# Task-Aware Functional Hypergraph Learning for Brain State Classification via Information Bottleneck

**Mingyang Xia**[1,2], **Yonggang Shi**[1,3†]

[1]Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, USC
[2]Ming Hsieh Department of Electrical and Computer Engineering, USC
[3]Alfred E. Mann Department of Biomedical Engineering, USC

†correspondence to yshi@loni.usc.edu

## Abstract

Functional connectivity networks (FCNs) are widely used in fMRI-based brain analysis. While most existing studies represent FCNs using graphs, traditional graph structures primarily focus on pairwise connections, overlooking higher-order relationships. Additionally, many methods construct graphs or hypergraphs independently of downstream tasks, which can result in suboptimal representations that fail to capture task-relevant structures. To address these limitations, we propose a novel approach that integrates task-specific information directly into the hypergraph construction process. Our method employs a learnable groupwise mask to construct a groupwise hypergraph structure across all subjects. To retain task-related brain regions and filter out irrelevant ones, we introduce an information bottleneck constraint to optimize our framework. Furthermore, to capture personalized information, we design a hypergraph multi-head attention mechanism that learns personalized hypergraph attention matrices. We apply our model to the ADNI-3 dataset and ABIDE dataset to classify brain states associated with Alzheimer's disease and autism. Our method outperforms competing approaches, achieving at least a $2.2\%$ improvement in accuracy.

**Keywords:** Hypergraph; Information Bottleneck; Brain Network.

## 1 Introduction

Functional magnetic resonance imaging (fMRI) has become a crucial tool for studying brain activity and understanding neurological disorders such as Alzheimer's disease and autism. A key challenge in fMRI-based analysis is accurately modeling the complex interactions between brain regions to classify different brain states. Traditional approaches often rely on graph-based methods Zhou et al. [2020] to represent functional connectivity. In graph representation, each node corresponds to a region of interest (ROI), and the edges reflect functional similarity or dependence between regions. The edges in the graph are typically computed using statistical measures such as Pearson correlation (PC) Song et al. [2008], Baldassarre et al. [2012], Geerligs et al. [2016], Lin et al. [2018], partial correlation Marrelec et al. [2006], coherence Sun et al. [2004], Bowyer [2016], and regularized techniques such as Lasso and Elastic Net Eavani et al. [2015], Zille et al. [2017]. However, graph-based methods focus on pairwise connections, which may fail to fully capture the higher-order relationships within brain networks. For example, multiple ROIs jointly participate in the default mode network (DMN), the sensory network, and the motor network Power et al. [2011].

A hypergraph, as a generalization of a graph, can effectively capture higher-order relationships by allowing any number of nodes to connect to an edge, known as a hyperedge. Therefore, hypergraph

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Data on the Brain & Mind.

may provide a more natural and flexible representation for modeling the high-order interactions inherent in functional brain networks. Existing hypergraph methods have achieved success in the diagnosis of neurological disease in recent years. For example, Ji et al. Ji et al. [2022] proposed FC-HAT, which demonstrates improved performance in classifying subjects with autism spectrum disorder. Similarly, Liu et al. Liu et al. [2023] constructed hypergraphs using multi-template based on sparse reconstruction and applied them to the classification of both Alzheimer's disease and autism. However, most existing hypergraph studies Xiao et al. [2019], Wang et al. [2023], Teng et al. [2024] first construct a hypergraph and subsequently utilize it for downstream tasks, such as analysis based on the hypergraph neural network. However, since hypergraph construction and downstream tasks are not integrated end-to-end, the hypergraph structure remains fixed and is not optimized during training. HYBRID Qiu et al. [2023] provides valuable information on constructing high-order relationships while selecting important regions based on the information bottleneck. However, it does not explicitly treat the structure as a hypergraph, nor does it leverage the properties of hypergraphs.

In this study, we propose an end-to-end framework that jointly constructs the hypergraph and optimizes the downstream task. To mitigate the impact of fMRI signal fluctuations, we design a groupwise task-related hypergraph structure while preserving personalized information through a multi-head hypergraph attention mechanism. Inspired by previous work, we further optimize our framework using the information bottleneck principle to enhance regions selection and representation learning. We validate our model in the Alzheimer's Disease Neuroimaging Initiative 3 (ADNI-3) Weiner et al. [2017] datasets and the Autism Brain Imaging Data Exchange (ABIDE) datasets Craddock et al. [2013] for the classification of brain states with/without neurological disease. Experimental results demonstrate that our method consistently outperforms competing approaches in all subsets. In addition, by incorporating tau PET data, we show that our model effectively identifies task-relevant brain regions during hypergraph construction, highlighting its potential to advance neuroimaging-based disease diagnosis.

## 2 Methods

To effectively characterize task-related brain regions and their underlying connectivity patterns, we propose a method that identifies the most informative regions and constructs their corresponding hypergraph representation. This structured representation serves as a foundation for downstream analyses.

Formally, given a dataset of $M$ subjects, each subject is represented as a tuple $(X_i, Y_i)$, where $X_i \in \mathbb{R}^{N\times P}$ denotes the input feature matrix. Here, $N$ represents the number of predefined parcellated brain regions, and $P$ corresponds to the feature dimension of each region. The label $Y$ encodes subject-specific attributes, such as brain states. Assume $\hat{H}$ is the underlying hypergraph representation, which could reflect brain states, and $E$ represents the number of hyperedges. We could view $(X, \hat{H}, Y)$ as a Markov chain, $X \to \hat{H} \to Y$. As in the information bottleneck principle, our goal is to find $\hat{H}$ that:

$$\arg\max_{\hat{H}} \; I(\hat{H}; Y) - \beta\, I(\hat{H}; X), \tag{1}$$

where $\beta$ is a trade-off parameter that controls the relative weight of the compression term $I(\hat{H}; X)$ against the predictive term $I(\hat{H}; Y)$.

The framework comprises four stages: (1) a learnable mask constructs a groupwise hypergraph structure, enabling the selection of critical regions and intra-connections; (2) a multi-head attention mechanism is applied to the hypergraph connections to quantify the importance of vertices within hyperedges; (3) learnable node features, shared across all subjects and processed through hypergraph convolution layers guided by a hypergraph attention matrix, extract discriminative features; (4) the extracted features are concatenated and subsequently passed through a two-layer MLP for prediction.

### 2.1 Hypergraph definition

As an extended definition of a graph, a hypergraph $\mathcal{H}$ can be defined as a structure in which an edge (a hyperedge) connects multiple nodes simultaneously. $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, v_2, ..., v_N\}$ is the vertex set and $\mathcal{E} = \{e_1, e_2, ..., e_E\}$ is the hyperedge set. A hypergraph incidence matrix

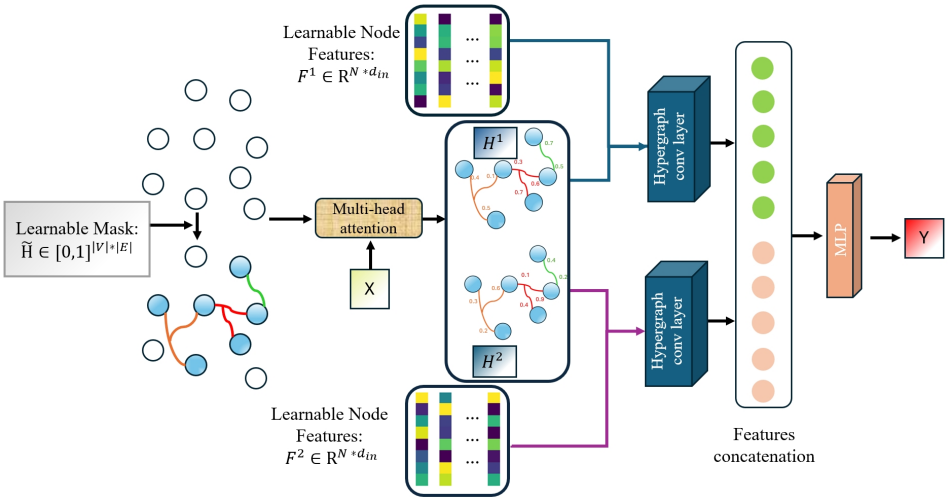



Figure 1: An illustration of the proposed method with $k = 2$ hypergraph attention heads, where the number of hyperedges $E = 3$ and each hyperedge is depicted in a different color. (1) The hypergraph structure is generated using a learnable probability mask. (2) After obtaining the hypergraph, a multi-head attention module produces $k$ hypergraph attention matrices. (3) Using learnable shared node features and the hypergraph attention matrices, a hypergraph convolution layer generates node features. (4) The resulting features are concatenated and then fed into an MLP to predict brain states.

$H \in \{0,1\}^{N\times E}$ represents membership:

$$H_{ij} = \begin{cases} 1, & \text{if } v_i \in e_j, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

### 2.2 Learnable hypergraph mask

Resting-state fMRI signals exhibit inherent instability. Consequently, constructing a personalized hypergraph structure is susceptible to fluctuations. To address this, we introduce a learnable probability mask, $\tilde{H} \in [0,1]^{N\times E}$, to represent a groupwise hypergraph structure shared across all subjects. Each element $\tilde{H}_{i,j}$ encodes the probability that the $i$-th region belongs to the $j$-th hyperedge:

$$H_{i,j} = \begin{cases} 1, & \text{if } \tilde{H}_{i,j} > 0.5, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The hard threshold in (3) is non-differentiable. We therefore employ the Straight-Through Estimator (STE) Van Den Oord et al. [2017] to approximate the gradient during backpropagation:

$$H_{\text{STE}} = \mathbb{I}[\tilde{H} > 0.5] + \big(\tilde{H} - \text{sg}(\tilde{H})\big), \tag{4}$$

where $\mathbb{I}(\cdot)$ is the indicator function and $\text{sg}(\cdot)$ denotes the stop-gradient operator.

### 2.3 Multi-head attention

After obtaining the hypergraph structure $H$, the attention scores are represented as $\hat{H} = [\hat{H}^1, \hat{H}^2, ..., \hat{H}^{N_h}]$, where $N_h$ is the number of heads. Each $\hat{H}^k$ reflects the importance of a vertex within a hyperedge and controls message passing. Hyperedge features are estimated by

$$X_{e_j} = \sum_{v_i \in e_j} \tilde{H}_{i,j}\, X_i. \tag{5}$$

We use $\tilde{H}$ as weights to compute $X_{e_j}$ rather than the binary $H$.

The attention score between hyperedge and vertex is

$$\hat{H}^k = \frac{\exp\left(\text{sim}(x_i W_v^k,\ X_{e_j} W_e^k)\right)}{\sum_{v_a \in e_j} \exp\left(\text{sim}(x_a W_v^k,\ X_{e_j} W_e^k)\right)} \cdot H, \tag{6}$$

with

$$\mathrm{sim}(a,b) = \frac{a^\top b}{\sqrt{d_k}}, \tag{7}$$

where $W_v^k$ and $W_e^k$ are learnable projections and $d_k$ is the feature dimension. The operator "$\cdot$" denotes elementwise masking that zeroes out entries where $H_{i,j} = 0$.

### 2.4 Feature extraction

Under the assumption $X \to \hat{H} \to Y$ (a Markov chain), once $\hat{H}$ is obtained, information from $X$ should not be retained. To ensure this, we use learnable parameters $X_L^k \in \mathbb{R}^{N \times d_{\text{in}}}$ as node features for the $k$-th head in $L$-th layer for downstream feature extraction. All samples share the same $X_L^k$ in each branch. A hypergraph convolution layer Bai et al. [2021] extracts features:

$$\begin{aligned} X_{\text{out}}^k &= \sigma\Big(\mathrm{HCN}(X_L^k, \hat{H}^k, \theta^k)\Big) \\ &= \sigma\Big((D_v^k)^{-\frac{1}{2}} \hat{H}^k W (D_e^k)^{-1} (\hat{H}^k)^\top (D_v^k)^{-\frac{1}{2}} X_L^k \theta^k\Big), \end{aligned} \tag{8}$$

where $D_v^k$ and $D_e^k$ are vertex and hyperedge degree matrices computed from $\hat{H}^k$, $\sigma$ is a nonlinearity, $\theta^k$ is the learnable weight, and $W$ is the hyperedge weight matrix (set to identity).

**Orthogonality regularization.** To prevent different heads from collapsing to similar subspaces, we add

$$\mathcal{L}_{\text{ortho}} = \sum_{1 \le k < k' \le N_h} \big\| (X_L^k)^\top X_L^{k'} \big\|_F^2. \tag{9}$$

### 2.5 Readout layer

Outputs from each head are vectorized and concatenated:

$$X_{\text{out}} = \big[\mathrm{Vec}(X_{\text{out}}^1) \parallel \cdots \parallel \mathrm{Vec}(X_{\text{out}}^{N_h})\big],$$

and fed to a two-layer MLP to predict $Y$:

$$Y_{\text{predict}} = \mathrm{MLP}(X_{\text{out}}). \tag{10}$$

### 2.6 Optimization objectives

Given input $X$ and target $Y$, the bottleneck objective is Eq. 1.

**Upper bound of $I(X, \hat{H})$.** We aim to minimize $I(X; \hat{H})$ so that only task-related structure remains:

$$\begin{aligned} I(X;\hat{H}) &\le \sum_{k=1}^{N_h} I(\hat{H}^k, X) \le \sum_{k=1}^{N_h} \sum_{j=1}^{|E|} I(\hat{H}_j^k, X) \\ &\le \sum_{k=1}^{N_h} \sum_{j=1}^{|E|} \sum_{i=1}^{N} I(\hat{H}_{i,j}^k, X_i) = N_h \sum_{j=1}^{|E|} \sum_{i=1}^{N} \mathcal{H}(X_i)\, \tilde{H}_{i,j}, \end{aligned} \tag{11}$$

where $\mathcal{H}(\cdot)$ is entropy and the last equality follows Kim et al. [2021] under independence assumptions. If node entropies are identical, this reduces to an $\ell_1$-norm.

**Lower bound of $I(Y, \hat{H})$.** Maximizing predictive power:

$$\begin{aligned} I(\hat{H};Y) &= \mathbb{E}_{p(\hat{H},Y)} \log \frac{p(Y|\hat{H})}{p(Y)} = \mathbb{E}_{q(\hat{H},Y)} \log \frac{p(Y|\hat{H})}{p(Y)} \\ &= \mathbb{E}_{q(\hat{H},Y)} \log p(Y|\hat{H}) - \mathbb{E}_{q(\hat{H},Y)} \log p(Y) \\ &\ge \mathbb{E}_{q(\hat{H},Y)} \log p(Y|\hat{H}), \end{aligned} \tag{12}$$

where $q(Y|\hat{H})$ is a variational approximation. This corresponds to maximizing likelihood in classification.

**Overall objective.**

$$\mathcal{L} = \underbrace{\mathbb{E}_{q(\hat{H},Y)} \log p(Y \mid \hat{H}) + \beta\, N_h \sum_{i=1}^{N} \sum_{j=1}^{E} \mathcal{H}(X_i)\, \tilde{H}_{i,j}}_{\mathcal{L}_{\text{task}}} + \lambda_{\text{ortho}}\, \mathcal{L}_{\text{ortho}}. \tag{13}$$

# 3 Experiments

## 3.1 Materials and image processing

We conduct experiments on two datasets: Alzheimer's Disease Neuroimaging Initiative 3 (ADNI-3) Weiner et al. [2017] and Autism Brain Imaging Data Exchange (ABIDE) Craddock et al. [2013].

**ADNI-3.** All T1-w sMRI and rs-fMRI data are preprocessed by a standard pipeline (including FreeSurfer Fischl [2012]), skull stripping, motion correction, normalization, and registration. Brains are parcellated with the Desikan–Killiany atlas Desikan et al. [2006], 34 cortical areas per hemisphere. Two subsets are used. (i) **ADNI-3 (CN, MCI)**: 145 mild cognitive impairment (MCI) and 145 cognitively normal (CN) subjects, one fMRI per subject. (ii) **ADNI-3 (Inferior Temporal)**: Tau-PET images are processed (FreeSurfer), intensity-normalized using inferior cerebellar gray matter to obtain SUVR images. Amyloid-positive subjects are partitioned by mean inferior temporal SUVR threshold 1.3 Biel et al. [2021], further constrained by inferior parietal SUVR $< 1.3$. This yields 89 "high-tau" vs. 100 "low-tau" subjects.

**ABIDE.** Preprocessed rs-fMRI are from PCP Craddock et al. [2013]. Scans with any ROI mean of zero across time are excluded. Remaining data: 442 ASD and 376 HC. CPAC pipeline outputs are used, and each brain is parcellated by AAL116 into 116 ROIs.

### 3.1.1 Implementation details

We adopt a five-fold cross-validation strategy, where nested cross-validation is employed to perform hyperparameter tuning via grid search. Pearson correlation (PC) matrices are used as the input features. To provide a meaningful initialization for the learnable probability mask $\tilde{H}$, we apply the Fuzzy C-Means algorithm Bezdek et al. [1984] on the PC matrix to estimate the initial membership probabilities. Model training is conducted using the Adam optimizer with a learning rate of $1 \times 10^{-3}$, weight decay of $1 \times 10^{-2}$, and a maximum of 200 training epochs. For evaluation, we report four widely used metrics in neuroimaging classification: accuracy (ACC), area under the ROC curve (AUC), sensitivity (SEN), and specificity (SPE).

### 3.1.2 Computation complexity

The multi-head attention and hypergraph convolution modules are $O(N_h N E d_{\text{in}})$ and $O(N_h N E d_\ell)$, respectively. Overall complexity is $O(k\, N_h N E d)$, where $d = \max(d_{\text{in}}, d_\ell)$, comparable to BrainGNN and BrainNetTF.

### 3.1.3 Competing methods

We compare our framework against six representative baselines that cover both traditional and state-of-the-art models for brain network analysis: (1) PC+MLP Popescu et al. [2009], a simple baseline that directly feeds Pearson correlation features into a multilayer perceptron; (2) GCN Kipf and Welling [2016], the classical graph convolutional network applied to brain connectivity graphs; (3) wHGNN Gao et al. [2022], a weighted hypergraph neural network that extends GCNs to capture higher-order interactions; (4) BrainNetTF Kan et al. [2022], a transformer-based model tailored for brain network analysis; (5) BrainGNN Li et al. [2021], an interpretable GNN framework that emphasizes ROI-level biomarkers; and (6) HYBRID Qiu et al. [2023], a recent method that leverages the information bottleneck principle to identify task-relevant regions.

Table 1: Performance comparison with different baselines.

| Dataset | Method | ACC | SPE | SEN | AUC |
|---|---|---|---|---|---|
| ADNI3 (CN, MCI) | PC+MLP | 70.8±3.0% | 75.1±3.8% | 66.5±4.5% | 72.2±3.8% |
| | GCN | 72.6±5.7% | 70.3±6.2% | 74.0±5.9% | 73.8±4.5% |
| | wHGNN | 74.3±5.1% | 72.8±6.8% | 75.9±4.9% | 75.2±4.2% |
| | BrainNetTF | 71.5±6.6% | 67.8±7.2% | 75.2±6.7% | 74.3±6.1% |
| | BrainGNN | 73.3±4.6% | 69.4±5.8% | **77.2±4.4%** | 75.8±3.6% |
| | HYBRID | 69.7±7.8% | 66.9±7.2% | 72.5±6.9% | 70.2±6.6% |
| | **Proposed** | **79.9±5.1%** | **83.7±4.8%** | 76.0±7.1% | **82.6±4.8%** |
| ADNI3 (Inferior Temporal) | PC+MLP | 75.9±6.0% | 78.7±5.2% | 73.1±6.5% | 77.0±4.7% |
| | GCN | 78.6±5.7% | 80.1±6.3% | 77.1±7.2% | 80.2±5.3% |
| | wHGNN | 80.4±3.6% | 82.8±3.8% | 78.0±6.1% | 81.2±3.2% |
| | BrainNetTF | 81.1±4.2% | 83.4±3.9% | 78.8±4.8% | 84.2±3.8% |
| | BrainGNN | 79.5±6.7% | 84.1±7.0% | 74.4±8.4% | 80.24±7.0% |
| | HYBRID | 78.4±6.8% | 82.9±7.1% | 73.3±9.1% | 78.2±6.2% |
| | **Proposed** | **84.0±4.9%** | **86.8±3.7%** | **81.3±5.2%** | **86.2±3.8%** |
| ABIDE | PC+MLP | 66.3±8.1% | 69.6±7.0% | 63.8±10.2% | 69.0±7.7% |
| | GCN | 64.8±7.2% | 70.1±7.5% | 59.5±8.3% | 67.4±7.5% |
| | wHGNN | 70.1±6.2% | 68.5±6.7% | 71.3±5.9% | 73.2±5.0% |
| | BrainNetTF | 69.9±8.7% | 71.2±9.2% | 68.0±8.3% | 69.8±7.8% |
| | BrainGNN | 67.8±9.1% | 71.0±11.2% | 65.2±7.6% | 68.0±8.1% |
| | HYBRID | 71.3±7.0% | **74.8±7.1%** | 69.6±8.0% | 72.5±6.4% |
| | **Proposed** | **73.1±7.2%** | 70.6±5.2% | **76.5±6.1%** | **75.5±5.6%** |

# 4 Results

## 4.1 Classification performance

The results in Table 1 show that PC+MLP is a strong baseline. wHGNN outperforms GCN, suggesting hypergraphs better represent brain networks. Our method achieves the best ACC and AUC, improving accuracy by $5.6\%$, $2.9\%$, and $2.2\%$ on the three datasets, respectively. Performance is higher on ADNI3 (Inferior Temporal) than on (CN, MCI), indicating reduced pathological heterogeneity improves discrimination. ABIDE remains challenging due to heterogeneity.

## 4.2 Ablation study: effectiveness of hypergraph

We compare our learnable mask against: (1) **Random Mask** (random binary mask), and (2) **Soft Mask** (use continuous $\tilde{H}$ directly). As shown in Fig. 2, Random Mask underperforms, confirming the importance of a task-related groupwise structure. Soft Mask achieves comparable averages, but the hard mask yields (i) sparser, more interpretable structure and (ii) stronger robustness on difficult splits.

## 4.3 Comparison of discrete masking methods

In addition to STE, Gumbel-Softmax Jang et al. [2016] provides continuous relaxation for discrete sampling. We compare STE and Gumbel-Softmax ($\tau = 1$) under different $\beta$. Fig. 3 (left) shows the accuracy; STE is higher in most cases. Although Gumbel-Softmax marginally leads at $\beta = 100$, it typically enforces sparsity less aggressively. We define average mask sparsity

$$\text{Sparsity}(H) = 1 - \frac{1}{N\,E} \sum_{i=1}^{N} \sum_{j=1}^{E} \mathbf{1}[H_{i,j} = 1]\,,$$

and plot it on ABIDE in Fig. 3 (right). At equal $\beta$, STE yields higher sparsity and smoother convergence; Gumbel-Softmax fluctuates more and can remain dense for small $\beta$.

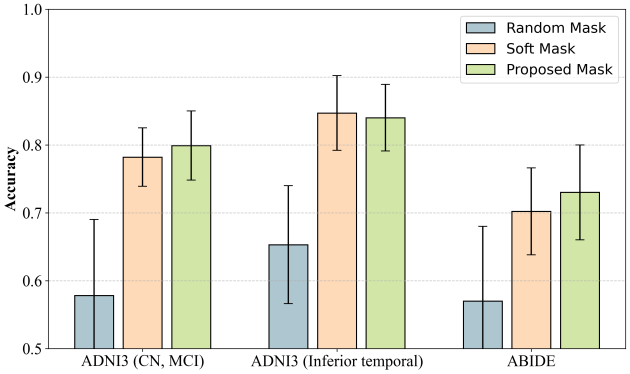
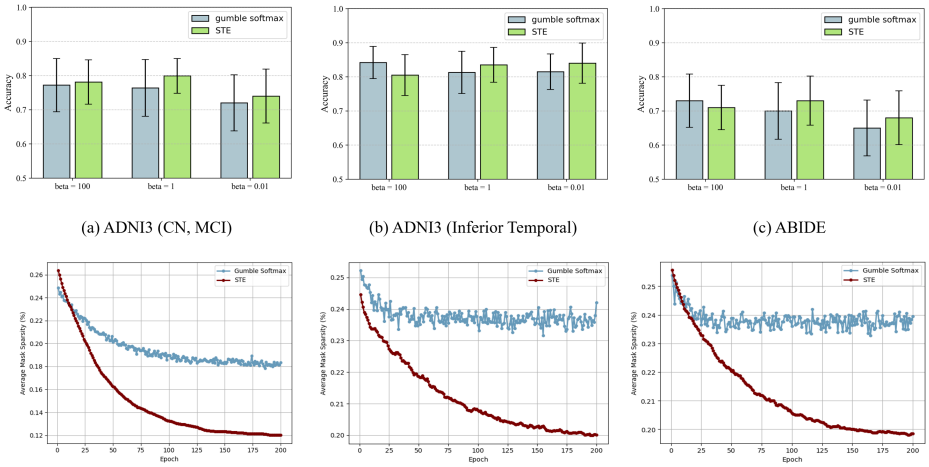



Figure 2: Ablation on masking strategies.




(a) ADNI3 (CN, MCI) (b) ADNI3 (Inferior Temporal) (c) ABIDE

(d) β = 100 (e) β = 1 (f) β = 0.01

Figure 3: (a)–(c) Accuracy under varying $\beta$ for STE vs. Gumbel-Softmax. (d)–(f) Mask sparsity curves under varying $\beta$ on ABIDE.

### 4.4 Impact of K-heads

We evaluate different head numbers $k$, with/without orthogonality regularization. When $k = 1$, no regularization is applied. Fig. 4 shows that orthogonality regularization improves performance on ADNI3. Without it, increasing $k$ does not help, indicating diversity across heads is beneficial.

### 4.5 Impact of number of hyperedges

We vary $E$ and evaluate accuracy. From Fig. 5, accuracy increases for $E < 10$, with peaks around $E = 20$–$30$. ABIDE uses 116 ROIs (vs. 68 in ADNI3), which may explain larger optimal $E$. For $E > 20$, gains plateau or drop due to noise/overfitting.

### 4.6 Vertex degree distribution

We compute vertex degree based on $H$ to assess ROI importance (higher degree $\Rightarrow$ more hyperedge participation). We set $E = 20$ for ADNI3 tasks and $E = 30$ for ABIDE. Regions with higher normalized degrees are likely more important.

For ADNI3 (CN, MCI), top regions include inferior parietal lobule, cuneus, inferior temporal gyrus, paracentral lobule, and right precuneus. Notably, inferior parietal and precuneus are DMN hubs; cuneus has been implicated in Alzheimer's pathology Wu et al. [2021]. For the inferior-temporal

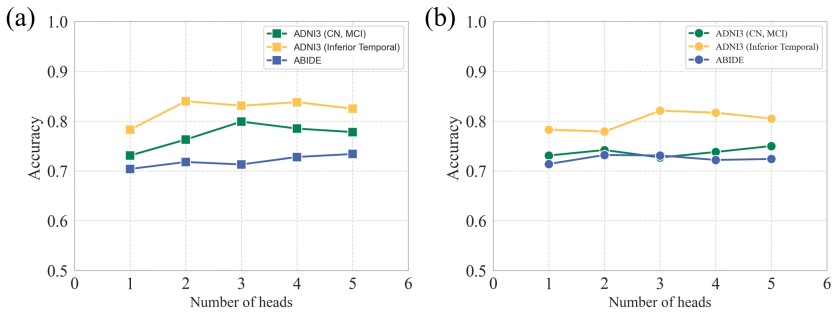



Figure 4: Performance vs. number of attention heads ($k$) with (a) and without (b) orthogonality regularization.

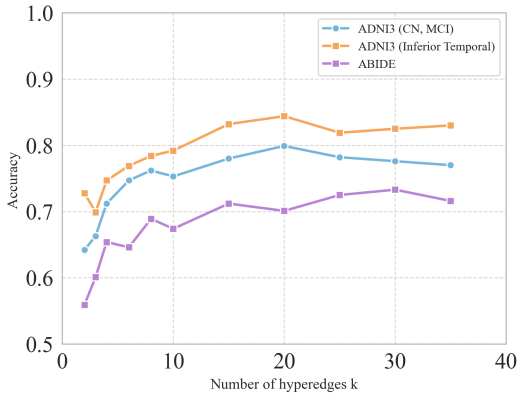



Figure 5: Performance under varying numbers of hyperedges $E$.

SUVR split, lateral orbitofrontal cortex, inferior temporal gyrus, posterior cingulate cortex, and paracentral lobule show elevated degrees; inferior temporal increases further, consistent with task focus, while inferior parietal decreases relative to (CN, MCI).

In ABIDE, the top-10 include left superior motor area, amygdala, superior frontal medial gyrus, left calcarine, left middle occipital gyrus, left paracentral, right thalamus, and right cerebellum. Prior work Baron-Cohen et al. [2000], Tomasi and Volkow [2019] links amygdala and thalamus to ASD; motor and visual network regions also appear. Several ROIs have zero degree (excluded by the mask).

## 5 Conclusion

We presented an end-to-end framework for constructing task-aware hypergraphs from fMRI data. It integrates: (1) a learnable groupwise probability mask with STE for discrete selection of task-relevant regions, (2) a multi-head hypergraph attention mechanism to weight region–hyperedge interactions, and (3) shared, head-specific node embeddings optimized under an information bottleneck objective with orthogonality regularization. Experiments on three public settings (ADNI-3 and ABIDE) show consistent gains over baselines; ablations confirm the contribution of the hard mask, the multi-head design, and the orthogonality constraint. Future work includes leveraging multimodal data (e.g., sMRI) and incorporating covariates (sex, age, site) to improve performance and generalization.

## References

Song Bai, Feihu Zhang, and Philip HS Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110:107637, 2021.

Antonello Baldassarre, Christopher M Lewis, Giorgia Committeri, Abraham Z Snyder, Gian Luca Romani, and Maurizio Corbetta. Individual variability in functional connectivity predicts perfor-

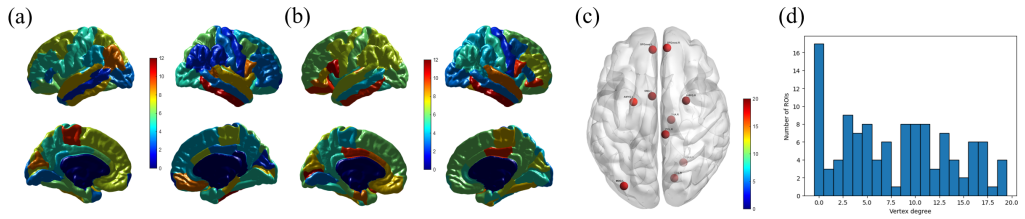



Figure 6: Vertex degree distribution: (a) ADNI-3 (CN, MCI). (b) ADNI-3 (Inferior Temporal). (c) Top-10 regions in ABIDE. (d) Histogram of vertex degree in ABIDE. Colorbar indicates degree.

mance of a perceptual task. *Proceedings of the National Academy of Sciences*, 109(9):3516–3521, 2012.

Simon Baron-Cohen, Howard A Ring, Edward T Bullmore, Sally Wheelwright, Chris Ashwin, and Steve CR Williams. The amygdala theory of autism. *Neuroscience & Biobehavioral Reviews*, 24 (3):355–364, 2000.

James C Bezdek, Robert Ehrlich, and William Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & geosciences*, 10(2-3):191–203, 1984.

Davina Biel, Matthias Brendel, Anna Rubinski, Katharina Buerger, Daniel Janowitz, Martin Dichgans, Nicolai Franzmeier, and Alzheimer's Disease Neuroimaging Initiative (ADNI). Tau-pet and in vivo braak-staging as prognostic markers of future cognitive decline in cognitively normal to demented individuals. *Alzheimer's research & therapy*, 13(1):137, 2021.

Susan M Bowyer. Coherence a measure of the brain networks: past and present. *Neuropsychiatric Electrophysiology*, 2(1):1, 2016.

Cameron Craddock, Yassine Benhajali, Carlton Chu, Francois Chouinard, Alan Evans, András Jakab, Budhachandra Singh Khundrakpam, John David Lewis, Qingyang Li, Michael Milham, et al. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics*, 7(27):5, 2013.

Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

Harini Eavani, Theodore D Satterthwaite, Roman Filipovych, Raquel E Gur, Ruben C Gur, and Christos Davatzikos. Identifying sparse connectivity patterns in the brain using resting-state fmri. *Neuroimage*, 105:286–299, 2015.

Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. Hgnn+: General hypergraph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3181–3199, 2022.

Linda Geerligs, Richard N Henson, et al. Functional connectivity and structural covariance between regions of interest can be measured more accurately using multivariate distance correlation. *NeuroImage*, 135:16, 2016.

Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016.

Junzhong Ji, Yating Ren, and Minglong Lei. Fc–hat: Hypergraph attention network for functional brain network classification. *Information Sciences*, 608:1301–1316, 2022.

Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer. *Advances in Neural Information Processing Systems*, 35:25586–25599, 2022.

Jaekyeom Kim, Minjung Kim, Dongyeon Woo, and Gunhee Kim. Drop-bottleneck: Learning discrete compressed representation for noise-robust exploration. *arXiv preprint arXiv:2103.12300*, 2021.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H Staib, Pamela Ventola, and James S Duncan. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical Image Analysis*, 74:102233, 2021.

Qi Lin, Monica D Rosenberg, Kwangsun Yoo, Tiffany W Hsu, Thomas P O'Connell, and Marvin M Chun. Resting-state functional connectivity predicts cognitive impairment related to alzheimer's disease. *Frontiers in aging neuroscience*, 10:94, 2018.

Jingyu Liu, Weigang Cui, Yipeng Chen, Yulan Ma, Qunxi Dong, Ran Cai, Yang Li, and Bin Hu. Deep fusion of multi-template using spatio-temporal weighted multi-hypergraph convolutional networks for brain disease analysis. *IEEE Transactions on Medical Imaging*, 43(2):860–873, 2023.

Guillaume Marrelec, Alexandre Krainik, Hugues Duffau, Mélanie Pélégrini-Issac, Stéphane Lehéricy, Julien Doyon, and Habib Benali. Partial correlation for functional brain interactivity investigation in functional mri. *Neuroimage*, 32(1):228–237, 2006.

Marius-Constantin Popescu, Valentina E Balas, Liliana Perescu-Popescu, and Nikos Mastorakis. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7): 579–588, 2009.

Jonathan D Power, Alexander L Cohen, Steven M Nelson, Gagan S Wig, Kelly Anne Barnes, Jessica A Church, Alecia C Vogel, Timothy O Laumann, Fran M Miezin, Bradley L Schlaggar, et al. Functional network organization of the human brain. *Neuron*, 72(4):665–678, 2011.

Weikang Qiu, Huangrui Chu, Selena Wang, Haolan Zuo, Xiaoxiao Li, Yize Zhao, and Rex Ying. Learning high-order relationships of brain regions. *arXiv preprint arXiv:2312.02203*, 2023.

Ming Song, Yuan Zhou, Jun Li, Yong Liu, Lixia Tian, Chunshui Yu, and Tianzi Jiang. Brain spontaneous functional connectivity and intelligence. *Neuroimage*, 41(3):1168–1176, 2008.

Felice T Sun, Lee M Miller, and Mark D'esposito. Measuring interregional functional connectivity using coherence and partial coherence analyses of fmri data. *Neuroimage*, 21(2):647–658, 2004.

Yingzhi Teng, Kai Wu, Jing Liu, Yifan Li, and Xiangyi Teng. Constructing high-order functional connectivity networks with temporal information from fmri data. *IEEE Transactions on Medical Imaging*, 2024.

Dardo Tomasi and Nora D Volkow. Reduced local and increased long-range functional connectivity of the thalamus in autism spectrum disorder. *Cerebral Cortex*, 29(2):573–585, 2019.

Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.

Junqi Wang, Hailong Li, Gang Qu, Kim M Cecil, Jonathan R Dillman, Nehal A Parikh, and Lili He. Dynamic weighted hypergraph convolutional network for brain functional connectome analysis. *Medical image analysis*, 87:102828, 2023.

Michael W Weiner, Dallas P Veitch, Paul S Aisen, Laurel A Beckett, Nigel J Cairns, Robert C Green, Danielle Harvey, Clifford R Jack Jr, William Jagust, John C Morris, et al. The alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's & Dementia*, 13(5):561–571, 2017.

Bang-Sheng Wu, Ya-Ru Zhang, Hong-Qi Li, Kevin Kuo, Shi-Dong Chen, Qiang Dong, Yong Liu, and Jin-Tai Yu. Cortical structure and the risk for alzheimer's disease: a bidirectional mendelian randomization study. *Translational psychiatry*, 11(1):476, 2021.

Li Xiao, Junqi Wang, Peyman H Kassani, Yipu Zhang, Yuntong Bai, Julia M Stephen, Tony W Wilson, Vince D Calhoun, and Yu-Ping Wang. Multi-hypergraph learning-based brain functional connectivity analysis in fmri data. *IEEE transactions on medical imaging*, 39(5):1746–1758, 2019.

Zhen Zhou, Xiaobo Chen, Yu Zhang, Dan Hu, Lishan Qiao, Renping Yu, Pew-Thian Yap, Gang Pan, Han Zhang, and Dinggang Shen. A toolbox for brain network construction and classification (brainnetclass). *Human brain mapping*, 41(10):2808–2826, 2020.

Pascal Zille, Vince D Calhoun, Julia M Stephen, Tony W Wilson, and Yu-Ping Wang. Fused estimation of sparse connectivity patterns from rest fmri—application to comparison of children and adult brains. *IEEE transactions on medical imaging*, 37(10):2165–2175, 2017.