G.B. Frisoni, H.
457 Hampel, W.J. Jagust, K.A. Johnson, D.S. Knopman, R.C. Petersen, P. Scheltens, R.A.
458 Sperling, B. Dubois, A/T/N: An unbiased descriptive classification scheme for Alzheimer
459 disease biomarkers, *Neurology* 87 (2016) 539–547.
460 <https://doi.org/10.1212/WNL.0000000000002923>.
- 461 [7] S. Sadigh-Eteghad, B. Sabermarouf, A. Majdi, M. Talebi, M. Farhoudi, J. Mahmoudi,
462 Amyloid-Beta: A Crucial Factor in Alzheimer's Disease, *Medical Principles and Practice* 24
463 (2014) 1–10. <https://doi.org/10.1159/000369101>.
- 464 [8] S.S. Sisodia, D.L. Price, Role of the β -amyloid protein in Alzheimer's disease, *The FASEB*
465 *Journal* 9 (1995) 366–370. <https://doi.org/10.1096/fasebj.9.5.7896005>.
- 466 [9] T. Grimmer, M. Riemenschneider, H. Förstl, G. Henriksen, W.E. Klunk, C.A. Mathis, T.
467 Shiga, H.-J. Wester, A. Kurz, A. Drzezga, Beta Amyloid in Alzheimer's Disease: Increased
468 Deposition in Brain Is Reflected in Reduced Concentration in Cerebrospinal Fluid,
469 *Biological Psychiatry* 65 (2009) 927–934. <https://doi.org/10.1016/j.biopsych.2009.01.027>.
- 470 [10] Y. Wang, E. Mandelkow, Tau in physiology and pathology, *Nat Rev Neurosci* 17 (2016)
471 22–35. <https://doi.org/10.1038/nrn.2015.1>.
- 472 [11] S. Muralidar, S.V. Ambi, S. Sekaran, D. Thirumalai, B. Palaniappan, Role of tau protein in
473 Alzheimer's disease: The prime pathological player, *International Journal of Biological*
474 *Macromolecules* 163 (2020) 1599–1617. <https://doi.org/10.1016/j.ijbiomac.2020.07.327>.
- 475 [12] L. Gjerum, B.B. Andersen, M. Bruun, A.H. Simonsen, O.M. Henriksen, I. Law, S.G.
476 Hasselbalch, K.S. Frederiksen, Comparison of the clinical impact of 2-[18F]FDG-PET and
477 cerebrospinal fluid biomarkers in patients suspected of Alzheimer's disease, *PLoS One* 16
478 (2021) e0248413. <https://doi.org/10.1371/journal.pone.0248413>.
- 479 [13] I. Bazarbekov, A. Razaque, M. Ipalakova, J. Yoo, Z. Assipova, A. Almisreb, A review of
480 artificial intelligence methods for Alzheimer's disease diagnosis: Insights from
481 neuroimaging to sensor data analysis, *Biomedical Signal Processing and Control* 92
482 (2024) 106023. <https://doi.org/10.1016/j.bspc.2024.106023>.
- 483 [14] Y. Tong, Z. Li, H. Huang, L. Gao, M. Xu, Z. Hu, Research of spatial context convolutional
484 neural networks for early diagnosis of Alzheimer's disease, *J Supercomput* 80 (2024)
485 5279–5297. <https://doi.org/10.1007/s11227-023-05655-9>.
- 486 [15] S.K. Arjaria, A.S. Rathore, D. Bisen, S. Bhattacharyya, Performances of Machine Learning
487 Models for Diagnosis of Alzheimer's Disease, *Ann. Data. Sci.* 11 (2024) 307–335.
488 <https://doi.org/10.1007/s40745-022-00452-2>.

- 489 [16] K. Chen, Y. Weng, A.A. Hosseini, T. Dening, G. Zuo, Y. Zhang, A comparative study of
490 GNN and MLP based machine learning for the diagnosis of Alzheimer's Disease involving
491 data synthesis, *Neural Netw* 169 (2024) 442–452.
492 <https://doi.org/10.1016/j.neunet.2023.10.040>.
- 493 [17] S. Al-Otaibi, M. Mujahid, A.R. Khan, H. Nobanee, J. Alyami, T. Saba, Dual Attention
494 Convolutional AutoEncoder for Diagnosis of Alzheimer's Disorder in Patients Using
495 Neuroimaging and MRI Features, *IEEE Access* 12 (2024) 58722–58739.
496 <https://doi.org/10.1109/ACCESS.2024.3390186>.
- 497 [18] S. Miao, Q. Xu, W. Li, C. Yang, B. Sheng, F. Liu, T.T. Bezabih, X. Yu, MMTFN: Multi-
498 modal multi-scale transformer fusion network for Alzheimer's disease diagnosis,
499 *International Journal of Imaging Systems and Technology* 34 (2024) e22970.
500 <https://doi.org/10.1002/ima.22970>.
- 501 [19] S. Suganyadevi, A.S. Pershiya, K. Balasamy, V. Seethalakshmi, S. Bala, K. Arora, Deep
502 Learning Based Alzheimer Disease Diagnosis: A Comprehensive Review, *SN COMPUT.*
503 *SCI.* 5 (2024) 391. <https://doi.org/10.1007/s42979-024-02743-2>.
- 504 [20] C.R. Jack, D.M. Holtzman, Biomarker Modeling of Alzheimer's Disease, *Neuron* 80 (2013)
505 1347–1358. <https://doi.org/10.1016/j.neuron.2013.12.003>.
- 506 [21] Bateman Randall J., Xiong Chengjie, Benzinger Tammie L.S., Fagan Anne M., Goate
507 Alison, Fox Nick C., Marcus Daniel S., Cairns Nigel J., Xie Xianyun, Blazey Tyler M.,
508 Holtzman David M., Santacruz Anna, Buckles Virginia, Oliver Angela, Moulder Krista,
509 Aisen Paul S., Ghetti Bernardino, Klunk William E., McDade Eric, Martins Ralph N.,
510 Masters Colin L., Mayeux Richard, Ringman John M., Rossor Martin N., Schofield Peter
511 R., Sperling Reisa A., Salloway Stephen, Morris John C., Clinical and Biomarker Changes
512 in Dominantly Inherited Alzheimer's Disease, *New England Journal of Medicine* 367 (2012)
513 795–804. <https://doi.org/10.1056/NEJMoa1202753>.
- 514 [22] R.P. Friedland, E. Koss, J.V. Haxby, C.L. Grady, J. Luxenberg, M.B. Schapiro, J. Kaye,
515 Alzheimer Disease: Clinical and Biological Heterogeneity, *Ann Intern Med* 109 (1988) 298–
516 311. <https://doi.org/10.7326/0003-4819-109-4-298>.
- 517 [23] D. Ferreira, A. Nordberg, E. Westman, Biological subtypes of Alzheimer disease,
518 *Neurology* 94 (2020) 436–448. <https://doi.org/10.1212/WNL.0000000000009058>.
- 519 [24] J.M. Ringman, A. Goate, C.L. Masters, N.J. Cairns, A. Danek, N. Graff-Radford, B. Ghetti,
520 J.C. Morris, Dominantly Inherited Alzheimer Network, Genetic Heterogeneity in Alzheimer
521 Disease and Implications for Treatment Strategies, *Curr Neurol Neurosci Rep* 14 (2014)
522 499. <https://doi.org/10.1007/s11910-014-0499-8>.
- 523 [25] H. Hampel, S. Lista, C. Neri, A. Vergallo, Time for the systems-level integration of aging:
524 Resilience enhancing strategies to prevent Alzheimer's disease, *Progress in Neurobiology*
525 181 (2019) 101662. <https://doi.org/10.1016/j.pneurobio.2019.101662>.
- 526 [26] Utility of the Apolipoprotein E Genotype in the Diagnosis of Alzheimer's Disease | *New*
527 *England Journal of Medicine*, (n.d.).
528 <https://www.nejm.org/doi/full/10.1056/nejm199802193380804> (accessed May 28, 2024).
- 529 [27] J.-C. Lambert, S. Heath, G. Even, D. Campion, K. Sleegers, M. Hiltunen, O. Combarros, D.
530 Zelenika, M.J. Bullido, B. Tavernier, L. Letenneur, K. Bettens, C. Berr, F. Pasquier, N.
531 Fiévet, P. Barberger-Gateau, S. Engelborghs, P. De Deyn, I. Mateo, A. Franck, S.
532 Helisalmi, E. Porcellini, O. Hanon, European Alzheimer's Disease Initiative Investigators,
533 M.M. de Pancorbo, C. Lendon, C. Dufouil, C. Jaillard, T. Leveillard, V. Alvarez, P. Bosco,
534 M. Mancuso, F. Panza, B. Nacmias, P. Bossù, P. Piccardi, G. Annoni, D. Seripa, D.
535 Galimberti, D. Hannequin, F. Licastro, H. Soininen, K. Ritchie, H. Blanché, J.-F. Dartigues,
536 C. Tzourio, I. Gut, C. Van Broeckhoven, A. Alperovitch, M. Lathrop, P. Amouyel, Genome-
537 wide association study identifies variants at CLU and CR1 associated with Alzheimer's
538 disease, *Nat Genet* 41 (2009) 1094–1099. <https://doi.org/10.1038/ng.439>.

- 539 [28] P. Hollingworth, D. Harold, R. Sims, A. Gerrish, J.-C. Lambert, M.M. Carrasquillo, R.
540 Abraham, M.L. Hamshere, J.S. Pahwa, V. Moskvina, K. Dowzell, N. Jones, A. Stretton, C.
541 Thomas, A. Richards, D. Ivanov, C. Widdowson, J. Chapman, S. Lovestone, J. Powell, P.
542 Proitsi, M.K. Lupton, C. Brayne, D.C. Rubinsztein, M. Gill, B. Lawlor, A. Lynch, K.S. Brown,
543 P.A. Passmore, D. Craig, B. McGuinness, S. Todd, C. Holmes, D. Mann, A.D. Smith, H.
544 Beaumont, D. Warden, G. Wilcock, S. Love, P.G. Kehoe, N.M. Hooper, E.R.L.C. Vardy, J.
545 Hardy, S. Mead, N.C. Fox, M. Rossor, J. Collinge, W. Maier, F. Jessen, E. R  ther, B.
546 Sch  rmann, R. Heun, H. K  lsch, H. van den Bussche, I. Heuser, J. Kornhuber, J. Wiltfang,
547 M. Dichgans, L. Fr  lich, H. Hampel, J. Gallacher, M. H  ll, D. Rujescu, I. Giegling, A.M.
548 Goate, J.S.K. Kauwe, C. Cruchaga, P. Nowotny, J.C. Morris, K. Mayo, K. Sleegers, K.
549 Bettens, S. Engelborghs, P.P. De Deyn, C. Van Broeckhoven, G. Livingston, N.J. Bass, H.
550 Gurling, A. McQuillin, R. Gwilliam, P. Deloukas, A. Al-Chalabi, C.E. Shaw, M. Tsolaki, A.B.
551 Singleton, R. Guerreiro, T.W. M  hleisen, M.M. N  then, S. Moebus, K.-H. J  ckel, N. Klopp,
552 H.-E. Wichmann, V.S. Pankratz, S.B. Sando, J.O. Aasly, M. Barcikowska, Z.K. Wszolek,
553 D.W. Dickson, N.R. Graff-Radford, R.C. Petersen, Alzheimer’s Disease Neuroimaging
554 Initiative, C.M. van Duijn, M.M.B. Breteler, M.A. Ikram, A.L. DeStefano, A.L. Fitzpatrick, O.
555 Lopez, L.J. Launer, S. Seshadri, CHARGE consortium, C. Berr, D. Campion, J. Epelbaum,
556 J.-F. Dartigues, C. Tzourio, A. Alperovitch, M. Lathrop, EADI1 consortium, T.M. Feulner, P.
557 Friedrich, C. Riehle, M. Krawczak, S. Schreiber, M. Mayhaus, S. Nicolhaus, S. Wagenpfeil,
558 S. Steinberg, H. Stefansson, K. Stefansson, J. Snaedal, S. Bj  rnsson, P.V. Jonsson, V.
559 Chouraki, B. Genier-Boley, M. Hiltunen, H. Soininen, O. Combarros, D. Zelenika, M.
560 Delphine, M.J. Bullido, F. Pasquier, I. Mateo, A. Frank-Garcia, E. Porcellini, O. Hanon, E.
561 Coto, V. Alvarez, P. Bosco, G. Siciliano, M. Mancuso, F. Panza, V. Solfrizzi, B. Nacmias,
562 S. Sorbi, P. Boss  , P. Piccardi, B. Arosio, G. Annoni, D. Seripa, A. Pilotto, E. Scarpini, D.
563 Galimberti, A. Brice, D. Hannequin, F. Licastro, L. Jones, P.A. Holmans, T. Jonsson, M.
564 Riemenschneider, K. Morgan, S.G. Younkin, M.J. Owen, M. O’Donovan, P. Amouyel, J.
565 Williams, Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are
566 associated with Alzheimer’s disease, *Nat Genet* 43 (2011) 429–435.
567 <https://doi.org/10.1038/ng.803>.
- 568 [29] A. Forte, S. Lara, C. Pe  a-Bautista, M. Baquero, C. Ch  fer-Peric  s, New approach for
569 early and specific Alzheimer disease diagnosis from different plasma biomarkers, *Clinica*
570 *Chimica Acta* 556 (2024) 117842. <https://doi.org/10.1016/j.cca.2024.117842>.
- 571 [30] M.I. Kamboh, F.Y. Demirci, X. Wang, R.L. Minster, M.M. Carrasquillo, V.S. Pankratz, S.G.
572 Younkin, A.J. Saykin, Alzheimer’s Disease Neuroimaging Initiative, G. Jun, C. Baldwin,
573 M.W. Logue, J. Buros, L. Farrer, M.A. Pericak-Vance, J.L. Haines, R.A. Sweet, M. Ganguli,
574 E. Feingold, S.T. Dekosky, O.L. Lopez, M.M. Barmada, Genome-wide association study of
575 Alzheimer’s disease, *Transl Psychiatry* 2 (2012) e117. <https://doi.org/10.1038/tp.2012.45>.
- 576 [31] S. Guo, J. Yang, Bayesian genome-wide TWAS with reference transcriptomic data of brain
577 and blood tissues identified 141 risk genes for Alzheimer’s disease dementia, *Alz Res*
578 *Therapy* 16 (2024) 120. <https://doi.org/10.1186/s13195-024-01488-7>.
- 579 [32] H. Mathys, J. Davila-Velderrain, Z. Peng, F. Gao, S. Mohammadi, J.Z. Young, M. Menon,
580 L. He, F. Abdurrob, X. Jiang, A.J. Martorell, R.M. Ransohoff, B.P. Hafler, D.A. Bennett, M.
581 Kellis, L.-H. Tsai, Single-cell transcriptomic analysis of Alzheimer’s disease, *Nature* 570
582 (2019) 332–337. <https://doi.org/10.1038/s41586-019-1195-2>.
- 583 [33] S. Horvath, K. Raj, DNA methylation-based biomarkers and the epigenetic clock theory of
584 ageing, *Nat Rev Genet* 19 (2018) 371–384. <https://doi.org/10.1038/s41576-018-0004-3>.
- 585 [34] A. Di Francesco, B. Arosio, A. Falconi, M.V. Micioni Di Bonaventura, M. Karimi, D. Mari, M.
586 Casati, M. Maccarrone, C. D’Addario, Global changes in DNA methylation in Alzheimer’s
587 disease peripheral blood mononuclear cells, *Brain, Behavior, and Immunity* 45 (2015) 139–
588 144. <https://doi.org/10.1016/j.bbi.2014.11.002>.

- 589 [35] X. Wei, L. Zhang, Y. Zeng, DNA methylation in Alzheimer's disease: In brain and
590 peripheral blood, *Mechanisms of Ageing and Development* 191 (2020) 111319.
591 <https://doi.org/10.1016/j.mad.2020.111319>.
- 592 [36] S. Qiu, M. Sun, Y. Xu, Y. Hu, Integrating multi-omics data to reveal the effect of genetic
593 variant rs6430538 on Alzheimer's disease risk, *Front. Neurosci.* 18 (2024).
594 <https://doi.org/10.3389/fnins.2024.1277187>.
- 595 [37] D. Das, J. Ito, T. Kadowaki, K. Tsuda, An interpretable machine learning model for
596 diagnosis of Alzheimer's disease, *PeerJ* 7 (2019) e6543.
597 <https://doi.org/10.7717/peerj.6543>.
- 598 [38] J. Zhang, X. Sun, X. Jia, B. Sun, S. Xu, W. Zhang, Z. Liu, Integrative multi-omics analysis
599 reveals the critical role of the PBXIP1 gene in Alzheimer's disease, *Aging Cell* 23 (2024)
600 e14044. <https://doi.org/10.1111/accel.14044>.
- 601 [39] T. Oka, Y. Matsuzawa, M. Tsuneyoshi, Y. Nakamura, K. Aoshima, H. Tsugawa, Multiomics
602 analysis to explore blood metabolite biomarkers in an Alzheimer's Disease Neuroimaging
603 Initiative cohort, *Sci Rep* 14 (2024) 6797. <https://doi.org/10.1038/s41598-024-56837-1>.
- 604 [40] A. Badhwar, G.P. McFall, S. Sapkota, S.E. Black, H. Chertkow, S. Duchesne, M. Masellis,
605 L. Li, R.A. Dixon, P. Bellec, A multiomics approach to heterogeneity in Alzheimer's
606 disease: focused review and roadmap, *Brain* 143 (2020) 1315–1331.
607 <https://doi.org/10.1093/brain/awz384>.
- 608 [41] R. Nativio, Y. Lan, G. Donahue, S. Sidoli, A. Berson, A.R. Srinivasan, O. Shcherbakova, A.
609 Amlie-Wolf, J. Nie, X. Cui, C. He, L.-S. Wang, B.A. Garcia, J.Q. Trojanowski, N.M. Bonini,
610 S.L. Berger, An integrated multi-omics approach identifies epigenetic alterations
611 associated with Alzheimer's disease, *Nat Genet* 52 (2020) 1024–1035.
612 <https://doi.org/10.1038/s41588-020-0696-0>.
- 613 [42] P. Kodam, R. Sai Swaroop, S.S. Pradhan, V. Sivaramakrishnan, R. Vadrevu, Integrated
614 multi-omics analysis of Alzheimer's disease shows molecular signatures associated with
615 disease progression and potential therapeutic targets, *Sci Rep* 13 (2023) 3695.
616 <https://doi.org/10.1038/s41598-023-30892-6>.
- 617 [43] D. Wang, J. Gu, Integrative clustering methods of multi-omics data for molecule-based
618 cancer classifications, *Quant Biol* 4 (2016) 58–67. <https://doi.org/10.1007/s40484-016-0063-4>.
- 620 [44] R. Shen, A.B. Olshen, M. Ladanyi, Integrative clustering of multiple genomic data types
621 using a joint latent variable model with application to breast and lung cancer subtype
622 analysis, *Bioinformatics* 25 (2009) 2906–2912.
623 <https://doi.org/10.1093/bioinformatics/btp543>.
- 624 [45] E.F. Lock, K.A. Hoadley, J.S. Marron, A.B. Nobel, Joint and individual variation explained
625 (JIVE) for integrated analysis of multiple data types, *Ann. Appl. Stat.* 7 (2013).
626 <https://doi.org/10.1214/12-AOAS597>.
- 627 [46] I. Gaynanova, G. Li, Structural learning and integrative decomposition of multi-view data,
628 *Biometrics* 75 (2019) 1121–1132. <https://doi.org/10.1111/biom.13108>.
- 629 [47] J. Bao, C. Chang, Q. Zhang, A.J. Saykin, L. Shen, Q. Long, for the Alzheimer's Disease
630 Neuroimaging Initiative, Integrative analysis of multi-omics and imaging data with
631 incorporation of biological information via structural Bayesian factor analysis, *Briefings in*
632 *Bioinformatics* 24 (2023) bbad073. <https://doi.org/10.1093/bib/bbad073>.
- 633 [48] N. Mahendran, D.R. Vincent P M, Deep belief network-based approach for detecting
634 Alzheimer's disease using the multi-omics data, *Computational and Structural*
635 *Biotechnology Journal* 21 (2023) 1651–1660. <https://doi.org/10.1016/j.csbj.2023.02.021>.
- 636 [49] S.R. Stahlschmidt, B. Ulfenborg, J. Synnergren, Multimodal deep learning for biomedical
637 data fusion: a review, *Briefings in Bioinformatics* 23 (2022) bbab569.
638 <https://doi.org/10.1093/bib/bbab569>.

- 639 [50] P. Lu, L. Hu, A. Mitelpunkt, S. Bhatnagar, L. Lu, H. Liang, A hierarchical attention-based
640 multimodal fusion framework for predicting the progression of Alzheimer's disease,
641 Biomedical Signal Processing and Control 88 (2024) 105669.
642 <https://doi.org/10.1016/j.bspc.2023.105669>.
- 643 [51] M. Trinh, R. Shahbaba, C. Stark, Y. Ren, Alzheimer's disease detection using data fusion
644 with a deep supervised encoder, Front. Dement. 3 (2024).
645 <https://doi.org/10.3389/frdem.2024.1332928>.
- 646 [52] J.L. Ávila-Jiménez, V. Cantón-Habas, M. del P. Carrera-González, M. Rich-Ruiz, S.
647 Ventura, A deep learning model for Alzheimer's disease diagnosis based on patient clinical
648 records, Computers in Biology and Medicine 169 (2024) 107814.
649 <https://doi.org/10.1016/j.compbimed.2023.107814>.
- 650 [53] V. Bollati, D. Galimberti, L. Pergoli, E. Dalla Valle, F. Barretta, F. Cortini, E. Scarpini, P.A.
651 Bertazzi, A. Baccarelli, DNA methylation in repetitive elements and Alzheimer disease,
652 Brain, Behavior, and Immunity 25 (2011) 1078–1083.
653 <https://doi.org/10.1016/j.bbi.2011.01.017>.
- 654 [54] M. Barrachina, I. Ferrer, DNA Methylation of Alzheimer Disease and Tauopathy-Related
655 Genes in Postmortem Brain, Journal of Neuropathology & Experimental Neurology 68
656 (2009) 880–891. <https://doi.org/10.1097/NEN.0b013e3181af2e46>.
- 657 [55] S. Scarpa, R.A. Cavallaro, F. D'Anselmi, A. Fuso, Gene silencing through methylation: An
658 epigenetic intervention on Alzheimer disease, Journal of Alzheimer's Disease 9 (2006)
659 407–414. <https://doi.org/10.3233/JAD-2006-9406>.
- 660 [56] M. Pammi, N. Aghaeepour, J. Neu, Multiomics, artificial intelligence, and precision
661 medicine in perinatology, Pediatr Res 93 (2023) 308–315. <https://doi.org/10.1038/s41390-022-02181-x>.
- 662 [57] S.M. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, (n.d.).
- 663 [58] B. Pavlyshenko, Using Stacking Approaches for Machine Learning Models, in: 2018 IEEE
664 Second International Conference on Data Stream Mining & Processing (DSMP), 2018: pp.
665 255–258. <https://doi.org/10.1109/DSMP.2018.8478522>.
- 666 [59] C. Cortes, V. Vapnik, Support-vector networks, Mach Learn 20 (1995) 273–297.
667 <https://doi.org/10.1007/BF00994018>.
- 668 [60] T.K. Ho, Random decision forests, in: Proceedings of 3rd International Conference on
669 Document Analysis and Recognition, 1995: pp. 278–282 vol.1.
670 <https://doi.org/10.1109/ICDAR.1995.598994>.
- 671 [61] G.I. Webb, Naïve Bayes, in: C. Sammut, G.I. Webb (Eds.), Encyclopedia of Machine
672 Learning, Springer US, Boston, MA, 2010: pp. 713–714. https://doi.org/10.1007/978-0-387-30164-8_576.
- 673 [62] A. Das, Logistic Regression, in: A.C. Michalos (Ed.), Encyclopedia of Quality of Life and
674 Well-Being Research, Springer Netherlands, Dordrecht, 2014: pp. 3680–3682.
675 https://doi.org/10.1007/978-94-007-0753-5_1689.
- 676 [63] G. Guo, H. Wang, D. Bell, Y. Bi, K. Greer, KNN Model-Based Approach in Classification,
677 in: R. Meersman, Z. Tari, D.C. Schmidt (Eds.), On The Move to Meaningful Internet
678 Systems 2003: CoopIS, DOA, and ODBASE, Springer, Berlin, Heidelberg, 2003: pp. 986–
679 996. https://doi.org/10.1007/978-3-540-39964-3_62.
- 680 [64] M.-C. Popescu, V.E. Balas, L. Perescu-Popescu, N. Mastorakis, Multilayer perceptron and
681 neural networks, WSEAS Trans. Cir. and Sys. 8 (2009) 579–588.
- 682 [65] F. Murtagh, Multilayer perceptrons for classification and regression, Neurocomputing 2
683 (1991) 183–197. [https://doi.org/10.1016/0925-2312\(91\)90023-5](https://doi.org/10.1016/0925-2312(91)90023-5).
- 684 [66] N. Omer Fadl Elssied, O. Ibrahim, A. Hamza Osman, A Novel Feature Selection Based on
685 One-Way ANOVA F-Test for E-Mail Spam Classification, RJASET 7 (2014) 625–638.
686 <https://doi.org/10.19026/rjaset.7.299>.
- 687
688

- 689 [67] D. Grün, L. Kester, A. van Oudenaarden, Validation of noise models for single-cell
690 transcriptomics, *Nat Methods* 11 (2014) 637–640. <https://doi.org/10.1038/nmeth.2930>.
- 691 [68] P. Refaeilzadeh, L. Tang, H. Liu, Cross-Validation, in: L. LIU, M.T. ÖZSU (Eds.),
692 *Encyclopedia of Database Systems*, Springer US, Boston, MA, 2009: pp. 532–538.
693 https://doi.org/10.1007/978-0-387-39940-9_565.
- 694 [69] S.M. Lundberg, B. Nair, M.S. Vavilala, M. Horibe, M.J. Eisses, T. Adams, D.E. Liston, D.K.-
695 W. Low, S.-F. Newman, J. Kim, S.-I. Lee, Explainable machine-learning predictions for the
696 prevention of hypoxaemia during surgery, *Nat Biomed Eng* 2 (2018) 749–760.
697 <https://doi.org/10.1038/s41551-018-0304-0>.
- 698 [70] L. Todorovski, S. Džeroski, Combining Classifiers with Meta Decision Trees, *Machine*
699 *Learning* 50 (2003) 223–249. <https://doi.org/10.1023/A:1021709817809>.
- 700 [71] V. Boateng, B. Yang, Ensemble Stacking with the Multi-Layer Perceptron Neural Network
701 Meta-Learner for Passenger Train Delay Prediction, in: *2023 IEEE Conference on Artificial*
702 *Intelligence (CAI)*, 2023: pp. 21–22. <https://doi.org/10.1109/CAI54212.2023.00017>.
- 703 [72] H. Mallick, A. Porwal, S. Saha, P. Basak, V. Svetnik, E. Paul, An integrated Bayesian
704 framework for multi-omics prediction and classification, *Statistics in Medicine* 43 (2024)
705 983–1002. <https://doi.org/10.1002/sim.9953>.
- 706 [73] D. Ouyang, Y. Liang, L. Li, N. Ai, S. Lu, M. Yu, X. Liu, S. Xie, Integration of multi-omics
707 data using adaptive graph learning and attention mechanism for patient classification and
708 biomarker identification, *Computers in Biology and Medicine* 164 (2023) 107303.
709 <https://doi.org/10.1016/j.combiomed.2023.107303>.
- 710 [74] D.A. Butterfield, D. Boyd-Kimball, A. Castegna, Proteomics in Alzheimer's disease: insights
711 into potential mechanisms of neurodegeneration, *J Neurochem* 86 (2003) 1313–1327.
712 <https://doi.org/10.1046/j.1471-4159.2003.01948.x>.
- 713

Table. 1 The demographic information of the Selected Participants. Data are mean \pm standard deviation (std). CN: Normal Controls; EMCI: Early Mild Cognitive Impairments; LMCI: Late Mild Cognitive Impairments; MCI: Mild Cognitive Impairments; AD: Alzheimer’s Diseases; F:

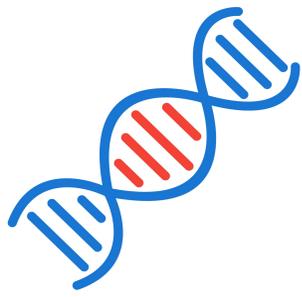
Diagnosis	Samples	Age (mean \pm std)	Sex (F/M)
CN	N = 203	74.45 \pm 5.78	101/102
EMCI	N = 180	71.44 \pm 7.11	81/ 99
LMCI	N = 113	72.74 \pm 7.67	45/68
AD	N = 95	74.28 \pm 7.59	35/60

Female; M: Male

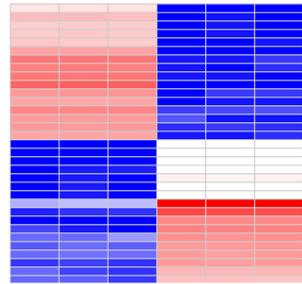
Table. 2 Comparing state-of-the-art methods. All model apply ADNI data as input source.

Groups	WIMOAD		IntegrationLearner [72]		MOGLAM [73]	
	Acc	AUC	Acc	AUC	Acc	AUC
AD vs. EMCI	0.776	0.882	0.712	0.686	0.333	0.531
AD vs. LMCI	0.862	0.946	0.698	0.743	0.450	0.495
AD vs. MCI	0.776	0.830	0.767	0.660	0.237	0.487
CN vs. AD	0.798	0.896	0.730	0.706	0.310	0.494
CN vs. EMCI	0.803	0.888	0.662	0.706	0.474	0.536
CN vs. LMCI	0.773	0.873	0.715	0.709	0.355	0.673
CN vs. MCI	0.743	0.845	0.671	0.678	0.592	0.574
CN vs. PT	0.709	0.810	0.685	0.671	0.733	0.489
EMCI vs. LMCI	0.740	0.847	0.695	0.685	0.621	0.556

Avg	0.776	0.869	0.704	0.694	0.456	0.537
-----	--------------	--------------	-------	-------	-------	-------



Gene Expression



Most Variable Genes



New Training Set1



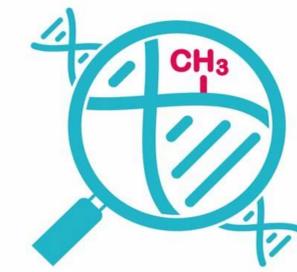
Meta Model_{exp}

α

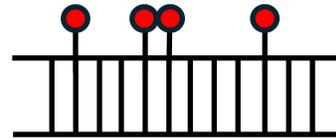


β

Meta Model_{methl}



Methylation



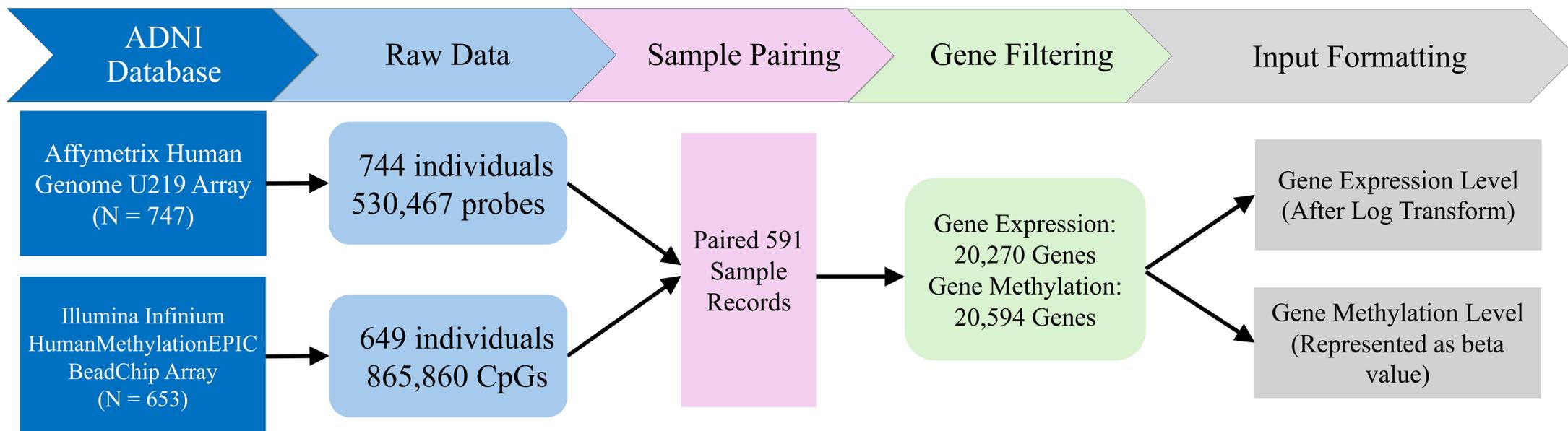
Most Variable Genes

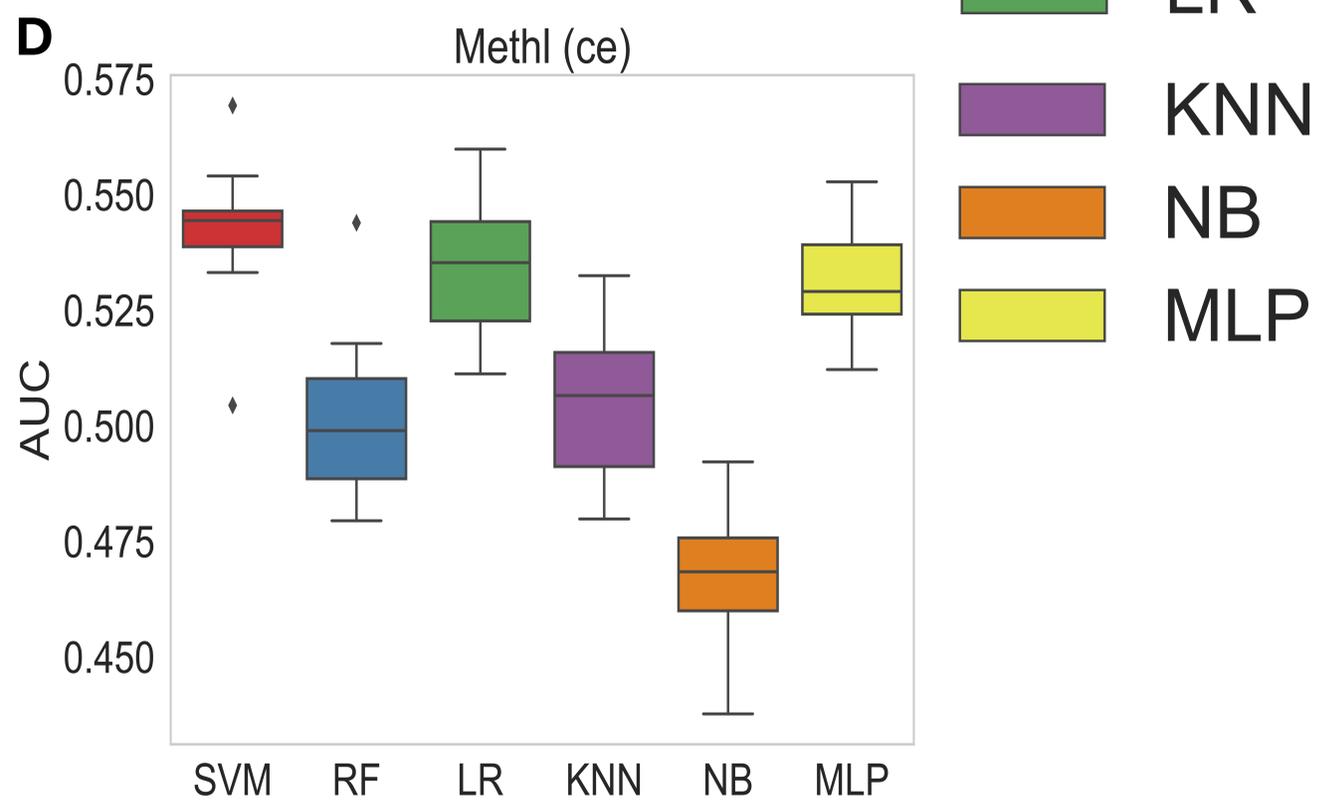
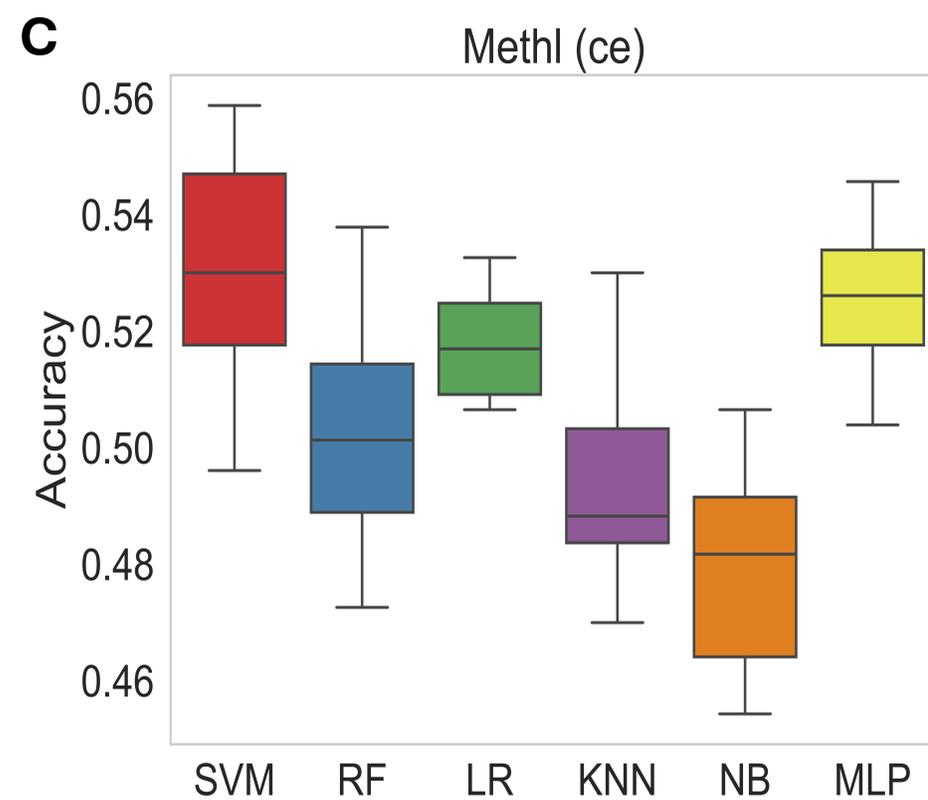
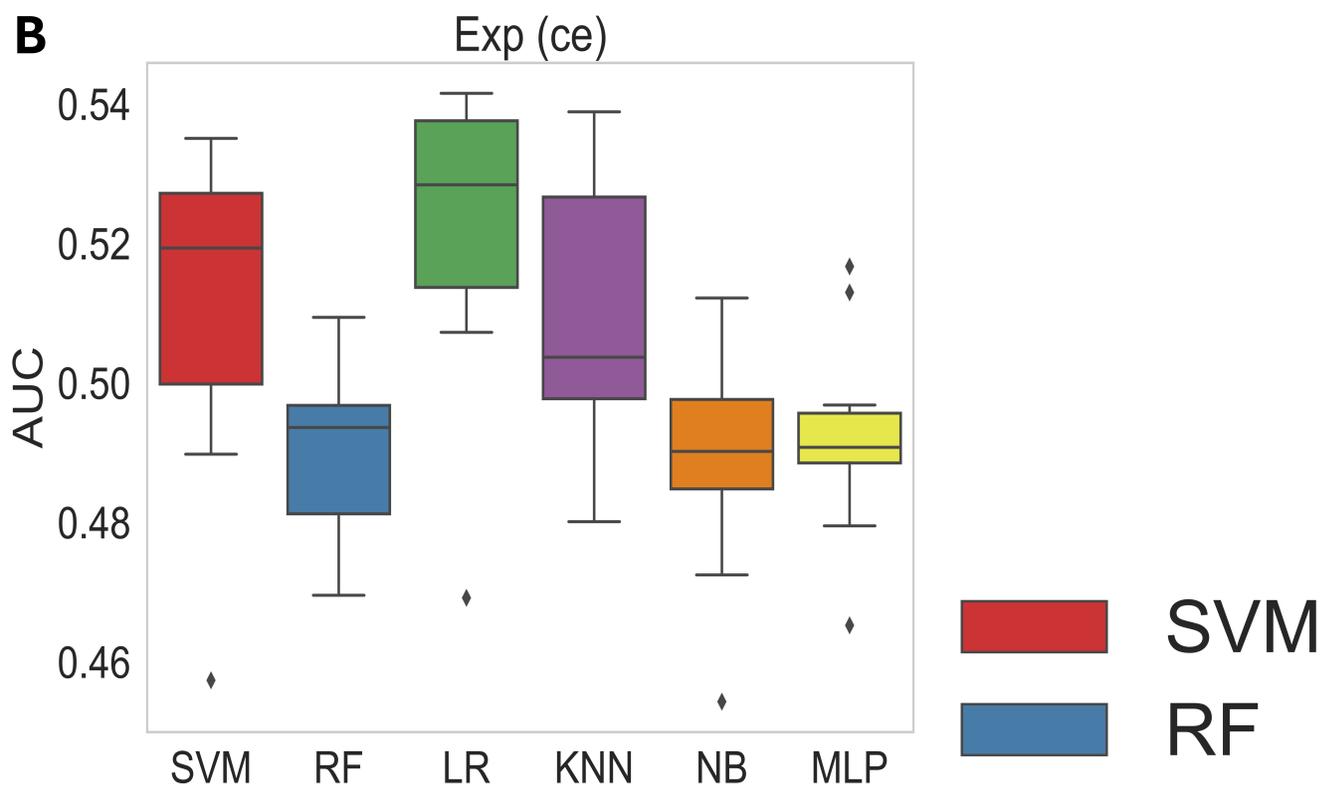
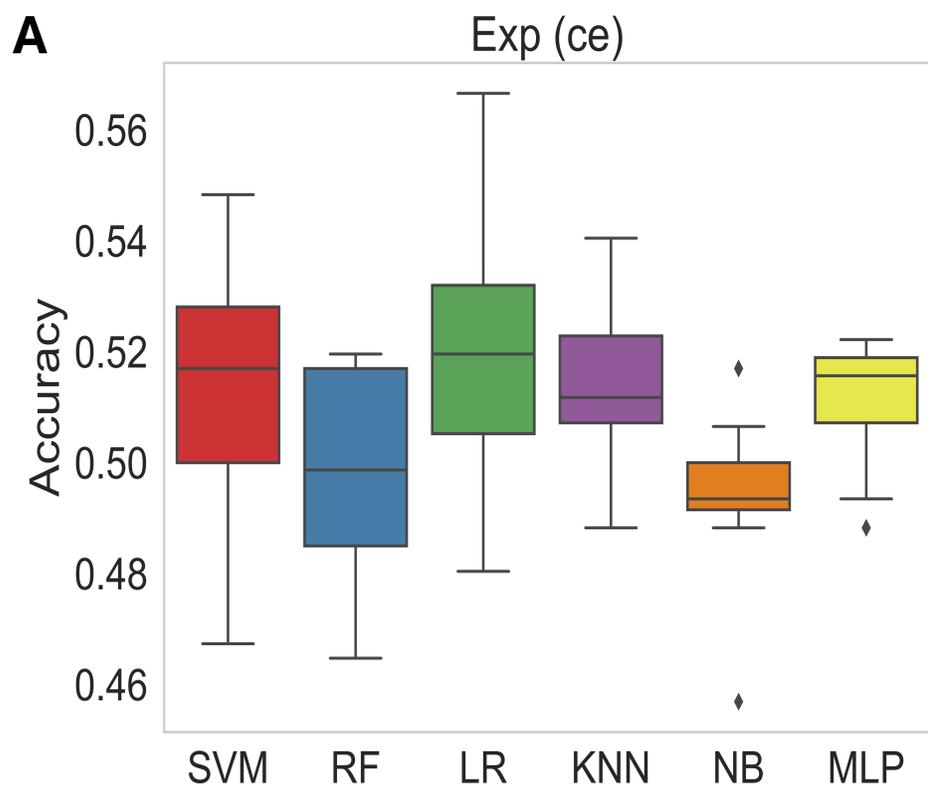


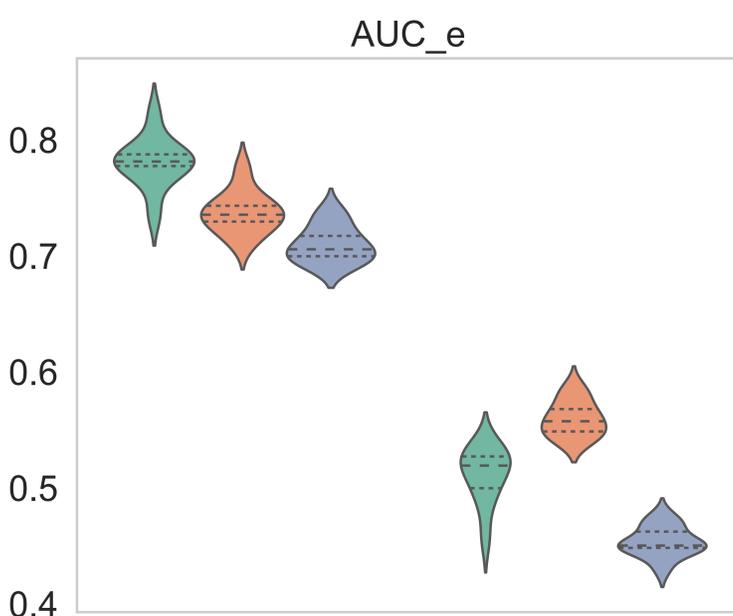
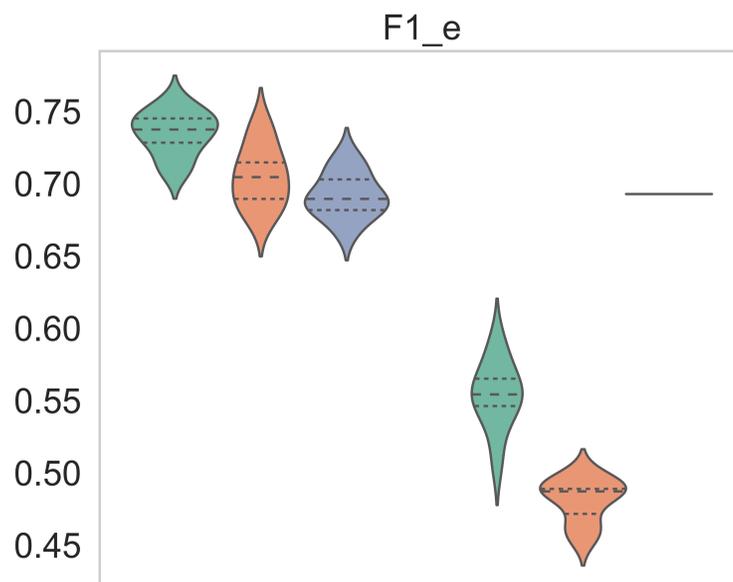
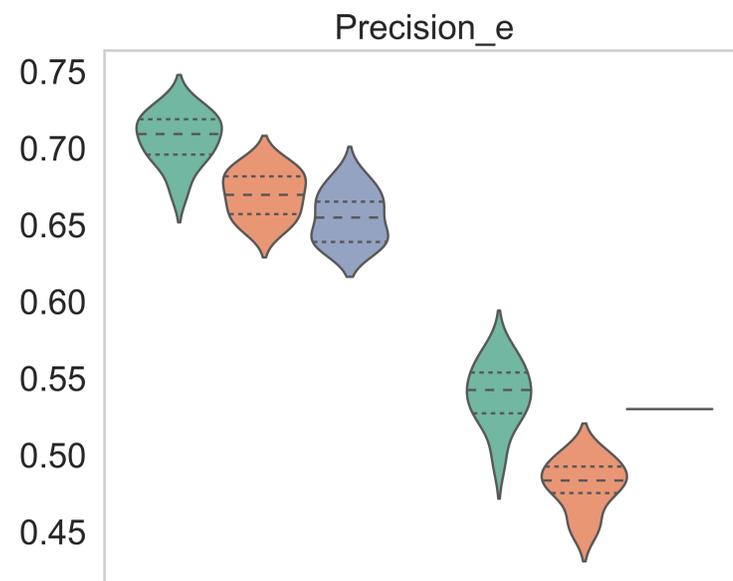
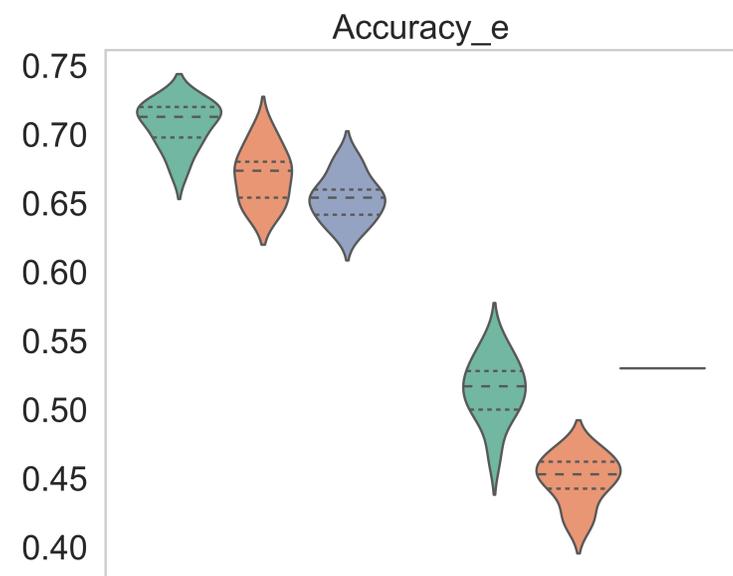
New Training Set2



AD Diagnosis

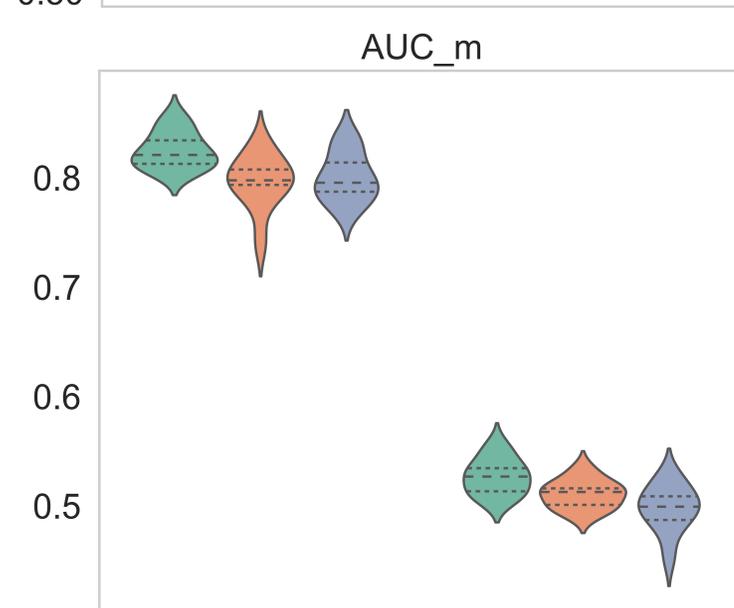
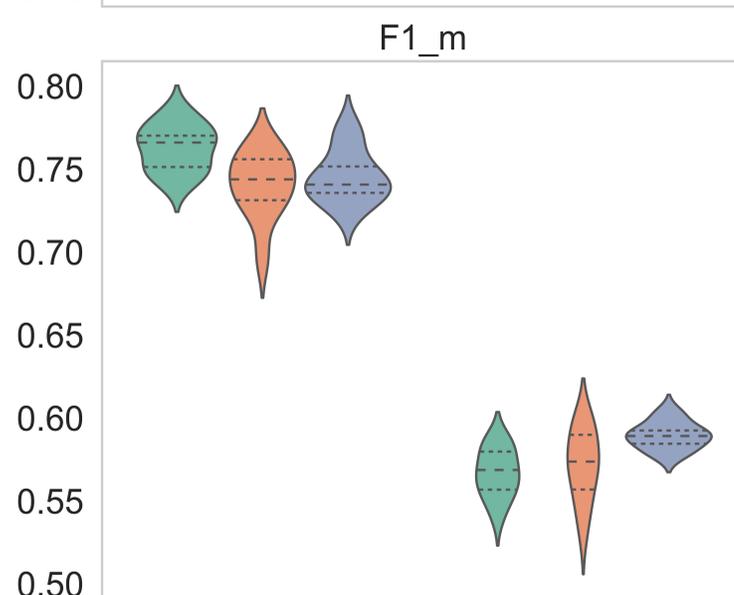
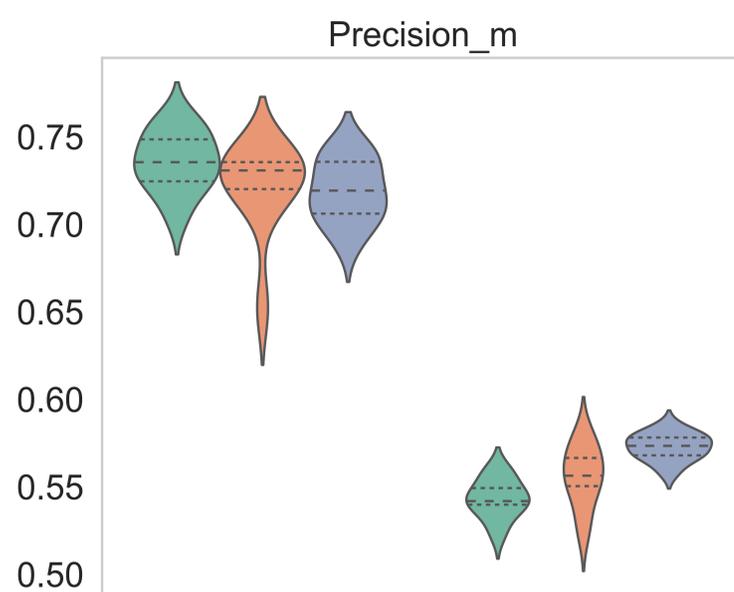
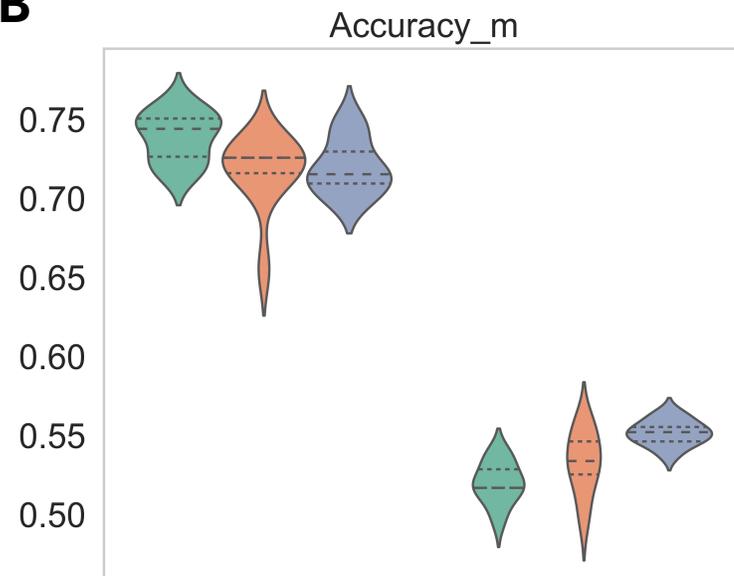




A

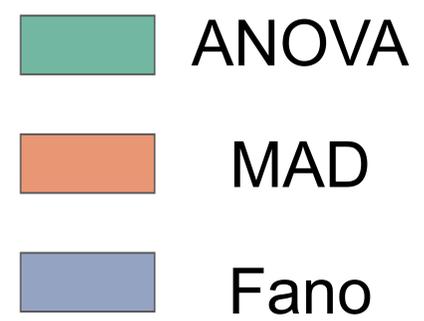
Stacking

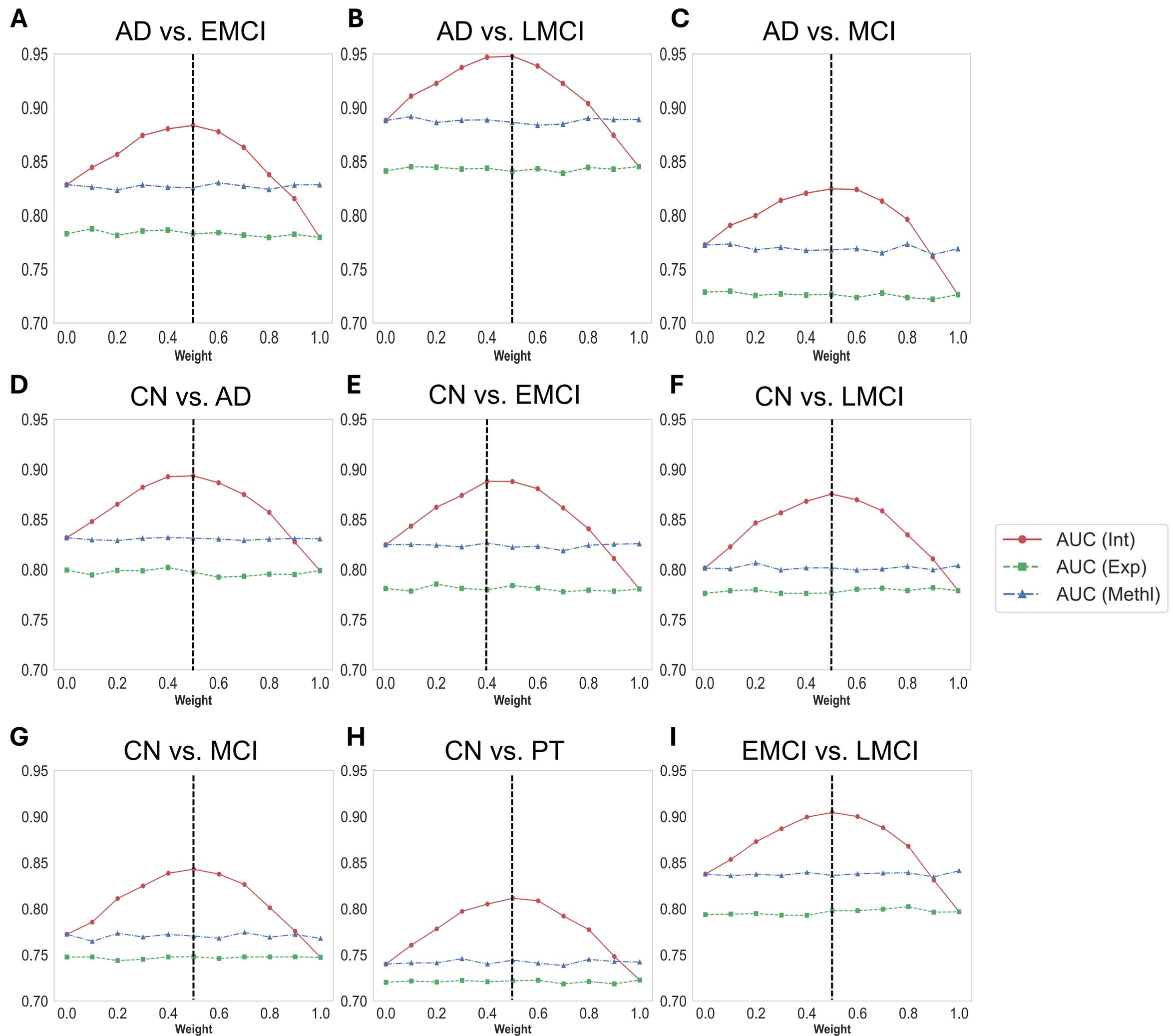
Ori.

B

Stacking

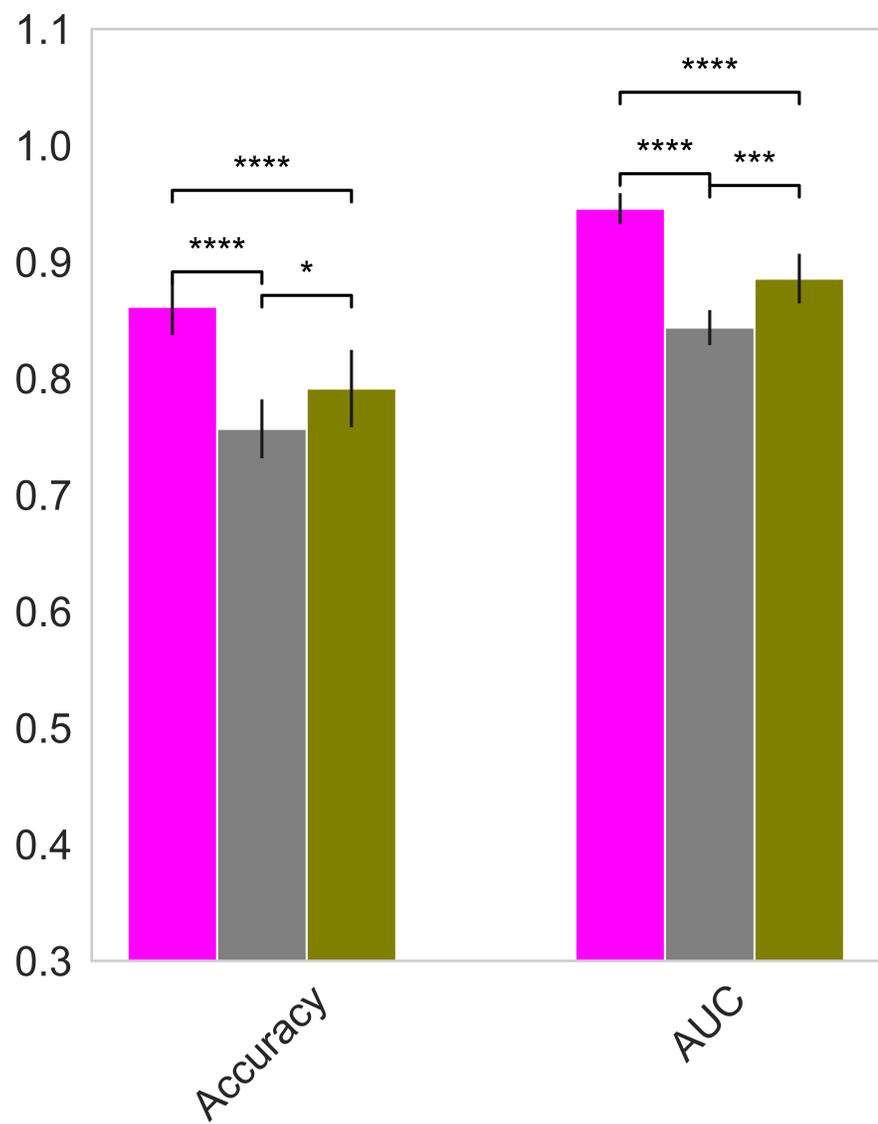
Ori.



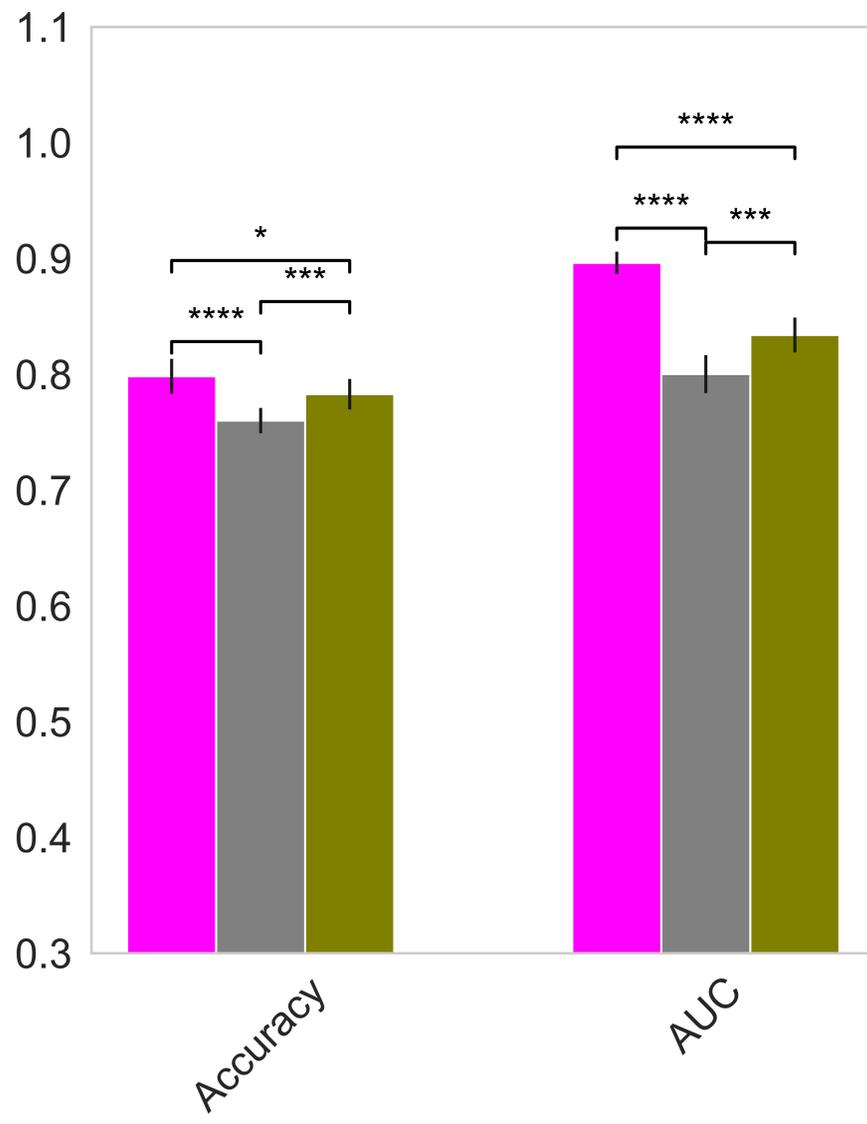


A

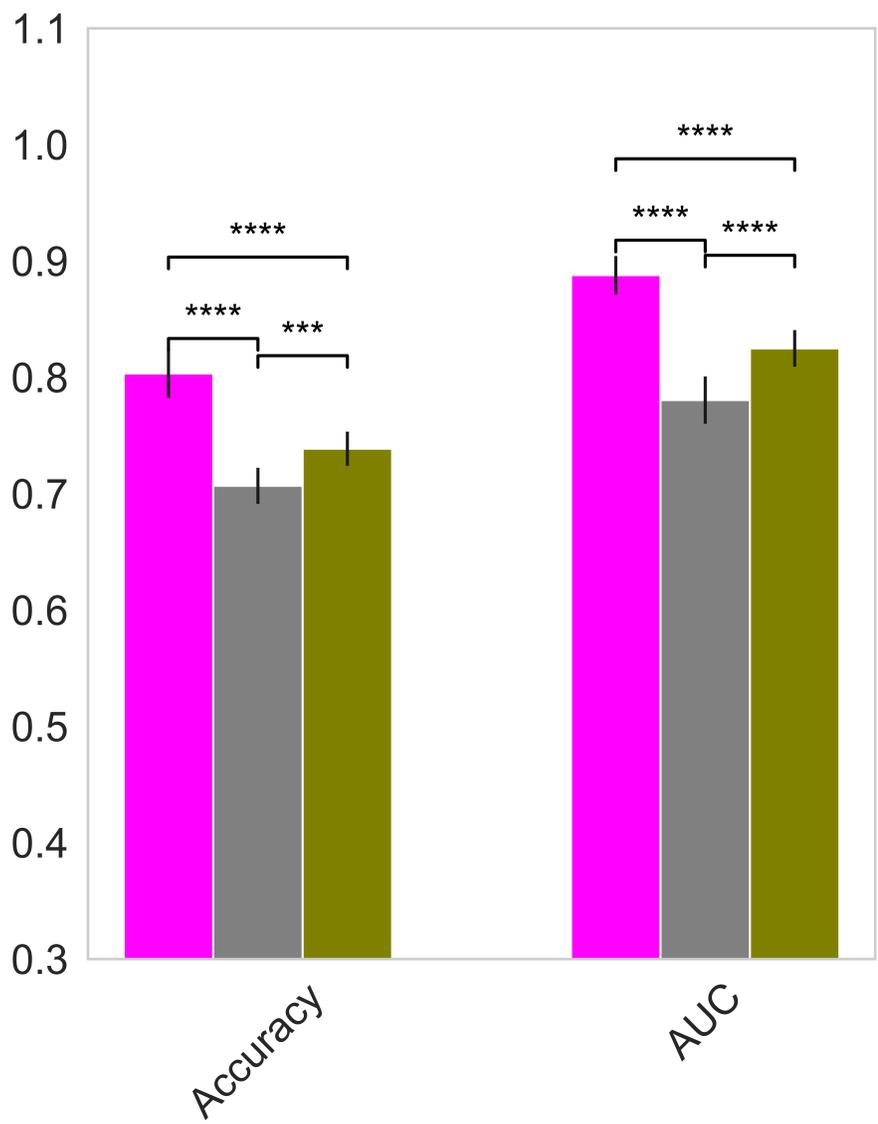
AD vs. LMCI

**B**

CN vs. AD

**C**

CN vs. EMCI

**D**

EMCI vs. LMCI

