From Blood to Brain: Uncertainty-Aware Adaptive Fusion for Alzheimer's Staging and Progression Under Incomplete Multimodal Profiles

Shahzad Ali[1,2][0000−0002−0608−9515] and Sara Garbarino[2,3][0000−0002−3583−3630]

[1] Dipartimento di Farmacia e Biotecnologie, Alma Mater Studiorum – Università di Bologna, Bologna, Italy
[2] Life Science Computational Laboratory, IRCCS Azienda Ospedaliera Metropolitana, Genova, Italy
[3] Dipartimento di Matematica, Università degli Studi di Genova, Genova, Italy
shahzad.ali6@unibo.it, sara.garbarino@unige.it

Abstract. Alzheimer's disease (AD) staging and progression prediction require multiple biomarker modalities, yet complete profiles are uncommon because examinations differ in accessibility, cost, and invasiveness. Existing multimodal studies often utilize a fixed, complete modality set or reconstruct unavailable examinations. We propose an uncertainty-aware adaptive fusion framework that operates directly on observed modalities. It integrates plasma biomarkers, structural MRI (sMRI), diffusion MRI (dMRI), fluorodeoxyglucose PET (FDG), amyloid-PET (AMY), and cerebrospinal fluid (CSF) biomarkers without synthesizing unavailable modalities. Modality-specific experts produce Platt-calibrated risk scores and predictive uncertainty estimates. An ElasticNet fusion layer combines these, encoding uncertainty, confidence, availability, and inter-modality disagreement explicitly. We used the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset to evaluate the proposed framework across three AD staging and progression prediction tasks using leakage-safe nested cross-validation. The proposed framework improves balanced accuracy (BACC) from 0.772 to 0.802 for mild cognitive impairment (MCI) versus dementia, reducing Brier score from 0.139 to 0.111 and expected calibration error (ECE) by 73% for cognitively normal (CN) versus dementia. Complete six-modality acquisition was not uniformly beneficial. In addition, the fusion coefficients and held-out SHAP explanations revealed task-specific evidence profiles. The framework reframes incomplete multimodal learning as calibrated evidence escalation rather than missing-test reconstruction. Code: https://github.com/alishahzi/BloodToBrain_AdaptiveFusion.

Keywords: Alzheimer's disease · Adaptive multimodal fusion · Missing modalities · Uncertainty estimation · Biomarker escalation · Explainability.

1 Introduction

Alzheimer's disease (AD) reflects molecular, metabolic, structural, and connectivity alterations that evolve across the clinical course. Accurate staging therefore requires diverse biomarkers that capture different aspects of disease biology [6]. sMRI measures neurodegeneration, FDG reflects glucose metabolism, AMY measures fibrillar amyloid deposition, dMRI characterizes white-matter integrity, and plasma or CSF assays provide molecular evidence. Recent blood-based biomarkers, particularly plasma p-tau217, have shown high diagnostic accuracy [12,3]. In practice, complete multimodal profiles are often unavailable because modality acquisition depends on cost, invasiveness, acquisition burden, and clinical indication.

This incomplete modality availability makes multimodal AD modelling difficult. Existing multimodal AD methods usually address missing modalities through dropout training [9], latent-space reconstruction [11], or synthesis of missing neuroimages [16,18]. Such approaches treat missingness mainly as a reconstruction or feature-completion problem. However, they do not determine whether additional biomarker acquisition is clinically necessary for a particular task or subject. Progressive confidence-gating approaches [9] still rely on cross-stage feature alignment and evaluate only subjects with a complete modality set. Moreover, existing studies rarely quantify predictive uncertainty or assess calibrated risk along clinically ordered acquisition pathways [8], yet clinical models should communicate uncertainty [7] and determine when additional testing adds value rather than providing only rankings [15].

To address these requirements, we propose an adaptive expert-fusion framework for incomplete multimodal cohorts. It predicts from any observed subset of six neuroimaging and biofluid sources without synthesizing missing tests, encodes calibrated expert risk with uncertainty, confidence, availability, and inter-modality disagreement as explicit fusion inputs, and provides two-level interpretability via ElasticNet attribution and held-out SHAP explanations. The central question shifts from how to reconstruct a missing modality to whether observed evidence is sufficient.

2 Methods

2.1 Problem Formulation

Let $\mathcal{M}$={sMRI, dMRI, FDG, AMY, Plasma, CSF}. For subject i, $\mathbf{x}_i^{(m)}$ is the feature vector for modality m, $a_i^{(m)} \in \{0,1\}$ the availability indicator, and $y_i \in \{0,1\}$ the task label. The proposed framework has two stages (Fig. 1). First, each modality-specific expert estimates a calibrated risk $\tilde{p}_i^{(m)}=c_m(f_m(\mathbf{x}_i^{(m)}))$. Second, an adaptive fusion model combines observed expert evidence to produce the final prediction $\hat{p}_i=g(\boldsymbol{\phi}_i)$.

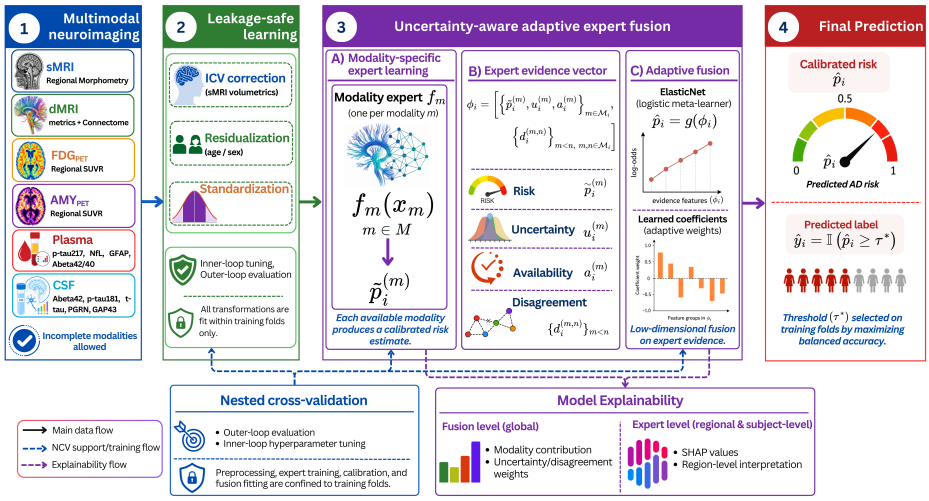


Fig. 1: Overview of the proposed uncertainty-aware adaptive fusion framework.

2.2 Modality-Specific Expert Learning and Uncertainty

A modality-specific XGBoost expert f_m is trained independently for each modality on subjects for whom that modality is available. No data from other modalities is imputed. Raw probabilities are Platt-calibrated within each training fold. Predictive uncertainty is quantified by binary entropy:

$$u_i^{(m)} = -\tilde{p}_i^{(m)} \log \tilde{p}_i^{(m)} - (1 - \tilde{p}_i^{(m)}) \log(1 - \tilde{p}_i^{(m)}) \tag{1}$$

Expert confidence is $\kappa_i^{(m)} = \max(\tilde{p}_i^{(m)}, 1 - \tilde{p}_i^{(m)})$ and pairwise inter-modality disagreement is $\Delta_i^{(m,n)} = |\tilde{p}_i^{(m)} - \tilde{p}_i^{(n)}|$ for all available pairs.

2.3 Adaptive Fusion and Calibration

The subject-level evidence vector contains calibrated risks, predictive uncertainty, confidence, availability, and pairwise disagreement from the observed modality experts:

$$\boldsymbol{\phi}_i = \left[\tilde{p}_i^{(m)},\ u_i^{(m)},\ \kappa_i^{(m)},\ a_i^{(m)},\ \Delta_i^{(m,n)}\right]_{m,n \in \mathcal{M}} \tag{2}$$

The fusion model is ElasticNet-regularized logistic regression, which produces the fused risk:

$$\hat{p}_i = \sigma\left(\alpha_0 + \boldsymbol{\phi}_i^\top \boldsymbol{\alpha}\right), \quad \mathcal{L}(\boldsymbol{\alpha}) = -\sum_i \ell(y_i, \hat{p}_i) + \lambda\left[\eta\|\boldsymbol{\alpha}\|_1 + \tfrac{1-\eta}{2}\|\boldsymbol{\alpha}\|_2^2\right] \tag{3}$$

The fusion layer meta-learner is trained on cross-fitted out-of-fold expert predictions, preventing it from observing predictions produced by experts trained

Table 1: Study cohort demographics and modality availability.

	CN (n=878)	**MCI** (n=1,078)	**Dementia** (n=403)	**sMCI** (n=780)	**pMCI** (n=298)	**All** (n=2,359)
Age (yrs)	72.3±6.6	72.7±7.6	74.9±7.8	72.4±7.8	73.7±7.2	72.9±7.4
Female	56 %	42 %	44 %	42 %	42 %	48 %
MMSE	29.1±1.2	27.6±1.9	23.1±2.2	27.9±1.8	26.9±1.8	27.4±2.7
Modality availability						
Plasma	548 (62%)	363 (34%)	123 (31%)	276 (35%)	87 (29%)	1,034 (44%)
sMRI	878 (100%)	1,078 (100%)	403 (100%)	780 (100%)	298 (100%)	2,359 (100%)
dMRI	346 (39%)	250 (23%)	97 (24%)	208 (27%)	42 (14%)	693 (29%)
FDG	399 (45%)	837 (78%)	296 (73%)	635 (81%)	202 (68%)	1,532 (65%)
AMY	379 (43%)	475 (44%)	143 (35%)	382 (49%)	93 (31%)	997 (42%)
CSF	597 (68%)	759 (70%)	262 (65%)	554 (71%)	205 (69%)	1,618 (69%)

on the same subjects. Missing modalities are not imputed. Absence is encoded via $a_i^{(m)}$; the fusion model operates on any observed subset without reconstruction. Final probabilities are Platt-calibrated, and the decision threshold is selected using training/inner-validation predictions only within each outer fold by maximising balanced accuracy (BACC).

2.4 Explainability

In addition, we examined how the fused prediction was formed at both the fusion and expert levels. At the fusion level, normalised absolute ElasticNet coefficients quantify the contribution of each modality group and feature family (risk, uncertainty, confidence, availability, disagreement). Subject-level dominant modality was defined as the most confident available expert. At the expert level interpretation, SHAP values [10] were computed only for held-out outer-fold subjects and aggregated by mean absolute contribution.

3 Experiments and Dataset

Cohort. We obtained data from ADNI (https://adni.loni.usc.edu/) [17]. The analysis cohort comprised 2,359 participants: 878 cognitively normal (CN), 1,078 MCI, and 403 dementia. The MCI participants were further stratified as stable (sMCI; $n = 780$) or progressive (pMCI; $n = 298$) according to conversion to dementia within 36 months. Three binary tasks were defined: CN vs. dementia (n=1,281; 878/403), MCI vs. dementia (n=1,481; 1,078/403), and sMCI vs. pMCI (n=1,078; 780 stable / 298 progressive; 36-month conversion label). Class imbalance ranged from 1:2.2 (CN vs. dementia) to 1:2.7 (MCI vs. dementia). Table 1 summarises cohort demographics. The modality availability reflected real cohort incompleteness: sMRI was available for 2,359 participants, FDG for 1,532, AMY for 997, dMRI for 693, plasma for 1,034, and CSF for 1,618. Only 50, 66, and 48 participants had all six sources in the three task cohorts, respectively.
Preprocessing and features. All neuroimaging scans were preprocessed using Clinica [13]. sMRI yielded 84 Desikan-Killiany (DK) atlas cortical thickness and

subcortical volume features. dMRI provided JHU atlas-based DTI measures (FA, MD, AD, and RD) and 84-node DK connectomes constructed using MRtrix3 tractography [14] and NetworkX [5]. FDG-PET and AMY-PET yielded 84 regional SUVR features each. Plasma included p-tau217, GFAP, NfL, and Aβ42/40; CSF included Aβ42, total tau, p-tau181, PGRN, and GAP43. Within each training fold, sMRI volumes were ICV-adjusted, imaging features were residualized for age and sex, and all features were standardized; transformations were fixed before applying to held-out data.

Experimental protocol. Support vector machine (SVM), random forest (RF), and XGBoost were first benchmarked across single-source and plasma-anchored escalation settings. XGBoost achieved the strongest overall baseline performance and was therefore retained as the fixed modality-specific expert learner and prespecified comparator for fusion evaluation. Model selection used stratified nested cross-validation (NCV) with five outer folds and three inner folds. Within each inner loop, 50 Optuna trials [1] were used to tune the baseline models and the fusion model. The plasma-anchored pathway sequentially added sMRI, dMRI, FDG-PET, AMY-PET, and CSF to assess ordered acquisition scenarios. We also evaluated complete six-modality fusion and any-available fusion. Performance metrics included BACC, area under the receiver operating characteristic curve (auROC), area under the precision–recall curve (auPRC), Brier score, and ECE; outer-fold means with 95% CIs are reported. Fold differences used paired sign-flip permutation tests (10,000 permutations).

4 Results

Baseline Performance. Table 2 summarises XGBoost reference performance along the escalation configurations. Plasma alone achieved a BACC of 0.796 for CN vs. dementia and 0.696 for sMCI vs. pMCI. Along the escalation pathway, performance generally improved as additional modalities were incorporated, although the optimal configuration was task-specific. The plasma+sMRI+FDG+AMY (+AMY) setting achieved peak BACC for CN vs. dementia (0.910) and MCI vs. dementia (0.772), while plasma+sMRI+FDG (+FDG) was best for sMCI vs. pMCI (0.732).

Adaptive Fusion Performance. Table 3 compares the proposed adaptive fusion framework with the selected XGBoost reference. For CN vs. dementia, the any-available fusion achieved BACC 0.918 vs. 0.910 for the reference, with higher auROC (0.968) and auPRC (0.948). For MCI vs. dementia, the proposed framework increased BACC from 0.772 to 0.802 (Δ+0.030, p=0.125) and auROC from 0.857 to 0.896. Brier score fell from 0.139 to 0.111. For sMCI vs. pMCI, the XGBoost reference is the best plasma-anchored configuration (Plasma+sMRI+FDG, BACC 0.732). The adaptive fusion at the +AMY step achieved BACC 0.755 (Δ+0.023, p=0.256) and improved auPRC from 0.569 to 0.608. Fold-wise sign-flip tests did not establish formal statistical superiority, consistent with five aligned evaluation folds. Directional improvements were observed across all tasks and metrics, except ECE for sMCI vs. pMCI. Probability

Table 2: Reference baseline performance along plasma-anchored escalation pathway.

	CN vs. Dementia			MCI vs. Dementia			sMCI vs. pMCI		
Setting	BACC	auROC	auPRC	BACC	auROC	auPRC	BACC	auROC	auPRC
Plasma	0.796	0.865	0.659	0.639	0.714	0.461	0.696	0.752	0.537
+sMRI	0.849	0.949	0.849	0.732	0.831	0.618	0.706	0.774	0.538
+dMRI	0.878	0.973	0.880	0.747	0.837	0.674	0.620	0.757	0.390
+FDG	0.857	0.954	0.928	0.749	0.860	0.688	**0.732**	0.791	0.569
+AMY	**0.910**	0.963	0.903	**0.772**	0.857	0.650	0.684	0.842	0.660
+CSF	0.897	0.979	0.938	0.754	0.825	0.586	0.702	0.855	0.683
Six-modality	0.844	0.896	0.917	0.632	0.775	0.738	0.625	0.853	0.704

Table 3: Performance comparison across adaptive fusion vs. XGBoost reference.

Task	**Model**	**BACC**	**auROC**	**auPRC**	$p_{\Delta\text{BACC}}$	**Brier**	**ECE**
CN vs. Dem.	Ref. (+AMY)	0.910±0.029	0.963±0.016	0.903±0.054	–	.077	.140
	Fusion (any)	**0.918±0.020**	**0.968±0.013**	**0.948±0.021**	0.625	**.053**	**.038**
MCI vs. Dem.	Ref. (+AMY)	0.772±0.038	0.857±0.027	0.650±0.127	–	.139	.157
	Fusion (+CSF)	**0.802±0.060**	**0.896±0.041**	**0.687±0.170**	0.125	**.111**	**.124**
sMCI vs. pMCI	Ref. (+FDG)	0.732±0.087	0.791±0.084	0.569±0.124	–	.175	.176
	Fusion (+AMY)	**0.755±0.118**	**0.820±0.136**	**0.608±0.202**	0.256	**.142**	.185

Note. Values are reported as mean ± 95 % CI across outer folds. ΔBACC = fusion − reference; p from sign-flip permutation. +FDG: Plasma+sMRI+FDG; +AMY: Plasma+sMRI+FDG+AMY; +CSF: Plasma+sMRI+FDG+AMY+CSF.

calibration improved for both staging tasks (Fig. 2): ECE fell by 73% for CN vs. dementia (0.140→0.038) and by 21% for MCI vs. dementia; Brier score improved across all three tasks. For sMCI vs. pMCI, ECE was marginally less favourable (0.185 vs. 0.176), consistent with the greater difficulty of calibrating progression estimates. Reliable risk probabilities are a separate and necessary requirement for escalation decisions [15,7].

Biomarker Escalation Analysis. Table 4 shows that predictive value did not increase monotonically with modality count. Plasma supplied useful initial evidence and sMRI produced the largest early gain. The optimal checkpoint was task-specific: CN vs. dementia peaked at any-available (0.788→0.918); MCI vs. dementia improved most after adding CSF (0.656→0.802); sMCI vs. pMCI peaked at +AMY (0.755). Complete six-modality fusion produced unstable estimates due to small eligible cohorts (n=48–66). This pattern clarifies that any-available fusion should be interpreted as a coverage-preserving operating mode rather than a universally optimal configuration: it was most effective for CN vs. dementia, whereas MCI vs. dementia and sMCI vs. pMCI benefited more from task-specific escalation to informative biomarker subsets.

Explainability. We provided model explanations and interpretability at two levels: fusion level and expert level. At the fusion level (Fig. 3), FDG risk contributed 20.4% and sMRI risk 15.2% of fusion weight for CN vs. dementia.

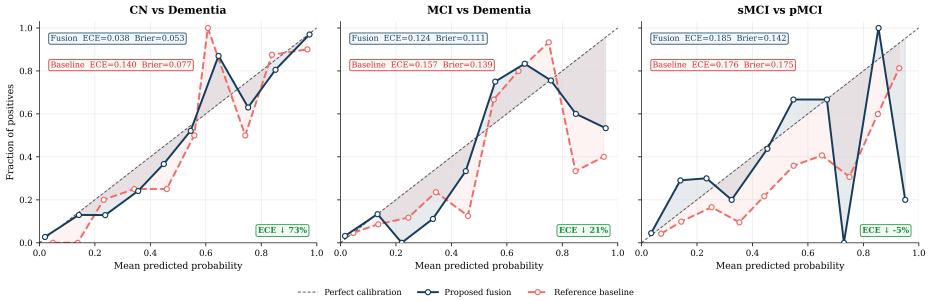


Fig. 2: Reliability diagrams comparing probability calibration between the proposed adaptive fusion model and reference baseline.

Table 4: Adaptive fusion performance across the plasma-anchored escalation pathway.

	CN vs. Dementia			MCI vs. Dementia			sMCI vs. pMCI		
Fusion setting	BACC	auROC	auPRC	BACC	auROC	auPRC	BACC	auROC	auPRC
Plasma only	0.788	0.865	0.659	0.656	0.701	0.452	0.691	0.752	0.537
+sMRI	0.868	0.946	0.860	0.754	0.820	0.613	0.746	0.802	0.619
+dMRI	0.904	0.955	0.870	0.769	0.850	0.678	0.521	0.683	0.407
+FDG	0.874	0.947	0.908	0.777	0.849	0.671	0.754	0.845	0.676
+AMY	0.890	0.959	0.919	0.725	0.853	0.677	**0.755**	0.820	0.608
+CSF	0.910	0.979	0.946	**0.802**	0.896	0.687	0.703	0.834	0.636
Six-modality	0.817	0.967	0.967	0.750	0.815	0.753	0.530	0.745	0.510
Any available	**0.918**	0.968	0.948	0.744	0.825	0.620	0.745	0.812	0.629

For MCI vs. dementia, FDG dominated at 44.6%. For sMCI vs. pMCI, fusion-summary features were most prominent at 62.9%, indicating strong reliance on combined evidence patterns. At the expert level (Fig. 4), we show the SHAP brain map for the best standalone expert per task (sMRI for CN vs. dementia; FDG for MCI vs. dementia; AMY for sMCI vs. pMCI). SHAP highlighted bilateral entorhinal cortex, hippocampus, and amygdala for the sMRI expert in CN vs. dementia — canonical cortical thinning and subcortical atrophy sites in early AD [4,6]. For MCI vs. dementia, FDG SHAP concentrated in the right isthmus cingulate, inferior temporal gyrus, pallidum, and inferior parietal cortex. For sMCI vs. pMCI, AMY SHAP centered on the right precuneus, left inferior temporal cortex, and left precuneus, aligning with early amyloid accumulation sites [2]. All SHAP values were computed on held-out subjects only and support model auditing rather than causal claims [10].

5 Discussion

The proposed framework shifts the incomplete-modality question from how to reconstruct a missing test to whether observed evidence is sufficient. Its

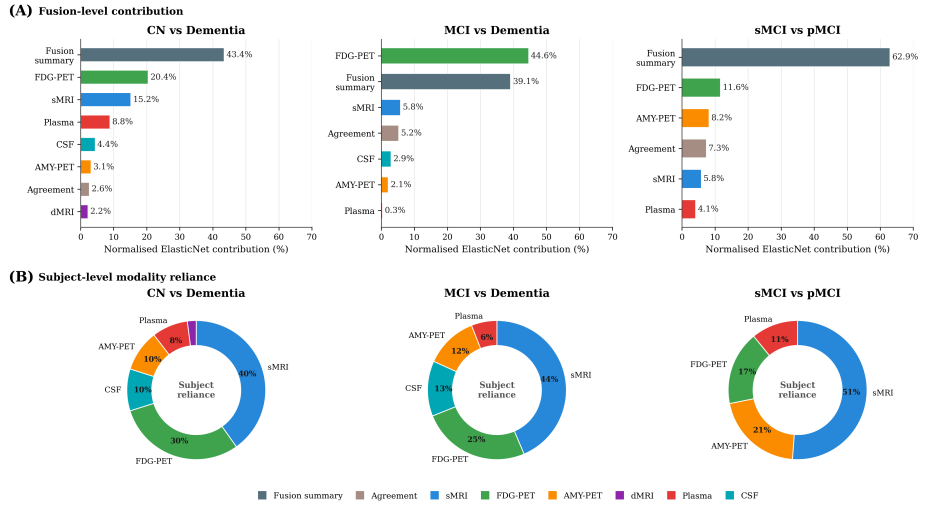
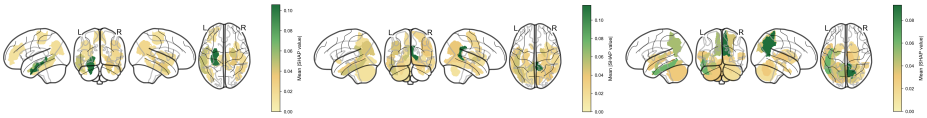


Fig. 3: Fusion-level contributions and subject-level dominant modalities across tasks.



(a) CN vs. Dementia (b) MCI vs. Dementia (c) sMCI vs. pMCI

Fig. 4: Expert-level SHAP evidence for the dominant modality in each task.

contribution lies not in universal discrimination superiority but in preserving cohort coverage, improving probability reliability, and revealing task-specific escalation patterns. The clearest discriminative gain was for MCI vs. dementia, where fusion reconciled disagreeing single-source estimates most effectively. For progression prediction, auPRC, Brier score and BACC improved, while ECE remained comparable.

The escalation analysis challenges the assumption that more modalities are always better. Complete profiles covered very small, selected cohorts and did not consistently improve performance. Plasma provided a useful entry point, sMRI delivered the largest early gain, and FDG, AMY, or CSF added task-dependent value. This reflects a trade-off between cohort coverage and evidence specificity: broader availability can support the easier CN vs. dementia contrast, whereas clinically ambiguous tasks may require targeted biomarker evidence that sharpens the inter-modality disagreement and consensus signals used by the fusion layer. These task-specific optima argue for uncertainty-driven escalation over fixed multimodal protocols [7]. The framework was evaluated on the ADNI cohort, though its large, multisite, longitudinal design partially mitigates generalisability

concerns. Independent external validation remains necessary. The present evaluation of the proposed framework was conducted within the same escalation setting and did not include direct comparisons with alternative missing-modality learning strategies. Future work should therefore validate the framework in independent cohorts, benchmark it against alternative missing-modality learning strategies, and incorporate acquisition cost into the escalation policy.

6 Conclusion

We presented an uncertainty-aware adaptive multimodal fusion framework for AD diagnosis and progression prediction. Unlike reconstruction-based approaches, it combines calibrated modality-specific risks with uncertainty, availability, and inter-modality disagreement — enabling prediction from any observed modality subset along a clinically ordered escalation pathway. The results show that the most informative biomarker checkpoint is task-dependent, and that complete-panel fusion can become unreliable at small cohort sizes. The framework provides a practical strategy for robust Alzheimer's risk stratification under real-world missing-modality conditions.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2623–2631 (2019). https://doi.org/10.1145/3292500.3330701
2. Ashton, N.J., Janelidze, S., Mattsson-Carlgren, N., Binette, A.P., Strandberg, O., Brum, W.S., Karikari, T.K., González-Ortiz, F., Di Molfetta, G., Meda, F.J., et al.: Differential roles of aβ42/40, p-tau231 and p-tau217 for alzheimer's trial selection and disease monitoring. Nature medicine **28**(12), 2555–2562 (2022). https://doi.org/10.1038/s41591-022-02074-w
3. Barthélemy, N.R., Salvadó, G., Schindler, S.E., He, Y., Janelidze, S., Collij, L.E., Saef, B., Henson, R.L., Chen, C.D., Gordon, B.A., et al.: Highly accurate blood test for alzheimer's disease is similar or superior to clinical cerebrospinal fluid tests. Nature medicine **30**(4), 1085–1095 (2024). https://doi.org/10.1038/s41591-024-02869-z
4. Braak, H., Braak, E.: Neuropathological stageing of alzheimer-related changes. Acta neuropathologica **82**(4), 239–259 (1991). https://doi.org/10.1007/BF00308809
5. Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring network structure, dynamics, and function using networkx. In: Proceedings of the 7th Python in Science Conference (SciPy 2008). pp. 11–15. Pasadena, CA, USA (2008). https://doi.org/10.25080/TCWV9851
6. Jack Jr, C.R., Andrews, J.S., Beach, T.G., Buracchio, T., Dunn, B., Graf, A., Hansson, O., Ho, C., Jagust, W., McDade, E., et al.: Revised criteria for diagnosis and staging of alzheimer's disease: Alzheimer's association workgroup. Alzheimer's & Dementia **20**(8), 5143–5169 (2024). https://doi.org/10.1002/alz.13859

7. Kompa, B., Snoek, J., Beam, A.L.: Second opinion needed: communicating uncertainty in medical machine learning. NPJ Digital Medicine **4**(1), 4 (2021). https://doi.org/10.1038/s41746-020-00367-3
8. Liu, J., Capurro, D., Nguyen, A., Verspoor, K.: Attention-based multimodal fusion with contrast for robust clinical prediction in the face of missing modalities. Journal of biomedical informatics **145**, 104466 (2023). https://doi.org/10.1016/j.jbi.2023.104466
9. Liu, L., Liu, S., Zhang, L., To, X.V., Nasrallah, F., Chandra, S.S.: Cascaded multimodal mixing transformers for alzheimer's disease classification with incomplete data. NeuroImage **277**, 120267 (2023). https://doi.org/10.1016/j.neuroimage.2023.120267
10. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I.: From local explanations to global understanding with explainable ai for trees. Nature machine intelligence **2**(1), 56–67 (2020). https://doi.org/10.1038/s42256-019-0138-9
11. Ou, Z., Jiang, C., Liu, Y., Zhang, Y., Cui, Z., Shen, D.: A graph-embedded latent space learning and clustering framework for incomplete multimodal multiclass alzheimer's disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 45–55. Springer (2024). https://doi.org/10.1007/978-3-031-72104-5_5
12. Palmqvist, S., Tideman, P., Mattsson-Carlgren, N., Schindler, S.E., Smith, R., Ossenkoppele, R., Calling, S., West, T., Monane, M., Verghese, P.B., et al.: Blood biomarkers to detect alzheimer disease in primary care and secondary care. Jama **332**(15), 1245–1257 (2024). https://doi.org/10.1001/jama.2024.13855
13. Routier, A., Burgos, N., Díaz, M., Bacci, M., Bottani, S., El-Rifai, O., Fontanella, S., Gori, P., Guillon, J., Guyot, A., et al.: Clinica: An open-source software platform for reproducible clinical neuroscience studies. Frontiers in neuroinformatics **15**, 689675 (2021). https://doi.org/10.3389/fninf.2021.689675
14. Tournier, J.D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.H., Connelly, A.: Mrtrix3: A fast, flexible and open software framework for medical image processing and visualisation. Neuroimage **202**, 116137 (2019). https://doi.org/10.1016/j.neuroimage.2019.116137
15. Van Calster, B., McLernon, D.J., Van Smeden, M., Wynants, L., Steyerberg, E.W.: Calibration: the achilles heel of predictive analytics. BMC medicine **17**(1), 230 (2019). https://doi.org/10.1186/s12916-019-1466-7
16. Wang, C., Piao, S., Huang, Z., Gao, Q., Zhang, J., Li, Y., Shan, H., Initiative, A.D.N., et al.: Joint learning framework of cross-modal synthesis and diagnosis for alzheimer's disease by mining underlying shared modality information. Medical Image Analysis **91**, 103032 (2024). https://doi.org/10.1016/j.media.2023.103032
17. Weiner, M.W., Aisen, P.S., Jack Jr, C.R., Jagust, W.J., Trojanowski, J.Q., Shaw, L., Saykin, A.J., Morris, J.C., Cairns, N., Beckett, L.A., et al.: The alzheimer's disease neuroimaging initiative: progress report and future plans. Alzheimer's & Dementia **6**(3), 202–211 (2010). https://doi.org/10.1016/j.jalz.2010.03.007
18. Xu, H., Wang, J., Feng, Q., Zhang, Y., Ning, Z.: Domain-specific information preservation for alzheimer's disease diagnosis with incomplete multi-modality neuroimages. Medical Image Analysis **101**, 103448 (2025). https://doi.org/10.1016/j.media.2024.103448