

Interpretable Graph Neural Network Integration of Multimodal

Biomarkers for Alzheimer's Disease Classification

Jazmin Lagier¹, Fahad Saeed¹

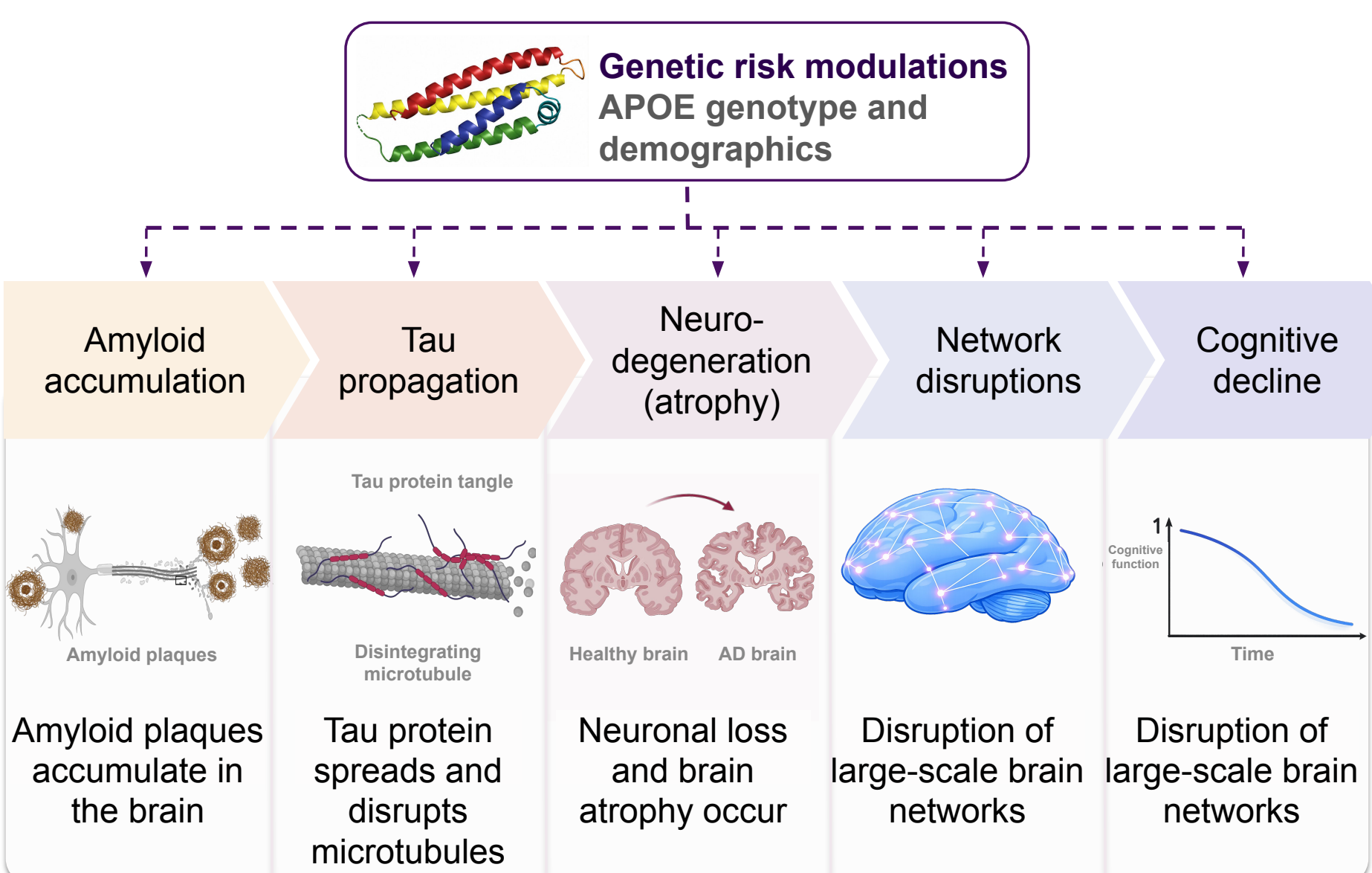
¹ Knight Foundation School of Computing and Information Science, Florida International University, Miami, FL

SaeedLab

1. Background

Alzheimer's disease (AD) is characterized by the progressive effects of protein pathology, regional neurodegeneration, and network-level functional disruption across the brain. Multimodal neuroimaging captures complementary disease aspects: positron emission tomography (PET) quantifies molecular burden, structural magnetic resonance imaging (sMRI) measures regional atrophy, and functional MRI (fMRI) reveals connectivity alterations [1]. While evidence shows that integrating multiple data modalities may result in improved diagnostic performance, most existing approaches combine biomarkers within black-box frameworks with limited explainability. **Objective:** Develop an interpretable graph-based framework that integrates multimodal biomarkers to classify cognitively normal (CN) vs. Alzheimer's disease (AD) and to identify the brain regions and biomarkers that drive predictions.

1. Biological cascade of Alzheimer's Disease



2. Multimodal biomarker representation

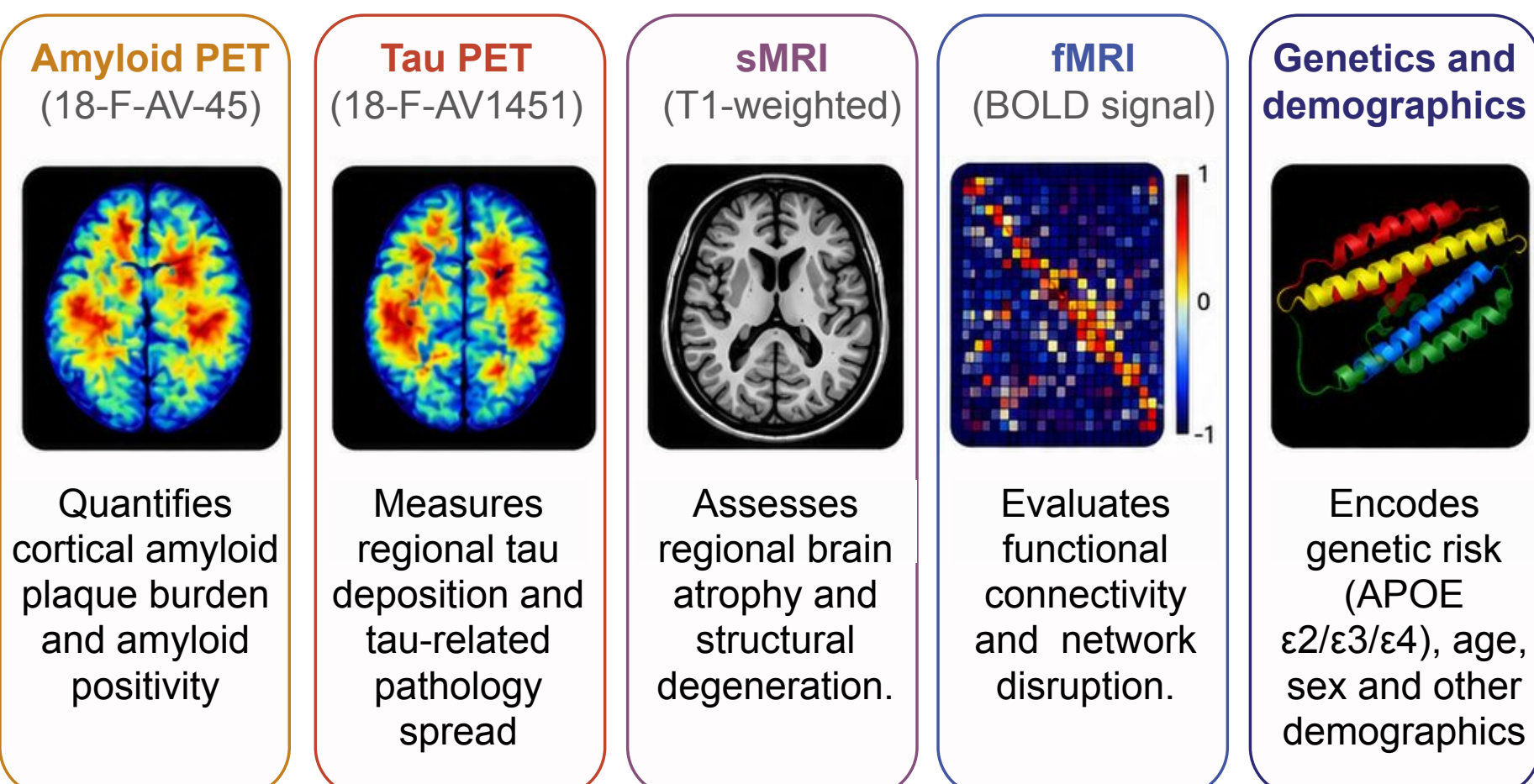


Figure 1: AD Multimodal Biomarker Representation

2. Dataset

560 sessions from the Alzheimer's Disease Neuroimaging Initiative were included, with cognitively normal and AD participants aged 55-90 years (mean age = 73.2). Inclusion required sMRI, fMRI, temporally matched amyloid and tau PET, APOE genotype, and demographic data. Unmatched or quality-control-failed sessions were excluded.

3. Methods

A. Graph Construction

Each brain is modeled as a graph $G = (V, E, X)$. Each node corresponds to one of the 112 brain regions defined by the Desikan-Killiany atlas [2]. Functional connectivity derived from rs-fMRI was used to construct the weighted adjacency matrix $A \in \mathbb{R}^{112 \times 119}$ computed using Fisher Z-transformed Pearson correlation coefficients between regional time series. Each node contains a 19-dimensional multimodal feature vector $X \in \mathbb{R}^{112 \times 19}$ integrating demographic variables, APOE genotype, PET-derived pathology indicators, and sMRI measures. This graph formulation enables the GNN to jointly model regional biological attributes and large-scale network interactions associated with AD.

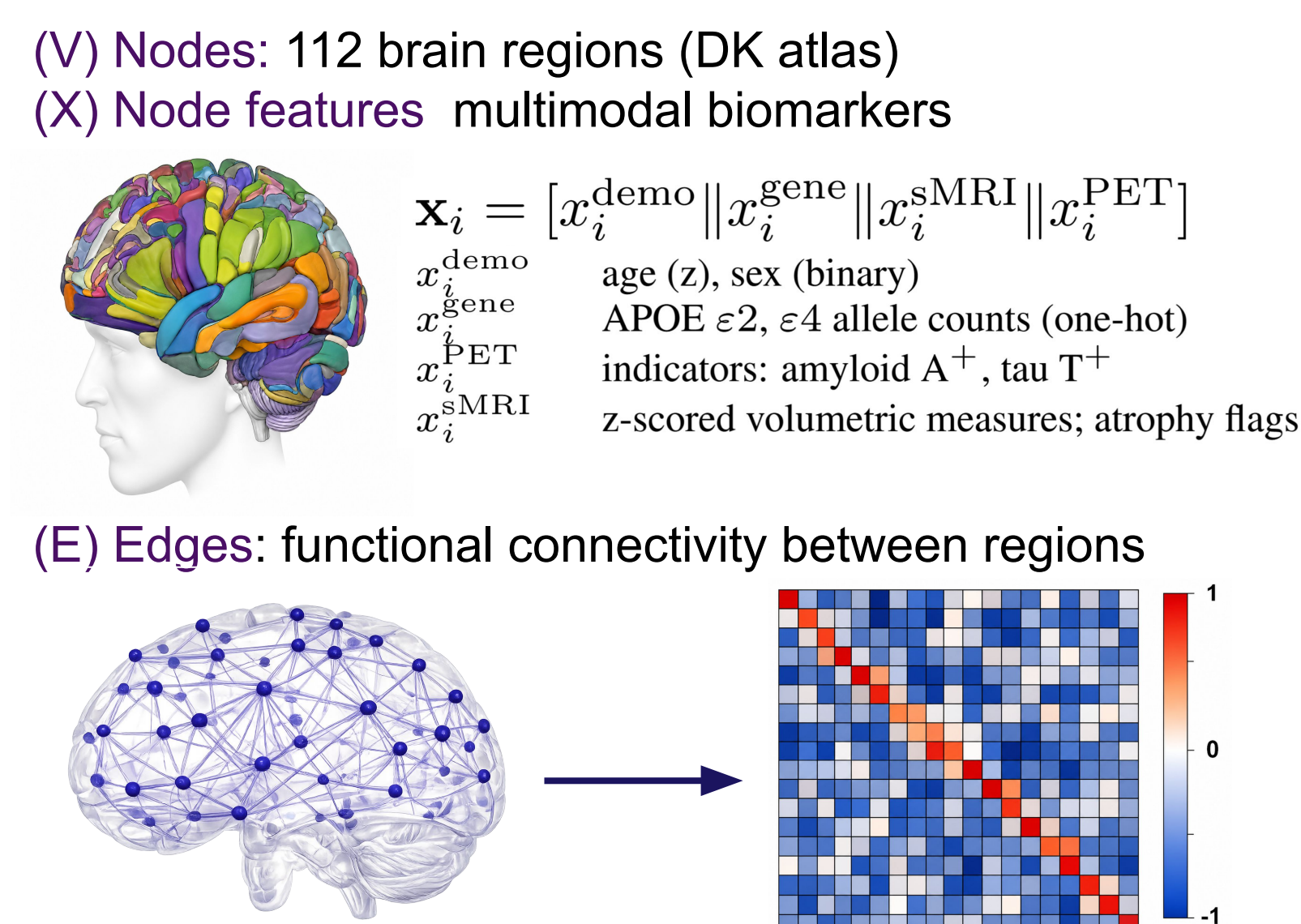


Figure 2: Multimodal graph construction

B. Model architecture

The GCN architecture [3] consisted of three graph convolutional layers with hidden dimensions (160, 160, 80). Each layer was followed by batch normalization, ReLU activation, and dropout. Subject-level graph embeddings were generated using global mean pooling, with seven-network pooling for network-level representations, and passed to a final linear classification head for CN vs. AD prediction.

C. Interpretability Framework

Feature-level attribution

Integrated Gradients (IG) quantify the contribution of each input feature to the model prediction by measuring the change in output along a path from a baseline input (x') to the actual input (x)

$$IG_i(x) = (x_i - x'_i) \int_0^1 \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} d\alpha$$

where F is the GNN prediction, x is the actual input, x' is the baseline input (zeros), and i indexes a feature.

Regional Attribution

Feature attributions are aggregated across features within each ROI to obtain a region-level importance score

$$R_r = \sum_{j=1}^F |IG_{r,j}|$$

where R_r is the importance of region r and F is the number of features per ROI.

Modality Ablation

Feature-set ablation quantifies modality contribution by measuring performance changes after removing multimodal feature groups.

$$\Delta AUC_m = AUC_{\text{full}} - AUC_m$$

where AUC_{full} is the performance of the full multimodal model and AUC_m is the performance when the modality m is removed.

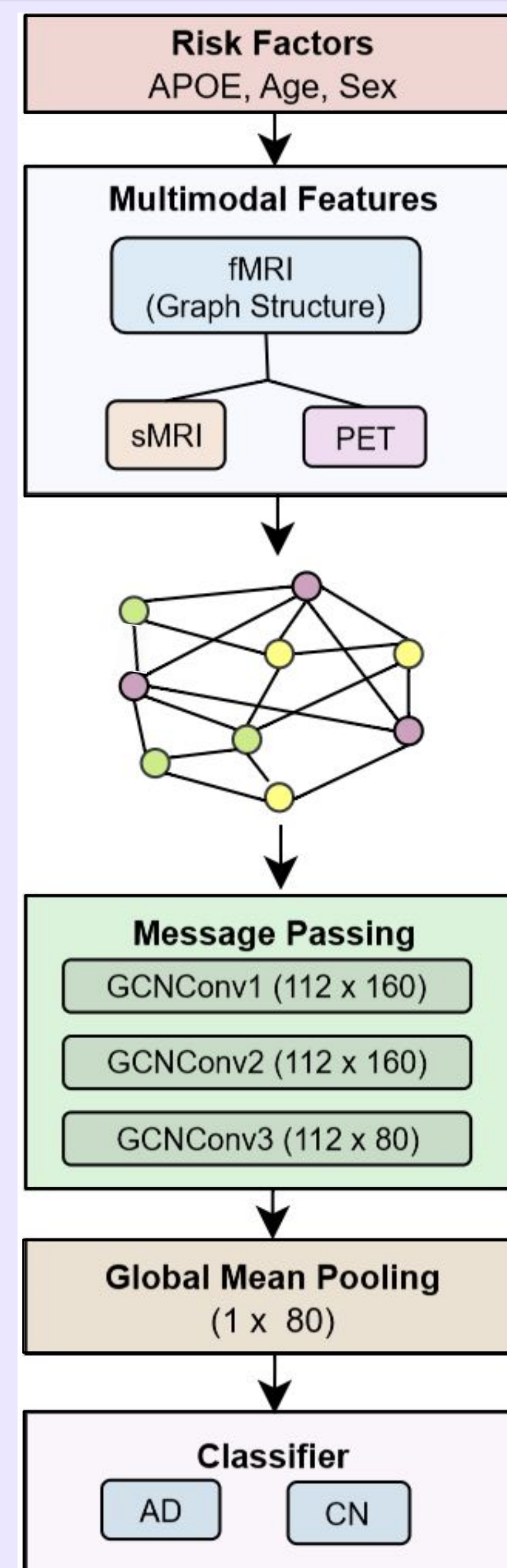


Figure 3: Model architecture

4. Results

Metric	Full multimodal	+PET	+sMRI	Baseline
AUC	0.94 ± 0.04	0.83 ± 0.09	0.82 ± 0.09	0.63 ± 0.04
BalAcc	0.87 ± 0.05	0.79 ± 0.06	0.74 ± 0.11	0.54 ± 0.04
F1-score	0.73 ± 0.09	0.59 ± 0.08	0.51 ± 0.16	0.17 ± 0.11
Sensitivity	0.82 ± 0.12	0.71 ± 0.11	0.63 ± 0.23	0.13 ± 0.10
Specificity	0.92 ± 0.06	0.86 ± 0.05	0.85 ± 0.08	0.95 ± 0.05

Table 1: Feature-set ablation analysis. Values denote mean $\pm$ s.d. across 5-fold cross-validation. All models build upon a baseline configuration using fMRI-derived graph topology and demographic variables (age, sex, APOE). Additional modalities were progressively added to quantify their contribution to classification performance. The full multimodal configuration achieved the highest overall performance, showing the complementary diagnostic value of integrating molecular pathology and structural neurodegeneration biomarkers with functional connectivity representations.

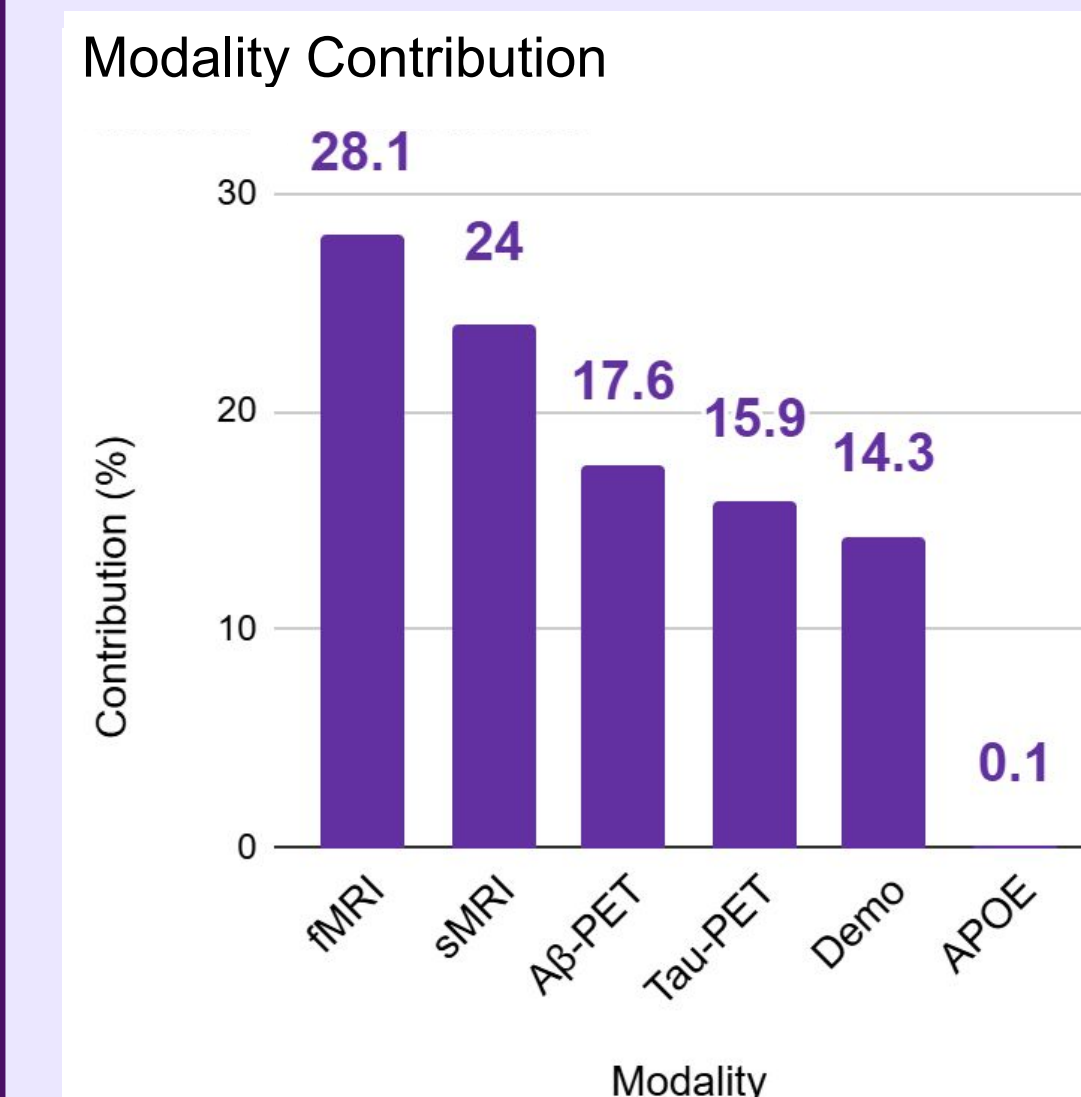


Figure 4: Modality contribution analysis. sMRI and fMRI provided the largest contributions to classification performance, while PET biomarkers further improved diagnostic sensitivity and overall performance. Results support the complementary role of structural neurodegeneration, molecular pathology, and network-level dysfunction in AD representation learning.

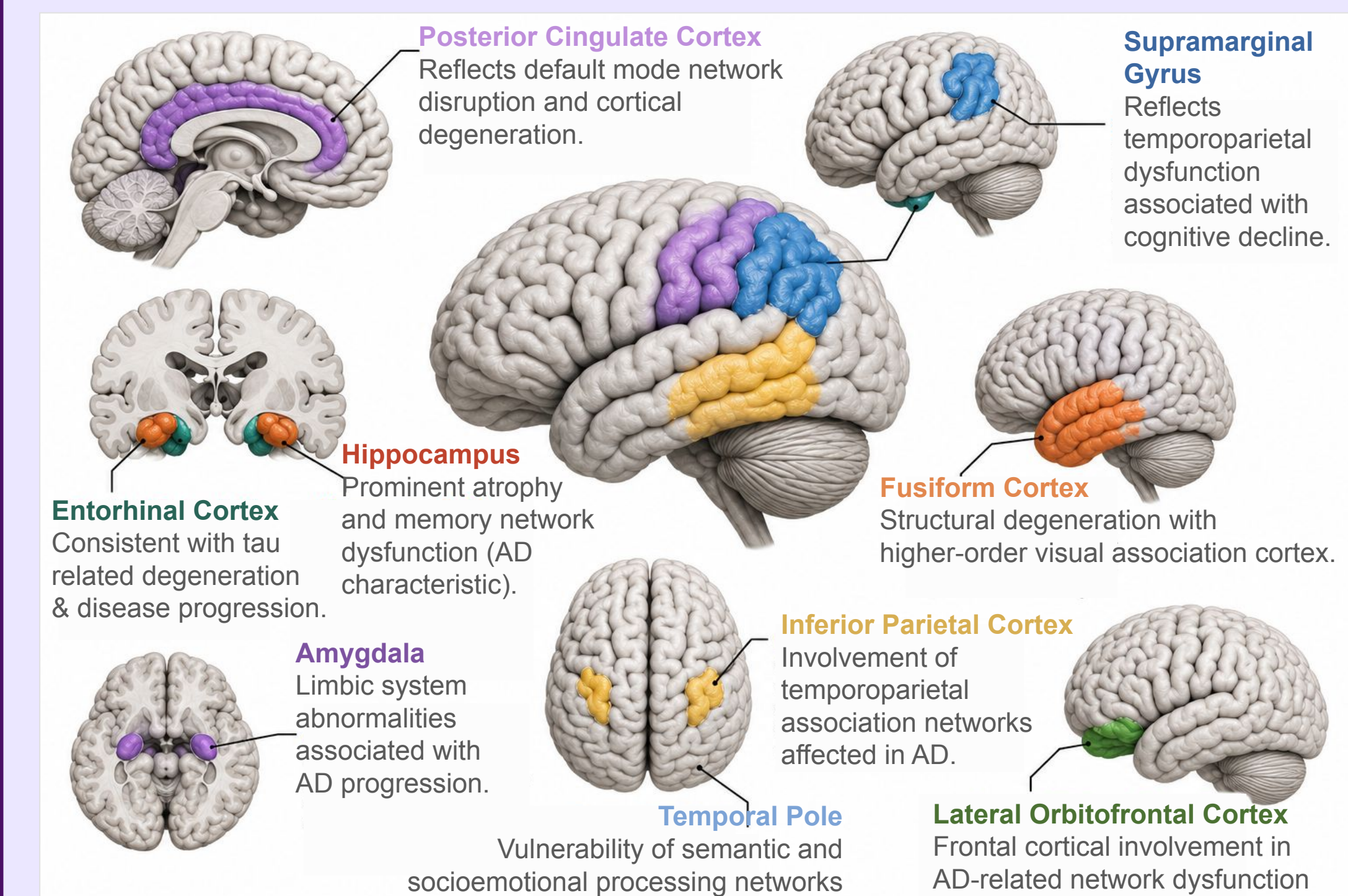


Figure 5: Integrated-gradient regional attribution analysis demonstrating the spatial distribution of multimodal feature importance learned by the proposed graph neural network. Highly contributing regions include hippocampal, entorhinal, posterior cingulate, temporal, parietal, and limbic structures, consistent with known Alzheimer's disease neurodegenerative pathways and default mode network disruption. Attribution values were aggregated across participant-level cross-validation folds.

Acknowledgements: Primary funding for this project is provided by National Science Foundation (NSF) OAC-2312599. The content does not necessarily represent the official views of the NSF. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). **References:** [1] Knopman et al., Nat Rev Dis Primers, 2021. [2] Desikan et al., NeuroImage, 2006. [3] Kipf & Welling, ICLR, 2017.

AIC-045