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

Mingyang Xia
Ming Hsieh Department of Electrical and Computer Engineering
University of Southern California
Los Angeles, CA 90007
xiamingy@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.

1 Introduction

Functional magnetic resonance imaging (fMRI) is widely used to study brain activity and neurological disorders such as Alzheimer's disease and autism. Unlike static imaging, fMRI produces a four-dimensional (3D space + time) signal that records blood-oxygen-level-dependent (BOLD) fluctuations across brain regions over time. A major challenge lies in modeling these temporal dynamics and their interactions across regions in order to classify different brain states. Conventional graph-based methods typically summarize fMRI time series into pairwise functional connectivity matrices using statistical measures such as Pearson/partial correlation or coherence [1–3]. However, these approaches only capture pairwise relations and thus overlook higher-order interactions, such as those underlying default mode or motor networks [4].

Hypergraphs provide a natural extension by allowing hyperedges to connect multiple regions simultaneously, thereby modeling higher-order brain connectivity. Recent studies [5, 6] have applied hypergraphs to disease classification, but most rely on fixed hypergraph constructions that are not optimized for downstream tasks. To address this, we propose an end-to-end framework that simultaneously constructs hypergraphs and optimizes prediction. Our model leverages a groupwise task-related hypergraph with multi-head attention and incorporates the information bottleneck to enhance region selection. Evaluated on ADNI-3 [7] and ABIDE [8], our method consistently outperforms baselines and identifies disease-relevant regions, demonstrating its potential in neuroimaging-based diagnosis.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: NeurIPS 2025 Workshop on Learning from Time Series for Health .

2 Methods

To capture task-related brain regions and their connectivity, we identify informative regions and construct a hypergraph representation as the basis for downstream analysis. For M subjects, each subject is represented as (X_i, Y_i), where $X_i \in \mathbb{R}^{N \times P}$ denotes features from N brain regions with dimension P, and Y_i encodes subject-specific states. Let $\hat{H}$ be the latent hypergraph with E hyperedges. Following the information bottleneck principle, we model $(X, \hat{H}, Y)$ as a Markov chain $X \to \hat{H} \to Y$ and aim to optimize

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

where β balances compression and predictive power.

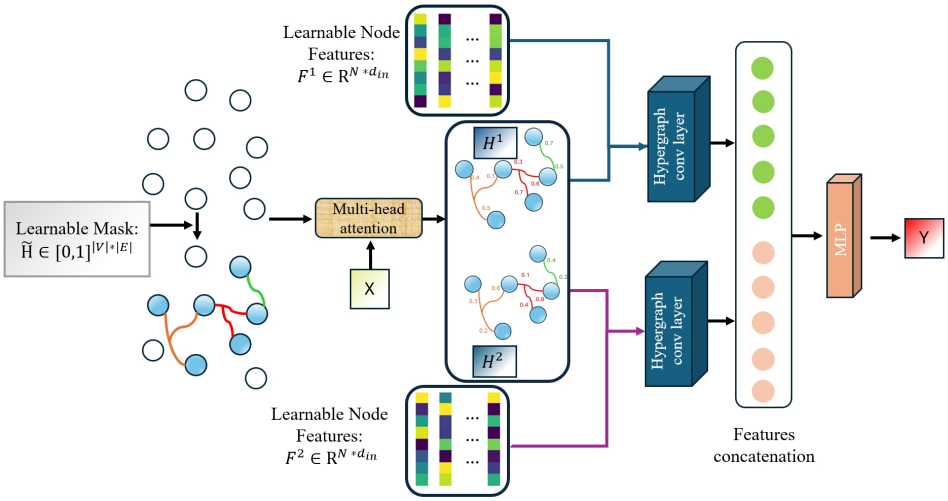

Figure 1: An illustration of the proposed method.

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

2.1 Hypergraph

A hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ is characterized by a vertex set $\mathcal{V} = \{v_1, v_2, ..., v_N\}$ and a collection of hyperedges $\mathcal{E} = \{e_1, e_2, ..., e_{|E|}\}$. Instead of describing relationships through pairwise connections, the structure can be compactly encoded in an incidence matrix H.

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

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, denoted $\tilde{H} \in [0, 1]^{N*E}$, to represent a groupwise hypergraph structure shared across all subjects. Each element $\tilde{H}_{i,j}$ encodes the probability that the i_{th} regions belong to the j_{th} hyperedge.

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

It should be noted that the hard threshold in (3) is non-differentiable and non-gradient flows through the step $H_{i,j} = 1[\tilde{H}_{i,j} > 0.5]$. To enable end-to-end learning, we therefore employ the Straight-Through Estimator (STE) [9] to approximate the gradient of this thresholding operation during backpropagation.

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

where $\mathbb{I}(\cdot)$ represents an indicator function and $\mathrm{sg}(\cdot)$ is the stop-gradient operator, which prevents gradients from flowing through its argument during backpropagation.

2.3 Multi-head attention

After obtaining the hypergraph structure H, the attention scores are represented as $\hat{H} = [\hat{H}^1, \hat{H}^2, ..., \hat{H}^{N_h}]$. N_h denotes the number of attention heads, which means that N_h sets of attention scores are computed. Each $\hat{H}^k$ is calculated to reflect the importance of a vertex within a hyperedge, which can also control the flow of message passing. The hyperedge features are estimated by $X_{e_j} = \sum_{v_i \in ej} \tilde{H}_{i,j} * X_i$.

X_{e_j} is the features of the hyperedge e_j. It should be mentioned that we use the probability $\tilde{H}$ as the weight to estimate hyperedge features rather than using the binary value in H.

The attention score between hyperedge and vertex is calculated as follows:

$$\hat{H}^k = \frac{exp(sim(x_i * W_v^k, X_{e_j} * W_e^k))}{\sum_{v_a \in ej} exp(sim(x_a * W_v^k, X_{e_j} * W_e^k))} \cdot H, sim(a,b) = \frac{a * b}{\sqrt{d_k}}, \tag{5}$$

where Wv^k and We^k are the learnable projection matrices that map the vertex features and hyperedges features to a new feature space. $\cdot$ represents the dot products, which only retain the value where $H_{i,j} = 1$. $Sim(,)$ is a similarity matrix that computes the distance between the vertex and the hyperedges and d_k is the feature dimension of W_v^k and W_e^k.

2.4 Prediction via Learnable Feature Extraction

In our assumption, $X \to \hat{H} \to Y$ forms a Markov chain, which means that once $\hat{H}$ is obtained, the information from X should not be retained. To ensure this, the learnable hyperparameters $X_L^k \in R^{N * d_{in}}$ are used as node features for the k_{th} head (branch) for subsequent extraction of high-level features. It should be mentioned that all samples share the same node features in each branch in the following steps. This method utilizes a multi-head attention hypergraph matrix. To extract features based on the hypergraph attention matrix, the hypergraph convolution layer [10] is applied to update and integrate information:

$$\begin{aligned} X_{out}^k &= \sigma(HCN(X_L^k, \hat{H}^k, \theta^k)) \\ &= \sigma((D_v^k)^{-\frac{1}{2}} \hat{H}^k W (D_e^k)^{-1} (\hat{H}^k)^T (D_v^k)^{-\frac{1}{2}} X_L^k \theta^k), \end{aligned} \tag{6}$$

The output features after each hypergraph convolution operation are then reshaped into a vector and concatenated together. $X_{out} = \Big[Vec(X_{out}^0)||Vec(X_{out}^1)||...||Vec(X_{out}^{N_h})\Big]$. A multilayer perceptron (MLP) is applied to predict the label Y.

2.5 Optimization Objectives

Given an input X and its corresponding target Y, a bottleneck framework formalizes the problem of obtaining the intermediate variable $\hat{H}$ could be written as follows:

$$\mathcal{L} = \underbrace{E_{q(\hat{H},Y)} \log p(Y \mid \hat{H}) + \beta\, N_h \sum_{i=1}^{N} \sum_{j=1}^{E} \mathcal{H}(X_i)\, \tilde{H}_{i,j}}_{\mathcal{L}_{\text{task}}} \quad + \lambda_{\text{ortho}} \sum_{1 \le k < k' \le N_h} \left\| (X_L^k)^\top X_L^{k'} \right\|_F^2, \tag{7}$$

where λ_{ortho} is a hyperparameter that balances classification performance against head-diversity.

3 Experiments

3.1 Materials and image processing

We conduct experiments on two datasets: Alzheimer's Disease Neuroimaging Initiative 3 (ADNI-3) [7] and Autism Brain Imaging Data Exchange (ABIDE) [8].

ADNI-3. All T1-w sMRI and rs-fMRI data are preprocessed by a standard pipeline , skull stripping, motion correction, normalization, and registration. Brains are parcellated with the Desikan–Killiany atlas, 34 cortical areas per hemisphere. The datasete includes 145 mild cognitive impairment (MCI) and 145 cognitively normal (CN) subjects, one fMRI scan per subject.

ABIDE. The preprocessed rs-fMRI is from PCP [8]. Scans with any ROI mean of zero across time are excluded. Remaining data: 442 subjects with autism disease (ASD) and 376 health control subjects (HC). The CPAC pipeline outputs are used and each brain is split by AAL116 into 116 ROIs.

Table 1: Performance comparison with different baselines.

Dataset	Method	ACC	SPE	SEN	AUC
ADNI3	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%**
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%**

As shown in Table.1, the proposed method achieves highest performance compared to the other methods. It can be observed that the proposed method outperforms all other methods in terms of ACC and AUC. The proposed method improves accuracy by 5.6%, and 2.2% on the two datasets, respectively.

4 Conclusion

In this work, we proposed an end-to-end framework that jointly constructs task-aware hypergraphs from fMRI data and optimizes downstream prediction. By integrating multi-head hypergraph attention with the information bottleneck principle, our method enhances region selection and representation learning. Experiments on ADNI-3 and ABIDE demonstrate consistent performance gains over baseline approaches, highlighting the potential of hypergraph-based modeling for neuroimaging-based disease diagnosis.

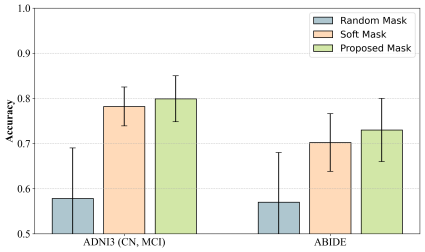

Figure 2: Ablation Study on Mask Construction.

Ablation on Masking Strategies
To validate the effectiveness of our proposed hypergraph mask, we compare it against two alternative masking strategies: **Random Mask**: replace the learnable mask with a randomly generated binary mask. **Soft Mask**: use the continuous probability mask $\tilde{H}$ directly in the attention computation (i.e. instead of binarization to H in (5)).

References

[1] Z. Zhou, X. Chen, Y. Zhang, D. Hu, L. Qiao, R. Yu, P.-T. Yap, G. Pan, H. Zhang, and D. Shen, "A toolbox for brain network construction and classification (brainnetclass)," *Human brain mapping*, vol. 41, no. 10, pp. 2808–2826, 2020.

[2] G. Marrelec, A. Krainik, H. Duffau, M. Pélégrini-Issac, S. Lehéricy, J. Doyon, and H. Benali, "Partial correlation for functional brain interactivity investigation in functional mri," *Neuroimage*, vol. 32, no. 1, pp. 228–237, 2006.

[3] M. Song, Y. Zhou, J. Li, Y. Liu, L. Tian, C. Yu, and T. Jiang, "Brain spontaneous functional connectivity and intelligence," *Neuroimage*, vol. 41, no. 3, pp. 1168–1176, 2008.

[4] J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar *et al.*, "Functional network organization of the human brain," *Neuron*, vol. 72, no. 4, pp. 665–678, 2011.

[5] J. Ji, Y. Ren, and M. Lei, "Fc–hat: Hypergraph attention network for functional brain network classification," *Information Sciences*, vol. 608, pp. 1301–1316, 2022.

[6] J. Liu, W. Cui, Y. Chen, Y. Ma, Q. Dong, R. Cai, Y. Li, and B. Hu, "Deep fusion of multi-template using spatio-temporal weighted multi-hypergraph convolutional networks for brain disease analysis," *IEEE Transactions on Medical Imaging*, vol. 43, no. 2, pp. 860–873, 2023.

[7] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. Jack Jr, W. Jagust, J. C. Morris *et al.*, "The alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement," *Alzheimer's & Dementia*, vol. 13, no. 5, pp. 561–571, 2017.

[8] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. S. Khundrakpam, J. D. Lewis, Q. Li, M. Milham *et al.*, "The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives," *Frontiers in Neuroinformatics*, vol. 7, no. 27, p. 5, 2013.

[9] A. Van Den Oord, O. Vinyals *et al.*, "Neural discrete representation learning," *Advances in neural information processing systems*, vol. 30, 2017.

[10] S. Bai, F. Zhang, and P. H. Torr, "Hypergraph convolution and hypergraph attention," *Pattern Recognition*, vol. 110, p. 107637, 2021.

A Appendix

A.1 Prove of the Objective Function

Given an input X and its corresponding target Y, a bottleneck framework formalizes the problem of obtaining the intermediate variable $\hat{H}$ could be written as Eq.1.

A.1.1 Upperbound of $I(X, \hat{H})$

$I(X, \hat{H})$ measures the mutual information between input and underlying hypergraph. The goal is to only keep the task-related structure; Hence, the mutual information between X and H should be minimized.

$$\begin{aligned} I(X; \hat{H}) &\leq \sum_{k=1}^{N_h} I(\hat{H}^k, X) \\ &\leq \sum_{k=1}^{N_h} \sum_{j=1}^{|E|} I(\hat{H}_j^k, X) \\ &\leq \sum_{k=1}^{N_h} \sum_{j=1}^{|E|} \sum_{i=1}^{N} I(\hat{H}_{i,j}^k, X_i) \\ &= N_h * \sum_{j=1}^{|E|} \sum_{i=1}^{N} \mathcal{H}(X_i) * \tilde{H}_{i,j} \end{aligned} \tag{8}$$

where $\mathcal{H}$ denotes the computation of entropy and $\tilde{H}_{i,j}$ represents the probability that i_{th} regions kept in the j_{th} hyperedge. The proof of the last equation is referenced in the previous work [**?**]. The inequality holds only when the nodes and the hyperedges are independent. In practice, this serves as a relatively loose upper bound. Moreover, when the entropy of each node is identical, the equation reduces to a L_1-norm.

A.1.2 Lowerbound of $I(Y, \hat{H})$

$I(Y, \hat{H})$ measures the predictive power of H for Y, which should be maximized.

$$\begin{aligned} I(\hat{H}; Y) &= E_{p(\hat{H},Y)} log \frac{p(Y|\hat{H})}{p(Y)} \\ &= E_{q(\hat{H},Y)} log \frac{p(Y|\hat{H})}{p(Y)} \\ &= E_{q(\hat{H},Y)} log(p(Y|\hat{H})) - E_{q(\hat{H},Y)} log(p(Y)) \\ &\geq E_{q(\hat{H},Y)} log(p(Y|\hat{H})), \end{aligned} \tag{9}$$

where $q(Y|\hat{H})$ is the variational approximation of $p(Y|\hat{H})$. The above equation corresponds to maximizing the likelihood in a classification task.

A.1.3 Objective function

In general, the objective function could be written as follows:

$$\mathcal{L} = E_{q(\hat{H},Y)} \log p(Y \mid \hat{H}) + \beta \, N_h \sum_{i=1}^{N} \sum_{j=1}^{E} \mathcal{H}(X_i) \, \tilde{H}_{i,j}. \tag{10}$$

A.2 Datasets

We conduct experiments on two datasets: Alzheimer's Disease Neuroimaging Initiative 3 (ADNI-3) [7] and Autism Brain Imaging Data Exchange (ABIDE) [8].

ADNI-3: All T1-w sMRI and rs-fMRI data are preprocessed by a standard pipeline (including FreeSurfer), skull stripping, motion correction, normalization, and registration. Brains are parcellated with the Desikan–Killiany atlas, 34 cortical areas per hemisphere. 145 mild cognitive impairment (MCI) and 145 cognitively normal (CN) subjects are included in the dataset.

ABIDE: Preprocessed rs-fMRI are from PCP. Scans with any ROI mean of zero across time are excluded. Remaining data: 442 Autism disease subjects (ASD) and 376 health control (HC). CPAC pipeline outputs are used, and each brain is parcellated by AAL116 into 116 ROIs.