

# Multimodal Decoupled Dynamic Graph Learning for Brain Disease Diagnosis

## To Reviewer #1, #2 and #3

**Limited theoretical motivation and novelty.** Our main contribution lies in a task-driven disentanglement tailored to modality quality imbalance and its tight integration with similarity-aware dynamic message passing to model disease-stage correlations. In addition, the shared/specific encoder design follows a standard inductive bias in multimodal learning and serves as a structural component of our disentanglement strategy, rather than an independent module.

**Presentation and clarity issues.** In the revised version, we will: (i) explicitly cross-reference Fig. 1 and Fig. 4 to clarify the before–after comparison of modality interactions and highlight how disentanglement alleviates modality neglect; (ii) clarify that modality interaction weights in Fig. 1 are computed via self-attention over modality tokens projected into a shared latent space (Sec. 3.4); (iii) add missing citations for all baseline methods in Tab. 1; (iv) more clearly direct readers to Sec. 4.1.3 for the transductive inference protocol and evaluation setup; (v) revise overly strong claims in Fig. 3 where uncertainty intervals overlap and adopt statistically appropriate wording; (vi) standardize loss-function notations in Sec. 3.2 and simplify ambiguous long sentences.

## To Reviewer #1

**Effectiveness of attending to low-quality modalities.** Since no single modality is sufficiently reliable (all Macro-F1 < 0.8), suppressing low-quality modalities risks losing complementary information. The ablation study (Tab. 2) shows that removing disentanglement consistently degrades performance, supporting the view that regulated utilization of low-quality modalities improves robustness by preserving complementary cross-modal cues under modality imbalance.

**Effectiveness of disentanglement.** Under class and modality imbalance, AUC is less sensitive; instead, the disentanglement module consistently improves weighted F1 by 2.2–5.5% (supplementary material, Tab. 13 and 14), indicating consistent gains in balanced performance and its role as a structural fusion mechanism rather than a standalone booster.

**Missing MM-GTUNets comparison.** MM-GTUNets relies on a reward–penalty graph and a different data-splitting protocol, making direct comparison under our unified cross-validation setting not directly comparable.

## To Reviewer #2

**Lack of statistical rigor.** We conducted statistical significance analysis and report the results in the supplementary material (Tab. 11), where most datasets and metrics show statistically significant improvements, together

with a combined significance level ( $p_{\text{combined}} < 0.05$ ) based on Fisher’s method, further supporting the effectiveness of MDDGL. We will tone down overly strong claims and explicitly reference these results to ensure statistically grounded interpretation.

**Dynamic vs. static message passing.** We evaluate (i) fully static settings, (ii) static message passing with an optimized initial graph, and (iii) layer-wise dynamic message passing with or without adaptive graph learning (Tab. 16). The results show that dynamic message passing consistently yields better performance than static message passing (e.g., (3) vs. (1)), with further gains achieved when combined with adaptive graph optimization ((4)). We will clarify these distinctions and definitions in the revised manuscript.

Table 16. Ablation study on the TADPOLE and ABIDE-5 datasets. (1) GCN; (2) GCN and Adj; (3) Dynamic Message Passing; (4) Dynamic Message Passing and Adj.

Model	Initial Graph	Message Passing	TADPOLE (AD,CN,sMCI)		ABIDE-5 (NC,ASD)	
			ACC(%)	AUC(%)	ACC(%)	AUC(%)
(1)	Static	Static	73.42±4.57	76.54±4.20	66.68±3.51	66.24±3.69
(2)	Dynamic	Static	81.28±4.61	86.72±3.68	82.19±4.57	83.98±4.70
(3)	-	Dynamic	81.46±4.86	90.60±2.89	87.51±2.68	90.55±3.72
(4)	Dynamic	Dynamic	84.13±4.06	91.95±2.96	88.54±3.02	91.30±3.21

## To Reviewer #3

**Small-scale datasets.** Although TADPOLE and ABIDE are small-scale, they exhibit strong modality heterogeneity and quality imbalance, which are precisely the challenges our method targets. Notably, performance gains increase on ABIDE-5 with more modalities, indicating favorable scalability with modality heterogeneity rather than dataset size alone.

**Circular definition of modality quality.** The modality quality score is a task-dependent empirical indicator of relative learning difficulty, computed independently of our model and used only for analysis, not supervision.

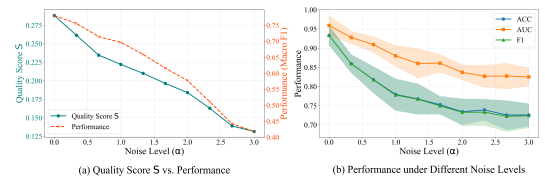


Figure 10. Robustness under progressive modality degradation.

**Robustness and inductive inference.** We further conduct controlled noise-injection experiments that demonstrate robust performance under progressive modality degradation (Fig. 10). Our framework is transductive by design and focuses on cohort-level diagnosis; fully inductive, single-patient inference is beyond the scope of this work and will be explicitly clarified in the revised manuscript.