

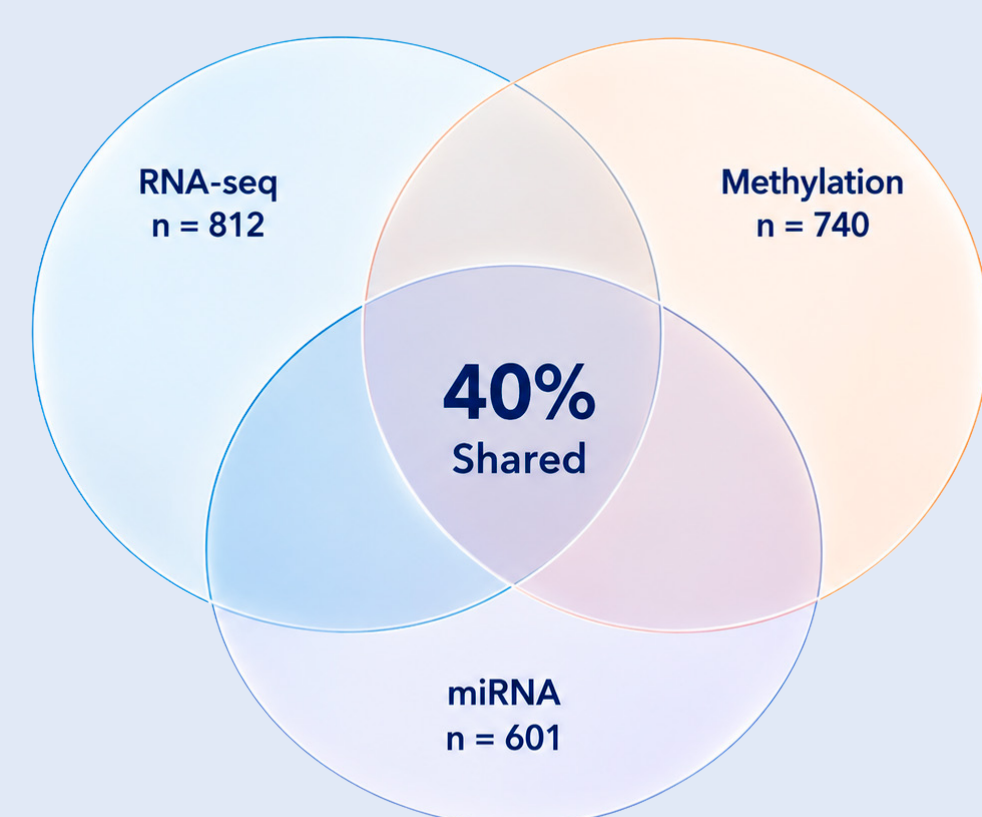
BACKGROUND

Integrative analysis of multi-omics data is critical for understanding the heterogeneous pathophysiology of Alzheimer's Disease (AD). While recent frameworks have addressed inference on incomplete data via contrastive alignment¹, current methods often treat patients as independent instances, neglecting the topological similarity between subjects. Furthermore, reliance on the strict intersection of samples across modalities limits feature learning robustness².

METHODS

We aimed to overcome these limitations by developing an asymmetrical, topology-aware framework that utilizes the "union" of available data to generate hypothesis-driven biomarkers from fragmented clinical cohorts. We utilized the ROSMAP3 aging population post-mortem brain dataset, where the intersection of samples represented only ~40% of the total available data (Figure 1). We designed a two-stage decoupled learning pipeline (Figure 2). First, we pre-trained modality-specific Graph Convolutional Network encoders on the maximal available cohort for each omic type. To capture patient topology, we constructed dynamic K-Nearest Neighbor graphs using cosine similarity for each batch. Second, we froze these encoders and trained a Graphformer4 fusion classifier on the common patient subset. Biological relevance was derived using Integrated Gradients⁵ on attention layers.

COHORT FRAGMENTATION — ROSMAP



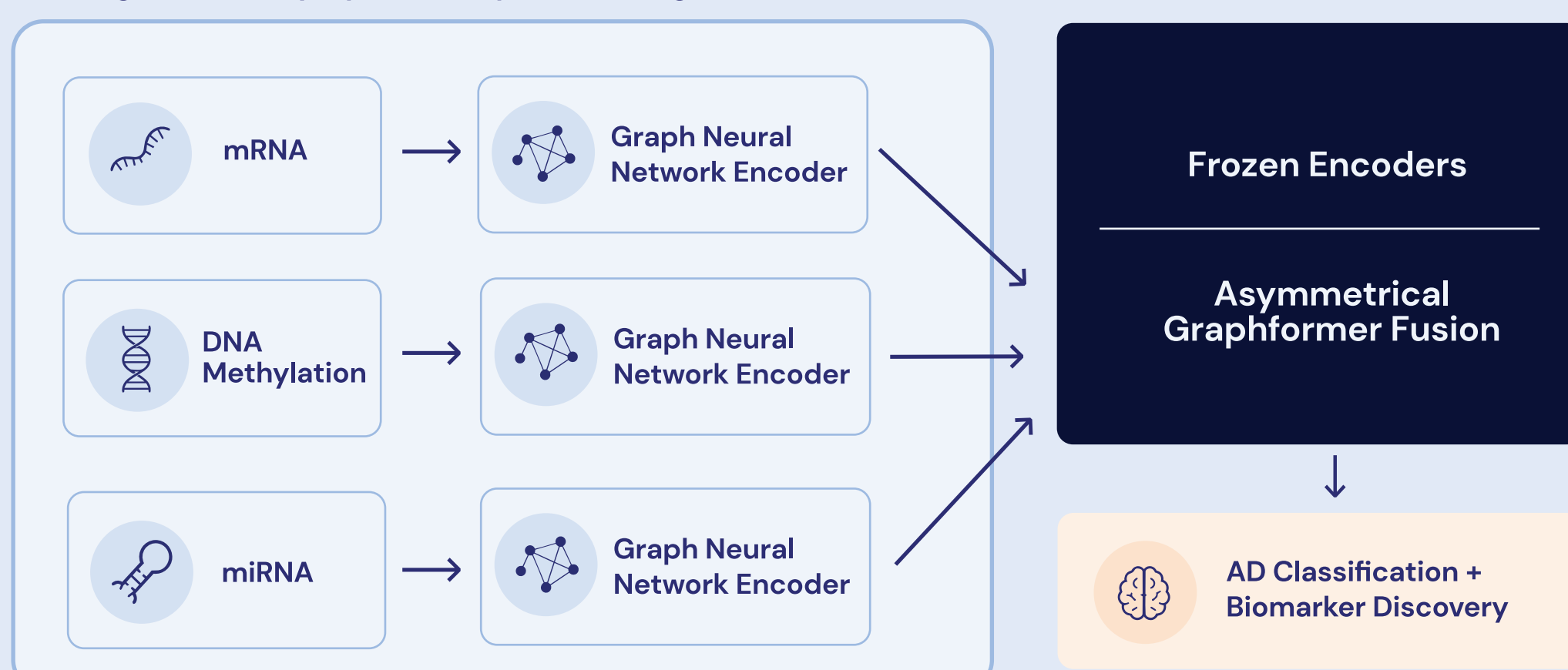
Intersection-based integration discards 60% of samples, severely limiting statistical power for discovery

Figure 1. Structural equation model illustrating multifactorial contributors to cognitive impairment. Psychosis, depression, head trauma, and education were modeled as latent predictors of cognitive impairment using CDR sum of boxes and MoCA as indicators. Standardized coefficients and p-values are shown.

TOPOLOGY-AWARE ASYMMETRICAL GRAPHFORMER FRAMEWORK

Stage 1 – Modality-Specific Graph Pretraining (Union Cohorts)

Stage 2 – Shared Subset Fusion



Interpretability Pipeline:
 Graph Attention Weights → Integrated Gradients → Feature Attribution → Pathway Enrichment → Mechanistic Hypotheses

Figure 2. Schematic of the multi-stage graph-based multi-omic fusion model, detailing the pretraining of modality-specific graph neural networks and subsequent joint model fusion for AD classification and biomarker discovery, including the connected mechanistic interpretability pipeline.

RESULTS

The asymmetrical Graphformer achieved an F1 score of 0.922 and accuracy of 0.920 (Figure 3). Interpretation of the top-ranked features revealed a pattern consistent with a "Broken Highway" hypothesis⁶, prioritizing a) the downregulation of kinesin motors (KIF5A, KIF1C)⁷, b) methylation sites linked to axonal regulators (OLFML3/cg17253459), and c) dysregulation of miRNAs (miR-10788, miR-29a9) known to regulate BACE1. These signals were distinct from metabolic stress indicators (e.g. DDIT410), suggesting a multi-stage cascade involving early transport and regulatory failure preceding impaired beta-amyloid clearance (Figure 4).

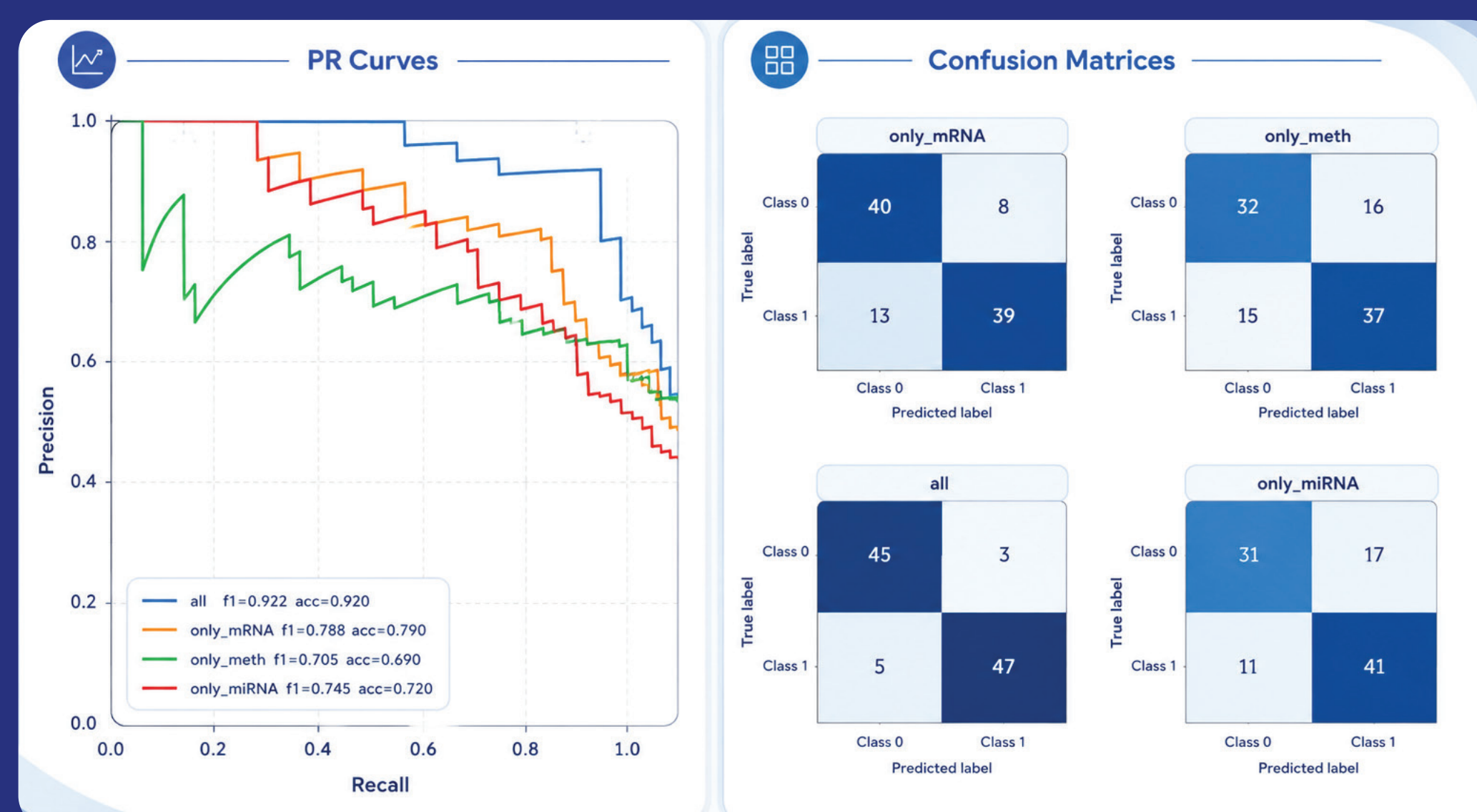


Figure 3. Performance evaluation of single-modality versus multi-omic fusion model for Alzheimer's disease binary classification. (Left) Precision-Recall (PR) curves demonstrate the superior predictive capability of the integrated model with an F1-score of 0.922 and accuracy of 0.920, significantly outperforming isolated modalities (mRNA, DNA methylation, and miRNA). (Right) Corresponding confusion matrices detail the distribution of true and false predictions across classes (0: Normal cognition, 1: Alzheimer's Disease) for each model, highlighting the increased accuracy and reduced misclassification achieved through multi-omic integration.

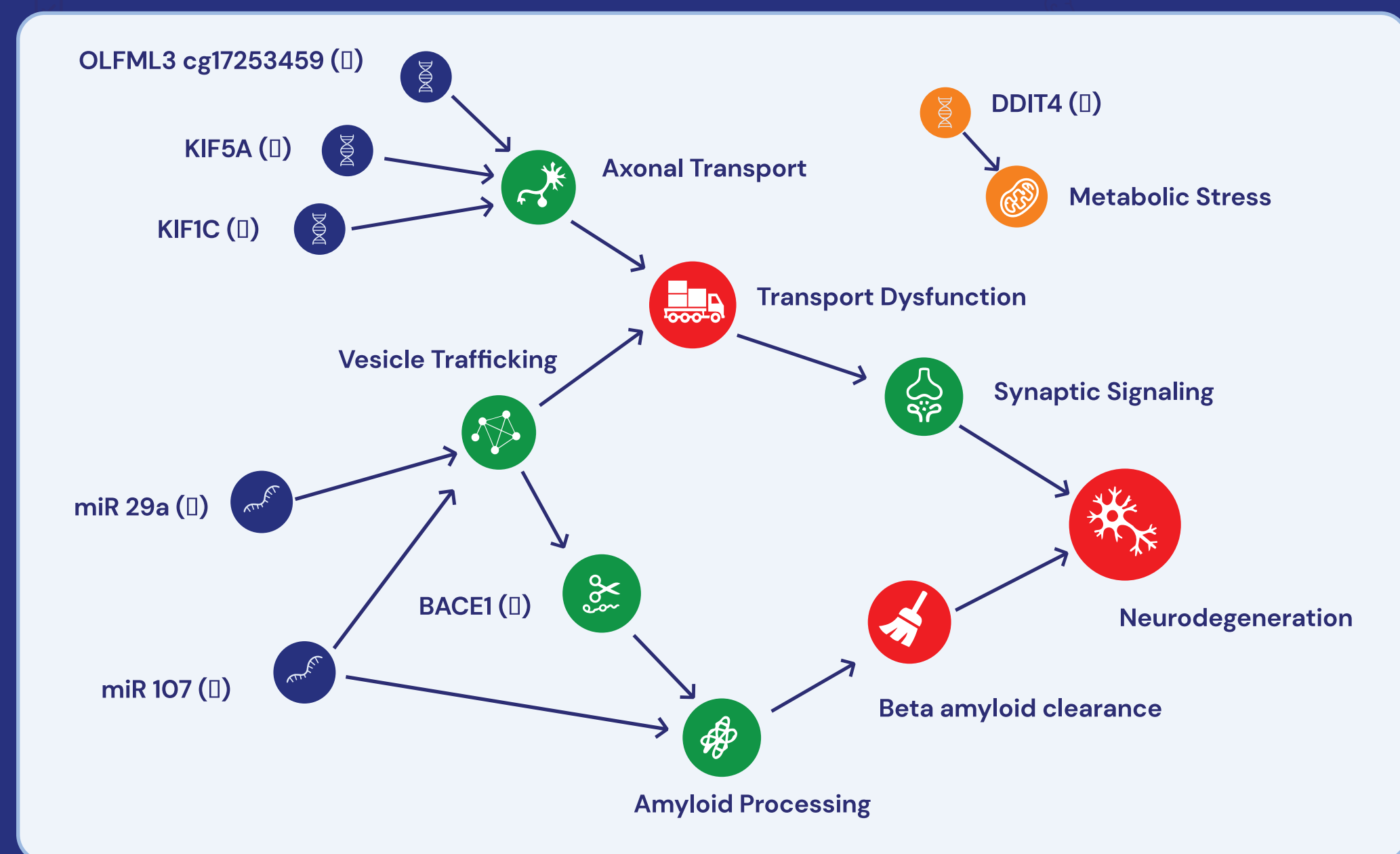
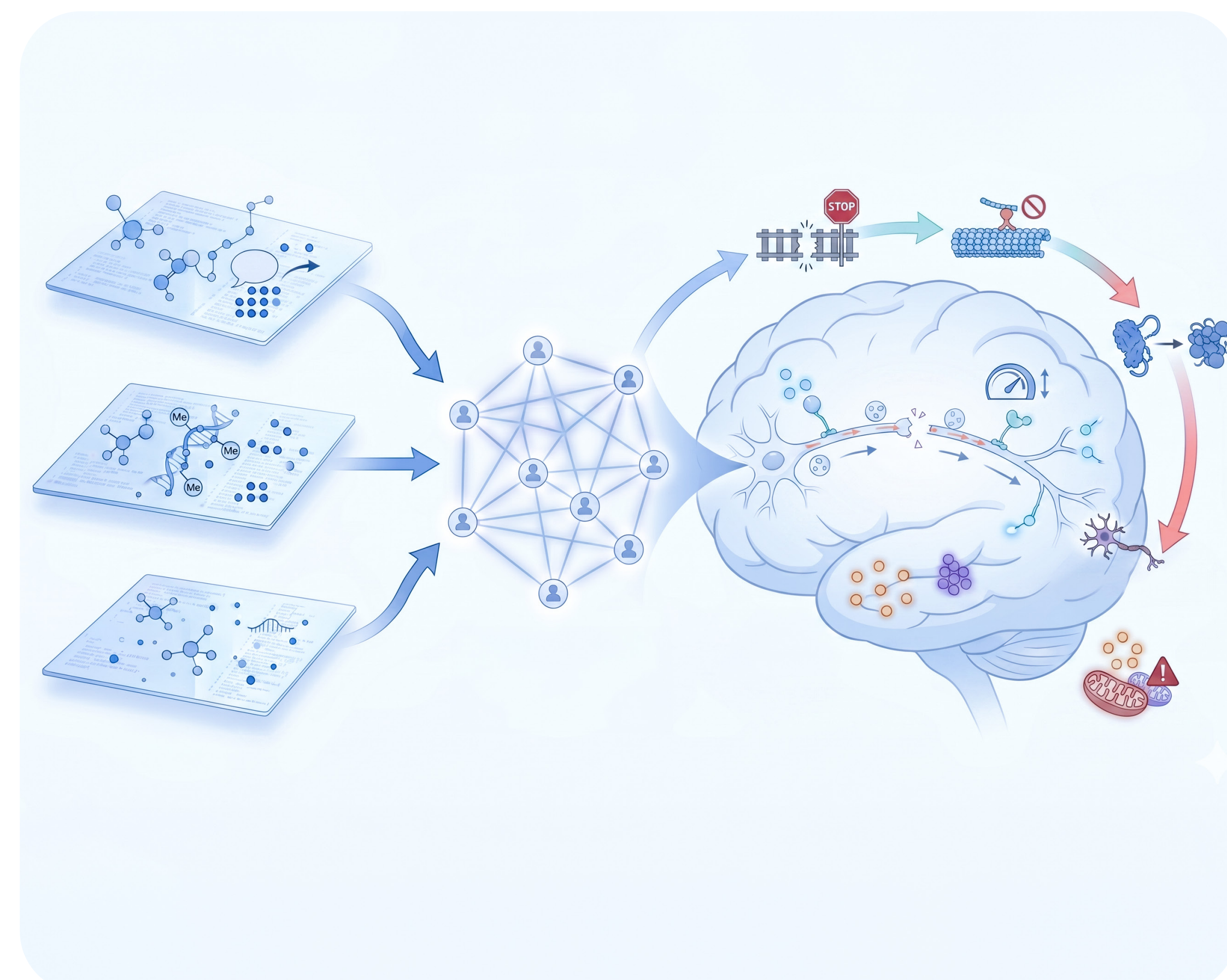


Figure 4. Performance evaluation of single-modality versus multi-omic fusion model for Alzheimer's disease binary classification. (Left) Precision-Recall (PR) curves demonstrate the superior predictive capability of the integrated model with an F1-score of 0.922 and accuracy of 0.920, significantly outperforming isolated modalities (mRNA, DNA methylation, and miRNA). (Right) Corresponding confusion matrices detail the distribution of true and false predictions across classes (0: Normal cognition, 1: Alzheimer's Disease) for each model, highlighting the increased accuracy and reduced misclassification achieved through multi-omic integration.



CONCLUSION

Our approach demonstrates that decoupling patient topology from multimodal fusion enables robust biomarker discovery and improves diagnostic precision in incomplete multi-omics datasets, a common limitation in AD research. By integrating graph-based representations with attention-guided fusion, the model prioritizes biologically plausible markers across molecular layers, including kinesin motors, miRNAs regulating BACE1, and methylation sites, offering a scalable, hypothesis-generating tool to probe mechanistic cascades. Notably, the model suggests a potential sequence in AD pathology, early cytoskeletal transport dysfunction followed by downstream metabolic failure in AD.



DISCLOSURE OF FUNDING

This research was funded by IGC Pharma, LLC.

REFERENCES

- Zhao, C., Liu, A., Zhang, X., Cao, X., Ding, Z., Sha, Q., Shen, H., Deng, H.-W., & Zhou, W. (2023). CLLSA: Cross-omics Linked embedding with Contrastive Learning and Self Attention for multi-omics integration with incomplete multi-omics data. arXiv. <https://doi.org/10.48550/arxiv.2304.05542>
- Wang T, Shao W, Huang Z, Tang H, Zhang J, Ding Z and Huang K (2021) MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. Nat. Commun. 12:3445. doi: 10.1038/s41467-021-23774-w
- Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA (2018) Religious Orders Study and Rush Memory and Aging Project. J Alzheimers Dis. 64(s1):S161-S189. doi: 10.3233/JAD-179939
- Ying, C., Cai, T., Luo, S., et al. (2021). Do Transformers Really Perform Bad for Graph Representation? NeurIPS 2021. <https://doi.org/10.48550/arxiv.2106.05234>
- Sundararajan M, Taly A and Yan Q (2017) Axiomatic attribution for deep networks. Proceedings of the 34th International Conference on Machine Learning (ICML) 70:3319-3328. doi: 10.48550/arxiv.1703.01365
- Millecamps S and Julien JP (2013) Axonal transport deficits and neurodegenerative diseases. Nat. Rev. Neurosci. 14:161-176. doi: 10.1038/nrn3380
- Wang Q, Li L and Ye Y (2019) Kinesin-1 in neurodegenerative diseases. Cell. Mol. Life Sci. 76:2853-2867. doi: 10.1007/s00018-019-03146-5
- Wang WX, Rajeev BW, Stromberg AJ, Ren N, Tang G, Huang Q, Rigoutsos I and Nelson PT (2008) Decreased expression of microRNAs in human Alzheimer's disease. PLoS One. 3:e1466. doi: 10.1371/journal.pone.0001466
- Hébert SS, Horré K, Nicolai L, Papadopoulou AS, Mandemakers W, Silahatoglu AN, Kauppinen S, Delacourte A and De Strooper B (2008) Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression. Proc. Natl. Acad. Sci. U.S.A. 105:6415-6420. doi: 10.1073/pnas.071026310
- Pérez-Ortiz JM, García-Gutiérrez MS, Navarrete M, Ginés S and Manzanares J (2024) Astrocyte protein RTP801 may contribute to cognitive decline in Alzheimer's disease. Glia. 72:1234-1245. doi: 10.1002/glia.24500



SCAN ME