Latent Pathology–Age Decoupling in Functional Connectivity Graphs for Multi-task Schizophrenia Assessment

Yige Yan[1], Yi Hao Chan[1], Jun Cheng[2], Qian Hui Chew[3], Kang Sim[3], and Jagath C. Rajapakse[1](✉)

[1] School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore
asjagath@ntu.edu.sg
[2] Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
[3] Research Division, Institute of Mental Health, Singapore, Singapore

Abstract. Accurate prediction of clinical symptoms and cognitive deficits from functional connectivity is essential for advancing precision psychiatry in schizophrenia (SZ). However, predicting these clinical and cognitive measures from brain functional connectivity remains challenging. A key barrier is that age-related brain changes substantially overlap with SZ pathology, predisposing models to exploit age-correlated variance rather than genuine pathological signals. At the same time, hard removal of age signals degrades performance on cognitive tasks where aging and pathology interact. To this end, we propose a Latent Pathology-Age Decoupling (LPAD) framework that enforces explicit latent constraints to decouple age-related variations from pathological features. The framework introduces a dual-stream architecture with pathology-age orthogonal decoupling and supervised contrastive regularization to suppress age-driven shortcuts, followed by a clinically-informed gated modulation mechanism that reintroduces age as a subordinate contextual reference. Experiments demonstrate that LPAD achieves performance improvements across all eight clinical symptom and cognitive deficit tasks, with external validation suggesting cross-site generalizability, and linear probe analysis and spatially complementary region-of-interest profiles between the two streams corroborating effective pathology-age decoupling. Code is available at `https://github.com/SCSE-Biomedical-Computing-Group/LPAD`.

Keywords: Functional Connectivity · Graph Neural Network · Multi-task Learning · Pathology-Age Decoupling · Schizophrenia

1 Introduction

Schizophrenia (SZ) exhibits substantial clinical heterogeneity extending beyond fluctuations in positive and negative symptoms, more profoundly reflected in

the complex entanglement between cognitive impairment and normal aging process [20,8]. Accurate prediction of these clinical and cognitive measures from functional connectivity (FC) is central to treatment stratification and prognosis in precision psychiatry. Although FC has been proposed as a potential biomarker for SZ, clinical symptoms, cognitive deficits, and natural brain aging are deeply intertwined, making it difficult to attribute observed functional alterations to disease pathology rather than normal aging [21]. Precisely disentangling disease-specific pathological signatures from age-driven functional changes in FC-based representations alone remains a challenge in precision psychiatry [26,29].

Resting-state FC has been shown to predict symptom trajectories and related outcomes in SZ [11]. Prior work on age-disease interactions operates at the group level [19,5,18] or treats age as one of multiple covariates [23,33], neither of which constrains individual latent representations for FC-based subject-wise prediction. A more direct strategy regresses out age or applies adversarial deconfounding [6,1]; however, such hard removal discards clinically relevant age signals, degrading performance on aging-sensitive cognitive tasks. Recent multi-task methods [30] improve interpretability and prediction performance but do not explicitly address age confounding in the latent space, leaving models susceptible to age-correlated shortcuts [10].

Building on this observation, we propose a Latent Pathology–Age Decoupling (LPAD) framework that enforces explicit latent constraints to decouple age-related variations from pathological factors. Specifically, we introduce a pathology-age orthogonal decoupling strategy with supervised contrastive regularization, employing a dual-stream architecture to reduce shared information between age-related and pathology-specific representations, suppressing age-driven shortcuts in the pathology stream while preserving age as a controlled contextual signal for tasks intrinsically modulated by aging. We further design a Clinically-Informed Gated Modulation (CIGM) mechanism that reintroduces age as a strictly subordinate contextual reference, enabling age to serve as a contextual modulator for pathological representations while limiting its direct contribution to final predictions. In summary, the contributions of this work are as follows:

1. We propose the LPAD framework that addresses the pathology-age confounding problem in FC-based SZ assessment through latent constraints;
2. We introduce a pathology-age orthogonal decoupling strategy with supervised contrastive regularization to enforce latent decoupling, and a CIGM mechanism to selectively reintroduce age as a contextual reference;
3. Experiments across eight clinical symptom and cognitive deficit tasks demonstrate consistent prediction improvements. Furthermore, external validation results suggest cross-site generalizability. Interpretability analyses, including linear probe analysis and spatially complementary region-of-interest profiles, provide converging evidence of pathology-age separation.

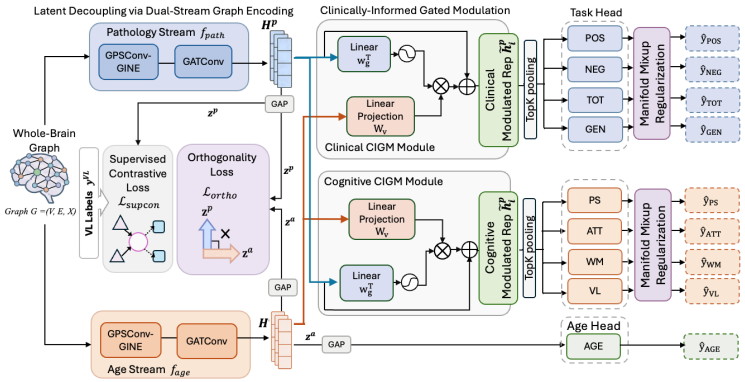


Fig. 1: Overview of the proposed LPAD framework. A dual-stream graph encoder processes a whole-brain graph, extracting node-level pathology $\mathbf{H}^p$ and age $\mathbf{H}^a$ representations. Global average pooling (GAP) aggregates them into graph-level embeddings ($\mathbf{z}^p$, $\mathbf{z}^a$) and computes $\mathcal{L}_{\text{ortho}}$ and $\mathcal{L}_{\text{supcon}}$ for latent decoupling. $\mathbf{H}^p$ and $\mathbf{H}^a$ enter clinical and cognitive CIGM modules, yielding modulated representations $\tilde{\mathbf{h}}_i^p$. Finally, the $\tilde{\mathbf{h}}_i^p$ undergo TopK pooling and manifold mixup for joint task prediction while $\mathbf{z}^a$ supervises the age head.

2 Method

As illustrated in Figure 1, LPAD addresses pathology-age decoupling through a dual-stage design: latent decoupling to suppress age as a confounder, clinically-informed gated modulation to reintroduce age as a subordinate contextual reference. The entire framework is jointly optimized via homoscedastic uncertainty weighting, with an auxiliary age regression task and manifold regularization (detailed target definitions are provided in Sec 3.1).

2.1 Latent Decoupling via Dual-Stream Graph Encoding

Dual-Stream Backbone We represent each subject as a whole-brain graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ with $|\mathcal{V}| = 264$ nodes [24], where $\mathbf{X} \in \mathbb{R}^{264\times 264}$ is the node feature matrix whose i-th row encodes the FC profile of node i. To decouple pathological features from age-related activity, we employ two graph encoders with non-shared weights: a pathology stream f_{path} and an age stream f_{age}. Both adopt a hybrid architecture in which a GPSConv-GINE layer—instantiating the GPS framework [25] with GINE [12,32] (an edge-aware extension of GIN) for local message passing and multi-head self-attention for global context—captures structural patterns via FC-derived edge weights $\mathbf{w} \in \mathbb{R}^{|\mathcal{E}|}$, motivated by the coexistence of local edge-feature patterns and long-range network interactions in FC. A subsequent GATConv [27] layer produces node-level representations $\mathbf{H}^p$ and $\mathbf{H}^a$.

Latent Orthogonality and Supervised Contrastive Regularization Given node-level representations $\mathbf{H}^p$ and $\mathbf{H}^a$, we introduce a dual regularization scheme

comprising latent orthogonality and supervised contrastive regularization to explicitly enforce latent decoupling. Specifically, a sample-wise cosine orthogonality penalty is applied over a mini-batch of size B:

$$\mathcal{L}_{\text{ortho}} = \frac{1}{B}\sum_{b=1}^{B}\left|\frac{\mathbf{z}_b^p \cdot \mathbf{z}_b^a}{\|\mathbf{z}_b^p\|_2\|\mathbf{z}_b^a\|_2}\right|, \tag{1}$$

where $\mathbf{z}_b^p = \frac{1}{|\mathcal{V}|}\sum_i \mathbf{H}_{b,i}^p$ and $\mathbf{z}_b^a$ are obtained analogously via global average pooling over node representations.

This regularization term encourages $\mathbf{z}^p$ to lie in the orthogonal complement of the age embedding space, thereby reducing reliance on age-related shortcuts. While $\mathcal{L}_{\text{ortho}}$ suppresses overlap between pathological and age embeddings, it does not constrain the internal structure of $\mathbf{z}^p$. To encourage the pathology stream to encode disease-relevant features, we apply supervised contrastive regularization using verbal learning (VL) labels as anchors since verbal learning is reliably impaired in schizophrenia and constitutes a core domain in standardized cognitive batteries [15,22]. Let $I = \{1, \ldots, B\}$ denote the index set of a mini-batch of size B. Positive pairs for anchor b are defined as

$$P(b) = \{j \in I \setminus \{b\} : |y_b^{\text{VL}} - y_j^{\text{VL}}| < \epsilon\},$$

where y_b^{VL} is the standardized VL score of subject b and ϵ is a proximity threshold set to 0.1. Inspired by the supervised contrastive learning framework of Khosla et al. [17], we adapt the loss formulation to continuous VL labels as:

$$\mathcal{L}_{\text{supcon}} = -\sum_{b\in I}\frac{1}{|P(b)|}\sum_{j\in P(b)}\log\frac{\exp(\text{sim}(\mathbf{z}_b^p, \mathbf{z}_j^p)/\tau)}{\sum_{k\in I\setminus\{b\}}\exp(\text{sim}(\mathbf{z}_b^p, \mathbf{z}_k^p)/\tau)}, \tag{2}$$

where sim is cosine similarity and τ is a temperature hyperparameter.

2.2 Clinically-Informed Gated Modulation (CIGM)

While latent decoupling suppresses age as a confounder, age remains a relevant contextual factor in SZ assessment, as cognitive and symptomatic presentations are known to vary with normal aging [20,8]. However, age serves as a subordinate reference rather than a primary diagnostic criterion, with pathological evidence remaining central to clinical evaluation. Motivated by this rationale, we design CIGM to instantiate this asymmetry: the pathology stream constitutes the primary representation, while age is selectively reintroduced as a gated contextual reference controlled by a learnable scalar gate.

Gating Implementation and Task Grouping For the i-th node, we assign asymmetric roles to the two streams: the pathological feature $\mathbf{h}_i^p \in \mathbb{R}^{64}$ (the i-th row of $\mathbf{H}^p$) serves as the primary representation, while the age feature $\mathbf{h}_i^a \in \mathbb{R}^{64}$ (the i-th row of $\mathbf{H}^a$) provides contextual information. We first project the age features as

$$\mathbf{v}_i = \mathbf{W}_v\mathbf{h}_i^a, \quad \mathbf{W}_v \in \mathbb{R}^{64\times 64}. \tag{3}$$

A node-level scalar gate is defined as

$$g_i = \sigma(\mathbf{w}_g^T \mathbf{h}_i^p), \quad \mathbf{w}_g \in \mathbb{R}^{64}, \tag{4}$$

where $\mathbf{w}_g$ is a learnable weight and $\sigma(\cdot)$ denotes the sigmoid function. The modulated representation is then obtained as

$$\tilde{\mathbf{h}}_i^p = \mathbf{h}_i^p + g_i \cdot \mathbf{v}_i. \tag{5}$$

To account for task heterogeneity, two separate CIGM modules are instantiated for clinical symptom and cognitive deficit tasks respectively, reflecting the distinct degree to which each task group benefits from age context.

2.3 Joint Optimization with Manifold Regularization

Unified Task Definition To address multi-task weight balancing, we adopt a homoscedastic uncertainty weighting loss [16]. Age prediction is included as an auxiliary regression target within the task loss $\mathcal{L}_{\text{task}}$, providing supervision for $\mathbf{z}^a$ as required by the orthogonality constraint. A learnable log-variance parameter $s_t := \log \sigma_t^2$ is introduced per task $t \in \{1, \ldots, 9\}$, covering the eight targets together with the auxiliary age regression task, where σ_t^2 denotes the task-specific homoscedastic noise variance. The joint regression loss is formulated as:

$$\mathcal{L}_{\text{task}} = \sum_{t=1}^{9} \left(e^{-s_t} \mathcal{L}_{\text{RMSE}}^{(t)}(\hat{y}_t, y_t) + s_t \right), \tag{6}$$

where $\mathcal{L}_{\text{RMSE}}^{(t)}(\hat{y}_t, y_t) = \sqrt{\frac{1}{B}\sum_{b=1}^{B}(\hat{y}_{t,b} - y_{t,b})^2}$ is the root mean square error for task t over a mini-batch of size B. Node-level representations from each stream are aggregated via TopK pooling [9] to obtain fixed-dimensional graph-level features $\mathbf{u} \in \mathbb{R}^D$. To mitigate overfitting, manifold mixup regularization [28] is applied to post-pooling features $\mathbf{u}$. Given a randomly sampled pair $(\mathbf{u}_b, y_b)$ and $(\mathbf{u}_{b'}, y_{b'})$, virtual training samples are constructed as $\hat{\mathbf{u}} = \alpha \mathbf{u}_b + (1-\alpha)\mathbf{u}_{b'}$ and $\hat{y} = \alpha y_b + (1-\alpha) y_{b'}$, where $\alpha \sim \mathcal{U}(0, 1)$.

Total Objective Function The total objective integrates task regression and latent constraints:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{task}} + \lambda_1 \cdot \mathcal{L}_{\text{ortho}} + \lambda_2 \cdot \mathcal{L}_{\text{supcon}}, \tag{7}$$

where $\lambda_1 = 0.05$ and $\lambda_2 = 0.1$ are scalar coefficients determined empirically.

3 Experiments and Results

3.1 Datasets and Implementation

Two datasets were used for the prediction of SZ clinical symptoms and cognitive deficits. The Center for Biomedical Research Excellence (COBRE) dataset [2]

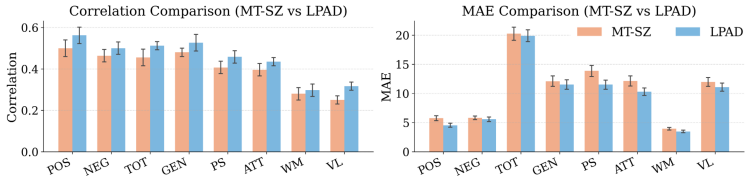


Fig. 2: Comparison of prediction performance between LPAD and MT-SZ on the external validation dataset (cognitive scores: BACS raw scores).

comprised 135 participants (60 SZ patients and 75 healthy controls; age 37.3±8.1 years, range 19–66), and a local clinical dataset comprised 35 participants (15 SZ patients and 20 healthy controls; age 36.6 ± 5.4 years, range 23–58). Clinical severity was assessed in both datasets using the Positive and Negative Syndrome Scale (PANSS) [14], yielding four prediction targets: positive symptoms (POS), negative symptoms (NEG), general psychopathology (GEN), and total score (TOT), with theoretical score ranges of 7–49, 7–49, 16–112, and 30–210, respectively. Cognitive function was evaluated using the MATRICS Consensus Cognitive Battery [22] (COBRE, T-scores) and the Brief Assessment of Cognition in Schizophrenia (BACS) [15] (local dataset, raw scores), covering four domains: processing speed (PS), attention (ATT), working memory (WM), and verbal learning (VL). Resting-state fMRI data were preprocessed using fMRIPrep [7], and FC matrices were constructed from 264 regions of interest (ROIs) defined by the Power atlas [24].

All experiments were conducted on an NVIDIA A100 GPU. The Adam optimizer was used with a learning rate of 0.0005, a batch size of 20, and 60 training epochs. Performance was evaluated via five-fold cross-validation on COBRE over 10 independent runs, with the local clinical dataset held out as an independent external validation cohort; the two datasets are never merged owing to differing acquisition protocols and cognitive batteries (MCCB vs. BACS). Prediction accuracy was assessed by Pearson correlation and mean absolute error (MAE).

3.2 Comparison with Competing Methods

As shown in Table 1, LPAD outperforms all competing methods across all eight tasks, with improvements over the strongest baseline MT-SZ showing statistical significance on 5 of 8 tasks (TOT, GEN, POS, ATT, VL; 1000 permutations, FDR-corrected $p < 0.05$). Following standard deconfounding practice, we construct two hard-removal baselines: MT-SZ$_{\text{ar}}$, removing age via linear regression prior to training [4], and MT-SZ$_{\text{adv}}$, applying adversarial deconfounding via a gradient reversal layer [6,1]. The former degrades performance broadly across most tasks, confirming that pre-training age regression indiscriminately removes both confounding and clinically relevant variance; the latter yields gains on most clinical symptom tasks and on processing speed and attention, but degrades on working memory and verbal learning, the two cognitive tasks most sensitive to aging–pathology interactions. LPAD instead decouples age from pathological

Table 1: Performance comparison with competing methods. MT, multi-task; AT, auxiliary task; -ar/-adv, age-regression and adversarial deconfounding variants of MT-SZ.

Metric	Method	Clinical Symptom Tasks				Cognitive Deficit Tasks			
		POS	NEG	TOT	GEN	PS	ATT	WM	VL
Pearson Corr (↑)	GAT-MT [27]	0.315±0.03	0.350±0.02	0.355±0.04	0.399±0.04	0.407±0.02	0.404±0.02	0.208±0.03	0.201±0.02
	GPS-MT [25]	0.418±0.03	0.408±0.02	0.367±0.04	0.417±0.03	0.397±0.05	0.399±0.03	0.226±0.03	0.211±0.03
	BNT-AT [13]	0.405±0.03	0.456±0.03	0.384±0.03	0.405±0.04	0.393±0.03	0.415±0.04	0.235±0.02	0.227±0.03
	ML-At [31]	0.402±0.06	0.462±0.03	0.439±0.03	0.448±0.03	0.406±0.04	0.398±0.03	0.258±0.11	0.233±0.07
	MT-SZ [30]	0.449±0.03	0.495±0.02	0.468±0.04	0.456±0.03	0.424±0.04	0.402±0.04	0.296±0.04	0.254±0.02
	MT-SZ$_{ar}$	0.463±0.04	0.462±0.02	0.430±0.03	0.414±0.02	0.402±0.04	0.382±0.03	0.269±0.03	0.211±0.03
	MT-SZ$_{adv}$	0.472±0.04	0.490±0.02	0.493±0.07	0.473±0.05	0.439±0.05	0.431±0.04	0.277±0.05	0.230±0.05
	LPAD	**0.490**±0.02	**0.516**±0.02	**0.523**±0.03	**0.499**±0.03	**0.460**±0.04	**0.448**±0.03	**0.319**±0.03	**0.286**±0.02
MAE (↓)	GAT-MT [27]	8.605±0.51	8.853±0.49	23.072±1.12	16.132±0.91	13.122±0.91	11.917±0.73	11.825±0.81	8.758±0.46
	GPS-MT [25]	8.415±0.47	6.701±0.39	22.780±1.23	13.849±0.95	12.501±0.91	11.945±0.72	11.783±0.77	8.636±0.44
	BNT-AT [13]	8.071±0.40	6.523±0.36	21.755±1.12	13.137±0.91	12.664±0.89	11.666±0.70	11.502±0.70	8.565±0.40
	ML-At [31]	8.044±0.41	6.449±0.40	21.080±1.06	12.601±0.84	12.809±0.90	11.130±0.85	11.542±0.72	8.865±0.37
	MT-SZ [30]	6.490±0.33	6.285±0.31	20.060±1.12	11.826±0.80	10.675±0.89	10.568±0.72	11.372±0.76	8.686±0.20
	MT-SZ$_{ar}$	6.390±0.39	6.496±0.42	20.943±1.01	12.012±0.97	11.251±0.73	11.009±0.65	11.209±0.82	8.879±0.18
	MT-SZ$_{adv}$	6.429±0.42	6.118±0.35	20.001±1.31	11.412±0.97	10.554±0.73	9.706±0.60	11.404±0.61	8.851±0.22
	LPAD	**6.115**±0.36	**5.532**±0.34	**19.798**±1.00	**11.244**±0.71	**9.899**±0.81	**9.239**±0.60	**11.021**±0.62	**7.835**±0.21

representations while selectively reintroducing it via CIGM. Figure 2 shows that the relative performance advantage of LPAD over the strongest baseline MT-SZ is preserved on a local clinical dataset, providing preliminary evidence that the method ranking generalizes across sites.

3.3 Ablation Study

Component-wise Ablation Table 2 confirms the complementary contribution of each component. Removing $\mathcal{L}_{\text{ortho}}$ causes broad Pearson correlation degradation across all eight tasks, consistent with age-related variance leaking into pathological representations. Removing $\mathcal{L}_{\text{supcon}}$ produces the most pronounced correlation drop on verbal learning, with comparatively smaller decreases on the remaining tasks. Removing CIGM decreases performance across both task groups. Working memory incurs the largest absolute cognitive decrease, consistent with the gating design's motivation to preserve aging-sensitive signals. Removing Mixup produces an overall correlation decrease.

Sensitivity Analysis Table 3 reports mean Pearson correlations across eight tasks under varying loss coefficients and contrastive anchor selection. The optimal configuration ($\lambda_1 = 0.05$, $\lambda_2 = 0.1$) achieves a mean correlation of 0.4426, with performance remaining stable across a moderate range of values. Verbal learning yields the best anchor performance, consistent with its standing as a core and reliably impaired cognitive domain in schizophrenia [15].

3.4 Interpretability Analysis

Linear Probe Analysis A RidgeCV regressor on frozen representations confirms age is recoverable from $\mathbf{z}^a$ ($R^2 = 0.477$) but not $\mathbf{z}^p$ ($R^2 = -0.176$), while

Table 2: Ablation study of individual model components.

Metric	Method	POS	NEG	TOT	GEN	PS	ATT	WM	VL
Pearson Corr (↑)	w/o $\mathcal{L}_{\text{ortho}}$	0.473±0.03	0.490±0.03	0.501±0.02	0.484±0.03	0.431±0.03	0.437±0.04	0.296±0.02	0.253±0.03
	w/o $\mathcal{L}_{\text{supcon}}$	0.489±0.02	0.501±0.02	0.506±0.03	0.491±0.04	0.443±0.02	0.440±0.02	0.306±0.03	0.251±0.04
	w/o CIGM	0.460±0.02	0.489±0.03	0.495±0.02	0.475±0.02	0.441±0.02	0.440±0.04	0.296±0.04	0.276±0.03
	w/o Mixup	0.488±0.03	0.497±0.02	0.478±0.02	0.471±0.02	0.449±0.02	0.443±0.02	0.279±0.03	0.266±0.04
	LPAD	**0.490**±0.02	**0.516**±0.02	**0.523**±0.03	**0.499**±0.03	**0.460**±0.04	**0.448**±0.03	**0.319**±0.03	**0.286**±0.02
MAE (↓)	w/o $\mathcal{L}_{\text{ortho}}$	6.539±0.37	6.105±0.26	20.401±0.91	11.947±0.47	10.464±0.38	9.954±0.38	11.344±0.36	8.202±0.28
	w/o $\mathcal{L}_{\text{supcon}}$	6.202±0.39	5.725±0.34	20.009±1.02	11.608±0.46	10.217±0.61	9.866±0.41	11.137±0.80	7.992±0.24
	w/o CIGM	6.597±0.20	6.223±0.31	20.357±0.89	11.371±0.44	10.223±0.46	10.375±0.58	11.645±0.44	8.051±0.24
	w/o Mixup	6.245±0.20	5.836±0.31	20.990±0.96	12.140±0.59	10.598±0.28	9.838±0.47	12.118±0.31	7.972±0.24
	LPAD	**6.115**±0.36	**5.532**±0.34	**19.798**±1.00	**11.244**±0.71	**9.899**±0.81	**9.239**±0.60	**11.021**±0.62	**7.835**±0.21

Table 3: Sensitivity analysis of loss coefficients and contrastive anchor selection.

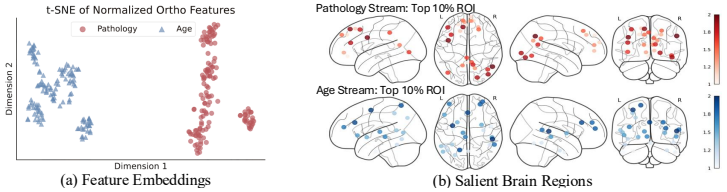
λ_1	λ_2			Anchor	Mean Corr
	0.05	0.1	0.2		
0.01	0.4262±0.02	0.4203±0.02	0.4361±0.03	POS	0.4329±0.02
0.05	0.4288±0.04	**0.4426±0.02**	0.4358±0.01	GEN	0.4382±0.01
0.10	0.4209±0.02	0.4303±0.04	0.4266±0.02	VL	**0.4426±0.02**

$\mathbf{z}^p$ remains predictive of PANSS-POS ($R^2 = 0.091$), confirming that orthogonal decoupling suppresses age-related variance without collapsing pathological structure. This is corroborated by Figure 3(a).

Discriminative ROIs for Prediction Node-importance scores from the TopK pooling layers are mapped onto the Power-264 atlas (top 10%, $K = 26$), yielding zero overlap between streams (Jaccard index = 0.000). Pathology-stream nodes concentrate in the Default Mode Network (DMN), consistent with established DMN disruption in SZ [3], while age-stream nodes are anchored to subcortical and temporal structures consistent with normal aging trajectories. The spatially complementary profiles further support effective pathology-age decoupling, as shown in Figure 3(b).

4 Conclusion

Predicting SZ clinical symptoms and cognitive deficits from FC is confounded by the overlap between age-related neural changes and disease pathology. LPAD addresses this by suppressing age as a confounder via orthogonal decoupling and supervised contrastive regularization, while reintroducing it as a subordinate contextual reference through clinically-informed gated modulation, yielding performance improvements across all eight prediction tasks. External validation on an independent clinical dataset further suggests cross-site generalizability, and interpretability analyses corroborate effective pathology-age separation. We acknowledge two limitations: the cosine orthogonality enforces only linear independence, leaving non-linear age leakage in $\mathbf{z}^p$ and residual pathology in $\mathbf{z}^a$ unaddressed; and sample sizes ($n = 135$ and $n = 35$) remain modest. Future work



(a) Feature Embeddings (b) Salient Brain Regions

Fig. 3: (a) t-SNE of normalized orthogonal features from pathology and age streams, showing clear cluster separation; (b) Top-10% discriminative ROIs mapped onto the Power-264 atlas for each stream.

will explore kernel-based independence criteria, larger longitudinal cohorts, and multimodal extensions.

Acknowledgments This research is partially funded by a Tier-1 grant RG15/24 of the Ministry of Education, Singapore.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Adeli, E., Zhao, Q., Pfefferbaum, A., Sullivan, E.V., Fei-Fei, L., Niebles, J.C., Pohl, K.M.: Representation learning with statistical independence to mitigate bias. In: WACV (2021)
2. Aine, C.J., Bockholt, H.J., Bustillo, J.R., et al.: Multimodal neuroimaging in schizophrenia: Description and dissemination. Neuroinformatics **15**(4), 343–364 (2017). https://doi.org/10.1007/s12021-017-9338-9
3. Anticevic, A., Cole, M.W., Murray, J.D., Corlett, P.R., Wang, X.J., Krystal, J.H.: The role of default network deactivation in cognition and disease. Trends Cogn. Sci. **16**(12), 584–592 (2012). https://doi.org/10.1016/j.tics.2012.10.008
4. Chyzhyk, D., Varoquaux, G., Thirion, B., Milham, M.: How to remove or control confounds in predictive models, with applications to brain biomarkers. Gigascience **11**, giac014 (2022). https://doi.org/10.1093/gigascience/giac014
5. Constantinides, C., Han, L.K., Alloza, C., et al.: Brain ageing in schizophrenia: evidence from 26 international cohorts via the enigma schizophrenia consortium. Mol. Psychiatry **28**(3), 1201–1209 (2023)
6. Dinsdale, N.K., Jenkinson, M., Namburete, A.I.: Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. NeuroImage **228**, 117689 (2021). https://doi.org/10.1016/j.neuroimage.2020.117689
7. Esteban, O., Markiewicz, C.J., Blair, R.W., et al.: fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods **16**(1), 111–116 (2019). https://doi.org/10.1038/s41592-018-0235-4
8. Fett, A.K.J., Reichenberg, A., Velthorst, E.: Lifespan evolution of neurocognitive impairment in schizophrenia: A narrative review. Schizophr. Res. Cogn. **28**, 100237 (2022). https://doi.org/10.1016/j.scog.2022.100237

9. Gao, H., Ji, S.: Graph u-nets. In: Chaudhuri, K., Salakhutdinov, R. (eds.) Proceedings of the 36th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 97, pp. 2083–2092. PMLR (2019), https://proceedings.mlr.press/v97/gao19a.html
10. Geirhos, R., Jacobsen, J.H., Michaelis, C., et al.: Shortcut learning in deep neural networks. Nat. Mach. Intell. **2**(11), 665–673 (2020). https://doi.org/10.1038/s42256-020-00257-z
11. Gifford, G., Crossley, N., Kempton, M.J., Morgan, S., Dazzan, P., Young, J., McGuire, P.: Resting state fMRI based multilayer network configuration in patients with schizophrenia. Neuroimage Clin. **25**, 102169 (2020). https://doi.org/10.1016/j.nicl.2020.102169
12. Hu, W., Liu, B., Gomes, J., et al.: Strategies for pre-training graph neural networks. In: ICLR (2020), https://openreview.net/forum?id=HJlWWJSFDH
13. Kan, X., Cui, H., Han, K., Guo, Y., Yang, C.: Multi-task learning for brain network analysis in the ABCD study. In: BHI (2024)
14. Kay, S.R., Fiszbein, A., Opler, L.A.: The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr. Bull. **13**(2), 261–276 (1987). https://doi.org/10.1093/schbul/13.2.261
15. Keefe, R.S.E., Goldberg, T.E., Harvey, P.D., Gold, J.M., Poe, M.P., Coughenour, L.: The brief assessment of cognition in schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. Schizophr. Res. **68**(2–3), 283–297 (2004). https://doi.org/10.1016/j.schres.2003.09.011
16. Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: CVPR. pp. 7482–7491 (2018)
17. Khosla, P., Teterwak, P., Wang, C., et al.: Supervised contrastive learning. In: NeurIPS. vol. 33, pp. 18661–18673 (2020)
18. Lee, W.H., Antoniades, M., Schnack, H.G., Kahn, R.S., Frangou, S.: Brain age prediction in schizophrenia: does the choice of machine learning algorithm matter? Psychiatry Res. Neuroimaging **310**, 111270 (2021)
19. Liu, X., Yang, H., Becker, B., et al.: Disentangling age-and disease-related alterations in schizophrenia brain network using structural equation modeling: A graph theoretical study based on minimum spanning tree. Hum. Brain Mapp. **42**(10), 3023–3041 (2021)
20. McCutcheon, R.A., Keefe, R.S.E., McGuire, P.K.: Cognitive impairment in schizophrenia: aetiology, pathophysiology, and treatment. Mol. Psychiatry **28**(5), 1902–1918 (2023). https://doi.org/10.1038/s41380-023-01949-9
21. Mehta, U.M., Ibrahim, F.A., Sharma, M.S., et al.: Resting-state functional connectivity predictors of treatment response in schizophrenia: A systematic review and meta-analysis. Schizophr. Res. **237**, 153–165 (2021). https://doi.org/10.1016/j.schres.2021.09.004
22. Nuechterlein, K.H., Green, M.F., Kern, R.S., et al.: The MATRICS consensus cognitive battery, part 1: test selection, reliability, and validity. Am. J. Psychiatry **165**(2), 203–213 (2008). https://doi.org/10.1176/appi.ajp.2007.07010042
23. Ouyang, J., Zhao, Q., Adeli, E., Zaharchuk, G., Pohl, K.M.: Disentangling normal aging from severity of disease via weak supervision on longitudinal mri. IEEE Trans. Med. Imaging **41**(10), 2558–2569 (2022)
24. Power, J.D., Cohen, A.L., Nelson, S.M., et al.: Functional network organization of the human brain. Neuron **72**(4), 665–678 (2011). https://doi.org/10.1016/j.neuron.2011.09.006
25. Rampášek, L., Galkin, M., Dwivedi, V.P., Luu, A.T., Wolf, G., Beaini, D.: Recipe for a General, Powerful, Scalable Graph Transformer. NeurIPS **35** (2022)

26. Sheffield, J.M., Rogers, B.P., Blackford, J.U., Heckers, S., Woodward, N.D.: Accelerated aging of functional brain networks supporting cognitive function in psychotic disorders. Biol. Psychiatry **86**(3), 240–248 (2019). https://doi.org/10.1016/j.biopsych.2018.12.016
27. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph Attention Networks. ICLR (2018), https://openreview.net/forum?id=rJXMpikCZ
28. Verma, V., Lamb, A., Beckham, C., et al.: Manifold mixup: Better representations by interpolating hidden states. In: ICML. Proceedings of Machine Learning Research, vol. 97, pp. 6438–6447. PMLR (2019)
29. Wang, X., Chang, Z., Wang, R.: Opposite effects of positive and negative symptoms on resting-state brain networks in schizophrenia. Commun. Biol. **6**(1), 262–279 (2023). https://doi.org/10.1038/s42003-023-04637-0
30. Xia, J., Chan, Y.H., Girish, D., et al.: Disentangling shared and unique brain functional changes associated with clinical severity and cognitive phenotypes in schizophrenia via deep learning. Commun. Biol. **8**(1), 1215 (2025)
31. Xia, J., Chen, N., Qiu, A.: Multi-level and joint attention networks on brain functional connectivity for cross-cognitive prediction. Med. Image Anal. **90**, 102921 (2023). https://doi.org/10.1016/j.media.2023.102921
32. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? In: ICLR (2019), https://openreview.net/forum?id=ryGs6iA5Km
33. Zhang, S., Jiang, Z., Shen, X., et al.: Graph disentanglement learning for fmri analysis: Decoupling disease, covariates, and individual variability. In: MICCAI (2025)