HIERARCHICAL HYPERGRAPH LEARNING WITH FUNCTIONAL ALIGNMENT FOR BRAIN DISORDER DIAGNOSIS

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph Neural Networks (GNNs) have shown promise in brain disorder diagnosis by modeling neurophysiological signals as graphs, such as Electroencephalogram (EEG) and functional Near-Infrared Spectroscopy (fNIRS). However, existing brain graph learning methods remain limited by hierarchical anatomical heterogeneity, manifested as inconsistencies for alterations caused by brain disorders across different anatomical levels. Specifically, this heterogeneity can be analyzed from two perspectives: (1) High-order interaction knowledge hierarchy heterogeneity, i.e., inconsistencies in interactions among brain elements across anatomical levels; the inadequate capture of level-specific knowledge leads to unsuitable modeling. (2) Functional role hierarchy heterogeneity, i.e., inconsistency in the functional roles of the same element when viewed from different levels, which leads to functional knowledge conflict and impairs GNN cross-level generalizability. To address these challenges, we propose **HHRA**, a framework that addresses dual hierarchical heterogeneity. First, it introduces **H**ierarchical **H**igh-order Knowledge Capturing (HHKC), which enables the joint learning on multiple anatomical levels. Specifically, by constructing both pairwise and hypergraph relations across levels, HHKC effectively enables capturing level-specific high-order interaction knowledge. Second, it introduces Hierarchical Functional **R**ole **A**lignment (HFRA), which promotes the coordination across multiple anatomical levels. Specifically, it estimates and aligns the functional importance of brain elements across levels for consistent knowledge and a cross-level generalizable GNN. Extensive experiments on EEG and fNIRS datasets validate the superiority of HHRA. The code is available at `https://anonymous.4open.science/r/HHRA-6656/`.

1 INTRODUCTION

Brain disorders (Tong et al., 2023; Wang et al., 2024b; Qiu et al., 2024; Yang et al., 2023), including mental disorders, neurodegenerative diseases, and others (Yu et al., 2022; Zhu et al., 2022; Zhao & Zhang, 2024; Dumitrascu et al., 2024), constitute a public health challenge and necessitate solutions. Existing research characterizes these disorders as manifestations of disrupted brain network function, thereby framing them as network diseases (Liu et al., 2017; Bassett et al., 2018; Bassett & Sporns, 2017; Burns et al., 2014). In this context, multichannel neurophysiological signals such as Electroencephalogram (EEG) and functional Near-Infrared Spectroscopy (fNIRS) (Graña & Morais-Quilez, 2023; Abadal et al., 2025; Hou et al., 2022; Zhao et al., 2025), are commonly modeling as graphs for brain disorder diagnosis. Therefore, Graph Neural Networks (GNNs) (Chen et al., 2025b; Wang et al., 2021; Lu et al., 2024), with their capacity for learning relationships and capturing dependencies(Duan et al., 2025; Qi et al., 2025; Luo et al., 2025; Ding et al., 2025), have emerged as paradigms for brain graph learning. While a comprehensive review of related work is deferred to Appendix A, we note that although notable progress has been made (Liu et al., 2024b; Huang et al., 2024), existing brain graph learning remains limited by **hierarchical anatomical heterogeneity**, manifested as inconsistencies for alterations caused by brain disorders across different anatomical levels (Alivisatos et al., 2013; Li et al., 2009; Wardlaw et al., 2020; Betzel & Bassett, 2017). Specifically, this heterogeneity can be analyzed from two perspectives: **high-order interaction knowledge hierarchical heterogeneity**, and **functional role hierarchical heterogeneity**.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

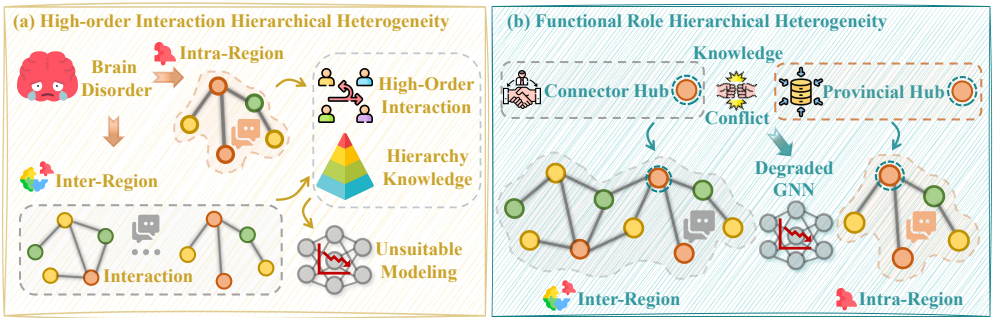


Figure 1: Problem illustration. Alterations induced by brain disorders exhibit hierarchical heterogeneity across anatomical levels, which can be analyzed from two perspectives: (a) High-order interaction knowledge among brain elements exhibits heterogeneity across levels. Therefore, the inadequate capture of level-specific knowledge leads to unsuitable modeling. (b) A given brain element may exhibit varying functional roles across different levels, resulting in functional role heterogeneity. It can introduce knowledge conflict and impair cross-level generalizability during learning, ultimately resulting in degraded GNN.

First, the functional interactions among brain network elements, such as brain regions or spatial areas captured by EEG channels, are high-order and complex rather than pairwise (Shen et al., 2024). Meanwhile, brain disorder–related alterations impact brain interactions differently across hierarchical anatomical levels, which we characterize as high-order interaction knowledge hierarchical heterogeneity. Consequently, existing single-level analyses often fail to capture level-specific pathological information, thereby limiting the capability of GNNs. Therefore, current methods based on pairwise relationships and single-view graph construction are insufficient to characterize the level-specific interaction dependencies(Do & Shin, 2024; 2023). In this context, although several brain graph learning studies have proposed multi-level learning, they primarily focus on collaboration across tasks or subject groups (Han et al., 2025), rather than on anatomical levels. Thereby failing to adequately capture level-specific high-order knowledge and leading to unsuitable modeling. Therefore, we raise the question 1): *How can **high-order** interaction knowledge be adequately captured in an anatomically **hierarchical** manner?*

Beyond the hierarchical heterogeneity in high-order interactions among multiple brain elements, each element may assume distinct functional roles across different levels. More concretely, an element may exhibit biased functional importance across levels during brain interactions, which leads to knowledge conflicts. For instance, a brain region may act as a provincial hub that primarily supports specialized information processing at a fine level, while emerging as a connector hub that integrates information at another level. Consequently, knowledge conflict induced by functional role hierarchy heterogeneity impairs cross-level generalizability and leads to degraded GNNs. Therefore, we pose the question 2): *How to align the **functional roles** across multiple levels against **heterogeneity**?*

To address question 1), we propose Hierarchical High-order Knowledge Capturing (HHKC). It constructs and jointly learns on three graphs from multiple anatomical levels, including a channel pairwise graph, a channel hypergraph, and a brain region hypergraph. The multi-level paradigm captures fine-grained hierarchical details, alleviating the adverse impact of heterogeneity in coarse learning for multiple anatomical levels. Moreover, the collaborative learning between pairwise and hypergraph relations effectively models the complex high-order interaction knowledge. Compared with existing hypergraph-based brain learning methods that focus solely on dynamic low-frequency (Ji et al., 2022) or attenuate attention to individual-level signals (Han et al., 2025; Zhang et al., 2022), HHKC naturally enables the anatomically hierarchical construction, enabling more effective learning of level-specific interaction knowledge.

To address question 2), we propose Hierarchical Functional Role Alignment (HFRA). Current neurophysiology research shows that variations in neural signals reflect the involvement of broader neuron populations, thereby indicating their functional role. Inspired by this phenomenon, HFRA quantifies node-wise signal variation in each view and combine it with structural information to estimate level-specific functional role. Subsequently, it aligns the role across multiple anatomical

levels, encouraging the brain element functional consistency in hierarchical collaborative learning. Consequently, HFRA enables GNNs to learn generalizable knowledge across levels and thereby mitigates knowledge conflicts under functional role heterogeneity.

In conclusion, our key contributions are:

- We identify the two key challenges under hierarchical anatomical heterogeneity in GNN-based brain disorder diagnosis.
- We propose HHRA, which enables hierarchical learning of the high-order knowledge and cross-level generalization.
- Extensive experiments on multiple EEG and fNIRS datasets demonstrate the superiority of our proposed framework.

2 METHODOLOGY

Overview. In Hierarchical Hypergraph Construction (HHKC), we jointly construct three anatomically hierarchical graphs: a channel-level pairwise graph that models direct functional interactions between channels, a channel-level hypergraph that groups each channel with its strongly coupled neighbors to capture high-order interactions, and a brain-region-level hypergraph that aggregates channels within anatomical regions and further explicitly models inter-region interactions based on region-wise functional similarity. Subsequently, we perform parallel message passing on each hierarchical view to learn level-specific representations, preserving the diversity of interaction patterns. Eventually, the learned representations from different views are retained independently and jointly concatenated for downstream brain disorder diagnosis. In Hierarchical Functional Role Alignment (HFRA), we first estimate the functional importance of each brain element under different hierarchical levels by jointly considering its signal diffusion variations and structural characteristics. Based on these estimations, we explicitly align the functional importance of the same element across levels, encouraging consistent functional roles in multi-level collaborative learning. Finally, the model is optimized under the joint supervision of the classification objective and the cross-level alignment objective, enabling effective hierarchical modeling and coordinated learning across anatomical levels.

Motivation of HHKC. Neurophysiological evidence suggests that alterations induced by brain disorders manifest differently across anatomical levels, while interactions among brain elements are often high-order rather than pairwise. Notably, brain disorder–related alterations impact brain interactions differently across hierarchical anatomical levels. Nevertheless, current brain graph learning either collapses multi-level information into a single view or coarse knowledge, thereby failing to capture level-specific high-order interaction patterns in a hierarchical manner. The limitations motivate the design of HHKC. By jointly constructing an anatomically hierarchical network from a unified functional connectivity source, HHKC explicitly enables collaborative learning on both pairwise and high-order interaction structures of multiple levels.

Motivation of HFRA. In addition to the hierarchical heterogeneity in high-order interactions among brain elements, individual elements can exhibit distinct functional roles across hierarchical levels. In particular, a brain element may exhibit inconsistent functional roles when observed at different anatomical levels, leading to bias that hinder cross-level generalizability. This motivates the design of HFRA. Instead of directly aligning node embeddings, HFRA focuses on aligning the functional role of brain elements across hierarchical views, which reflects their function in interactions. Therefore, it encourages coherent functional interpretations and mitigates conflicts induced by functional heterogeneity across levels. Consequently, HFRA enables effective hierarchical coordination, allowing HHRA to leverage level-specific knowledge while maintaining cross-level consistency, thereby addressing the second challenge under hierarchical anatomical heterogeneity.

2.1 HIERARCHICAL HYPERGRAPH CONSTRUCTION

Considering a multi-channel neurophysiological recording segment from one trial as a matrix $\mathbf{X} \in \mathbb{R}^{T\times K}$, where T is the number of temporal samples and K is the number of channels. We denote the number of brain regions as R. For a mini-batch of B trials, we write $\{\mathbf{X}^{(b)}\}_{b=1}^{B}$ with $\mathbf{X}^{(b)} \in \mathbb{R}^{T\times K}$. Note that the following graph construction is performed independently for each trial in the batch. First, we extract the instantaneous phase for each channel using the analytic signal obtained by the Hilbert

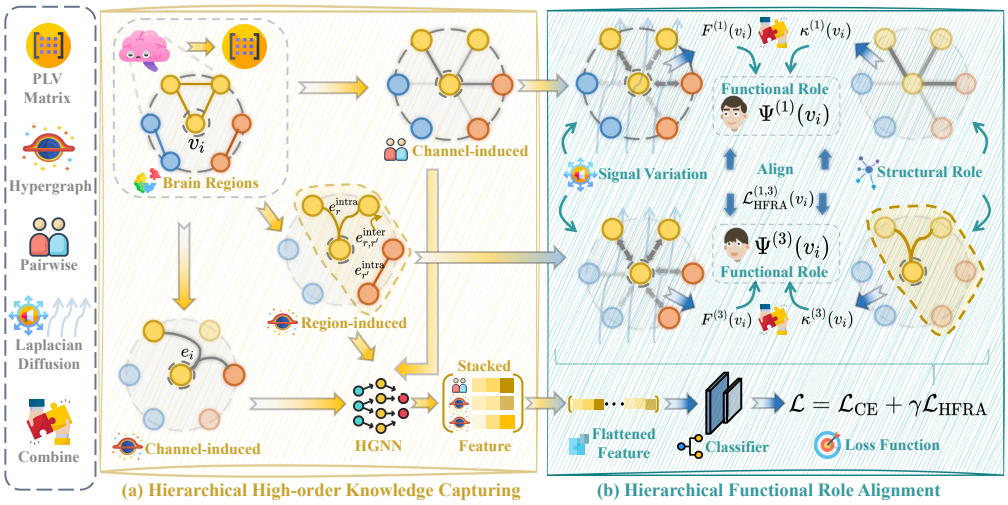


Figure 2: Architecture illustration. (a) Hierarchical hypergraphs are constructed from PLV-based functional connectivity at the pairwise channel, high-order channel, and high-order brain-region levels. We perform parallel message passing on multiple anatomical levels for level-specific brain interaction knowledge. (b) We evaluate and align the functional role that combines fractional Laplacian–based Fisher-inspired signal variation sensitivity and Forman–Ricci curvature across levels.

transform. Let $x_i(t)$ denote the signal of channel i at time index t, and let $z_i(t) = x_i(t) + j\mathcal{H}(x_i(t))$ be its analytic form, where $\mathcal{H}(\cdot)$ denotes the Hilbert transform and $j = \sqrt{-1}$. The instantaneous phase is $\phi_i(t) = \angle z_i(t)$. The Phase Locking Value (PLV) between channels i and j is computed as:

$$\mathrm{PLV}_{ij} = \left| \frac{1}{T} \sum_{t=1}^{T} \exp\left(j(\phi_i(t) - \phi_j(t))\right) \right|. \tag{1}$$

This equation measures how consistently the phase difference $\phi_i(t) - \phi_j(t)$ remains aligned over time, producing a similarity score in $[0, 1]$. Aggregating all channel pairs yields the PLV similarity matrix $\mathbf{S} \in \mathbb{R}^{K\times K}$ with entries $\mathbf{S}_{ij} = \mathrm{PLV}_{ij}$, which serves as the connectivity source for all three subsequent levels. We first construct a global channel-level propagation operator based on PLV similarities to capture pairwise functional relationships. Specifically, we apply symmetric normalization to the PLV similarity matrix:

$$\mathbf{A}^{(1)} = \left(\mathbf{D}^{(1)}\right)^{-\frac{1}{2}} \mathbf{S} \left(\mathbf{D}^{(1)}\right)^{-\frac{1}{2}}, \tag{2}$$

where $\mathbf{D}^{(1)} \in \mathbb{R}^{K\times K}$ is the diagonal degree matrix with $\mathbf{D}^{(1)}_{ii} = \sum_j \mathbf{S}_{ij}$. $\mathbf{A}^{(1)} \in \mathbb{R}^{K\times K}$ is the normalized propagation operator.

Subsequently, we construct a channel-level hypergraph that forms high-order relations. Specifically, we treat each channel as a graph node, and denote by v_i the node corresponding to channel i. For each node v_i, we select its $k^{(2)}$ nearest neighbors using PLV and define a hyperedge, where $k^{(2)} = 5$ in our implementation.

$$e_i = \{v_i\} \cup \mathcal{N}_{k^{(2)}}(v_i), \quad \mathcal{N}_{k^{(2)}}(v_i) = \mathrm{TopK}\left(\{\mathbf{S}_{ij}\}_{j\neq i}\right), \tag{3}$$

where $\mathcal{N}_{k^{(2)}}(v_i)$ denotes the set of indexes of the $k^{(2)}$ channels with the largest PLV similarities to v_i, and $\mathrm{TopK}(\cdot)$ returns the indices corresponding to these values. This equation induces a hyperedge set $\mathcal{E}^{(2)} = \{e_i\}_{i=1}^{K}$ that explicitly models local functional assemblies around each channel. Let $\mathbf{H}^{(2)} \in \{0, 1\}^{K\times|\mathcal{E}^{(2)}|}$ be the corresponding incidence matrix, where $\mathbf{H}^{(2)}_{i,e} = 1$ if node v_i belongs to hyperedge e and 0 otherwise. We then obtain a normalized hypergraph propagation operator for the second level:

$$\mathbf{A}^{(2)} = \left(\mathbf{D}^{(2)}_v\right)^{-\frac{1}{2}} \mathbf{H}^{(2)} \mathbf{W}^{(2)} \left(\mathbf{D}^{(2)}_e\right)^{-1} \left(\mathbf{H}^{(2)}\right)^{\top} \left(\mathbf{D}^{(2)}_v\right)^{-\frac{1}{2}}. \tag{4}$$

where $\mathbf{D}_v^{(2)}$ and $\mathbf{D}_e^{(2)}$ are diagonal node-degree and hyperedge-degree matrices induced by $\mathbf{H}^{(2)}$, and $\mathbf{W}^{(2)} \in \mathbb{R}^{|\mathcal{E}^{(2)}|\times|\mathcal{E}^{(2)}|}$ is a fixed diagonal hyperedge-weight matrix with all diagonal entries set to 1, where $|\mathcal{E}^{(2)}|$ is hyperedges scale in the channel-level hypergraph. We further build a brain-region hypergraph. Let $\mathcal{V}_r$ denote the set of nodes (channels) assigned to region r. We first enforce intra-region cohesion by creating one hyperedge per region:

$$e_r^{\text{intra}} = \mathcal{V}_r, \quad r = 1, \ldots, R. \tag{5}$$

where e_r^{intra} aggregates all channels within region r into a single high-order relation. Subsequently, we lift channel-level PLV to a region-level similarity by averaging pairwise PLV values across all channel pairs between two regions:

$$\mathbf{S}_{rs}^{\text{reg}} = \frac{1}{|\mathcal{V}_r||\mathcal{V}_s|} \sum_{i\in\mathcal{V}_r} \sum_{j\in\mathcal{V}_s} \mathbf{S}_{ij}. \tag{6}$$

where $\mathbf{S}^{\text{reg}} \in \mathbb{R}^{R\times R}$ is the region-level similarity matrix, with $r, s \in \{1, \ldots, R\}$ indexing brain regions. This equation provides a compact measure of functional synchrony between regions by pooling the underlying PLV interactions of their constituent channels. For each region r, we identify its $k^{(3)}$ most similar regions based on region-level PLV similarity, where $k^{(3)} = 3$:

$$\mathcal{N}_{k^{(3)}}(r) = \text{TopK}\left(\{\mathbf{S}_{rs}^{\text{reg}}\}_{s\neq r}\right). \tag{7}$$

Instead of aggregating all neighboring regions into a single hyperedge, we construct pairwise inter-region hyperedges in a one-to-one manner. Specifically, for each $s \in \mathcal{N}_{k^{(3)}}(r)$, we define:

$$e_{r,s}^{\text{inter}} = \mathcal{V}_r \cup \mathcal{V}_s. \tag{8}$$

where ties are broken by selecting the region with the smaller index. It injects coarse cross-region interactions by explicitly modeling pairwise functional couplings between a central region and each of its most strongly synchronized neighboring regions. For clarity, the third level still performs message propagation over channel nodes, while the hyperedges are induced by region assignments, including both intra-region hyperedges and pairwise inter-region hyperedges. Let $\mathcal{E}^{(3)} = \{e_r^{\text{intra}}\}_{r=1}^{R} \cup \{e_{r,s}^{\text{inter}} \mid r = 1, \ldots, R,\ s \in \mathcal{N}_{k^{(3)}}(r)\}$ be the resulting region-level hyperedge set, and let $\mathbf{H}^{(3)} \in \{0,1\}^{K\times|\mathcal{E}^{(3)}|}$ be its incidence matrix. Analogous to the channel-level case, we define the normalized region-level propagation operator for the third level:

$$\mathbf{A}^{(3)} = \left(\mathbf{D}_v^{(3)}\right)^{-\frac{1}{2}} \mathbf{H}^{(3)}\mathbf{W}^{(3)} \left(\mathbf{D}_e^{(3)}\right)^{-1} \left(\mathbf{H}^{(3)}\right)^{\top} \left(\mathbf{D}_v^{(3)}\right)^{-\frac{1}{2}}. \tag{9}$$

where $\mathbf{D}_v^{(3)}$ and $\mathbf{D}_e^{(3)}$ are the node-degree and hyperedge-degree matrices for $\mathbf{H}^{(3)}$, and $\mathbf{W}^{(3)} \in \mathbb{R}^{|\mathcal{E}^{(3)}|\times|\mathcal{E}^{(3)}|}$ is a fixed diagonal hyperedge-weight matrix with unit weights for all hyperedges.

Before applying graph neural networks, we directly use the PLV matrix as the node feature source. Specifically, for each channel, its PLV-based connectivity profile with respect to all other channels is treated as the initial node feature vector. This yields an initial node feature matrix $\mathbf{X}_0 = \mathbf{S} \in \mathbb{R}^{K\times K}$. After constructing the three propagation operators $\{\mathbf{A}^{(m)}\}_{m=1}^{3}$, we employ a single HGNN to learn multi-scale representations by performing message passing in parallel on all views. Specifically, given the initial features $\mathbf{X}_0 \in \mathbb{R}^{K\times F}$, where $K = F$, for view m, the node representation at layer ℓ is updated by:

$$\mathbf{X}_{\ell+1}^{(m)} = \sigma\Big(\mathbf{A}^{(m)}\mathbf{X}_\ell^{(m)}\mathbf{W}_\ell\Big), \quad m \in \{1,2,3\},\ \ell = 0, \ldots, L-1. \tag{10}$$

where $\mathbf{W}_\ell \in \mathbb{R}^{d_\ell\times d_{\ell+1}}$ is the trainable weight matrix of layer ℓ, that this weight matrix are shared the across views. $\sigma(\cdot)$ denotes the ReLU activation function, and L is the total number of message-passing layers (set to $L = 2$ in our experiments). The final hidden dimension is denoted as $F' = d_L$. Subsequently, we explicitly preserve level-specific node tokens and concatenate all node embeddings from different views. Let $\mathbf{X}_L^{(m)} \in \mathbb{R}^{K\times F'}$ denote the final-layer node embedding matrix of view m. We first stack the three embedding matrices along the node dimension to obtain a unified set:

$$\mathbf{H} = \begin{bmatrix} \mathbf{X}_L^{(1)} \\ \mathbf{X}_L^{(2)} \\ \mathbf{X}_L^{(3)} \end{bmatrix} \in \mathbb{R}^{3K\times F'}. \tag{11}$$

where $\mathbf{H}$ contains $3K$ node tokens, i.e., K nodes from each view. We then flatten $\mathbf{H}$ into a single vector and feed it into the classification head as: $\mathbf{h} = \mathrm{vec}(\mathbf{H}) \in \mathbb{R}^{3KF'}$, where $\mathrm{vec}(\cdot)$ denotes the vectorization operator that concatenates all rows of $\mathbf{H}$ into one long vector. Finally, the predicted class probability vector $\hat{\mathbf{y}} \in \mathbb{R}^C$ is obtained by an MLP classifier $g(\cdot)$: $\hat{\mathbf{y}} = \mathrm{softmax}(g(\mathbf{h}))$, where C is the number of classes, and $g(\cdot)$ is a two-layer MLP, ReLU activation, and dropout. The model is trained leveraging the cross-entropy loss:

$$\mathcal{L}_{\mathrm{CE}} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c), \tag{12}$$

where $\mathbf{y} = [y_1, \ldots, y_C]^\top$ is the one-hot encoded ground-truth label vector. This objective encourages the proposed multi-view hypergraph GNN to learn discriminative multi-scale representations for downstream classification.

2.2 Hierarchical Functional Role Alignment

To further promote consistency across the hierarchical views, we introduce an alignment mechanism that quantifies node roles using spectral and geometric properties and minimizes their discrepancies. For each view $m \in \{1, 2, 3\}$, we first compute the symmetric normalized hypergraph Laplacian $\mathbf{L}^{(m)} = \mathbf{I} - \mathbf{A}^{(m)}$, where $\mathbf{I}$ is the $K \times K$ identity matrix and $\mathbf{A}^{(m)}$ is the propagation operator defined earlier. This Laplacian encodes the structural connectivity for diffusion processes. Given the moderate number of channels K, we perform full spectral decomposition of the Laplacian:

$$\mathbf{L}^{(m)} = \mathbf{U}^{(m)} \mathbf{\Lambda}^{(m)} (\mathbf{U}^{(m)})^\top, \tag{13}$$

where $\mathbf{U}^{(m)} \in \mathbb{R}^{K \times K}$ is the matrix of orthonormal eigenvectors and $\mathbf{\Lambda}^{(m)} = \mathrm{diag}(\lambda_1^{(m)}, \ldots, \lambda_K^{(m)})$ with eigenvalues $\lambda_n^{(m)} \geq 0$ for $n = 1, \ldots, K$. This decomposition diagonalizes the Laplacian into its frequency components. We then construct the fractional-order Laplacian to emphasize long-range dependencies:

$$(\mathbf{L}^{(m)})^\alpha = \mathbf{U}^{(m)} (\mathbf{\Lambda}^{(m)})^\alpha (\mathbf{U}^{(m)})^\top, \tag{14}$$

where $(\mathbf{\Lambda}^{(m)})^\alpha = \mathrm{diag}((\lambda_1^{(m)})^\alpha, \ldots, (\lambda_K^{(m)})^\alpha)$ and $\alpha \in (0, 1)$ is a fractional exponent. Using the learned node embeddings as features, we compute a Fisher-inspired diffusion sensitivity score for each node v_i ($i = 1, \ldots, K$):

$$F^{(m)}(v_i) = \left\| [(\mathbf{L}^{(m)})^\alpha \mathbf{X}_L^{(m)}]_{i,:} \right\|_2^2, \tag{15}$$

where $[\cdot]_{i,:}$ extracts the i-th row as a vector in $\mathbb{R}^{F'}$ and $\| \cdot \|_2^2$ is the squared Euclidean norm. This quantity measures the energy of the node representation under fractional Laplacian diffusion, reflecting the degree to which node features vary with respect to the non-local structural operator. It measures the aggregated feature dispersion under fractional diffusion across all feature dimensions, quantifying how the node representation is affected by non-local structure.

Complementarily, we incorporate discrete geometric insights via Forman-Ricci curvature. For views $m \in \{2, 3\}$, $\mathcal{E}^{(m)}$ denotes the corresponding hyperedge set defined in the hierarchical construction. For the first view ($m = 1$), although $\mathbf{A}^{(1)}$ is constructed from pairwise PLV similarities, we interpret it as a 2-uniform hypergraph by treating each graph edge as a hyperedge connecting two channel nodes. Accordingly, we define $\mathcal{E}^{(1)} = \{(v_i, v_j) \mid \mathbf{S}_{ij} > 0,\ i < j\}$, and construct the incidence matrix $\mathbf{H}^{(1)} \in \{0, 1\}^{K \times |\mathcal{E}^{(1)}|}$, where $\mathbf{H}_{u,e}^{(1)} = 1$ if node u belongs to edge e and 0 otherwise. We set $\mathbf{W}^{(1)}$ as a diagonal matrix with unit weights. Let $\mathcal{E}^{(m)}(i) = \{e \in \mathcal{E}^{(m)} : v_i \in e\}$ denote the set of hyperedges incident to v_i. For each such pair (v_i, e), the node-hyperedge curvature is:

$$\kappa^{(m)}(v_i, e) = w_i w_e^{(m)} + \sum_{v_j \in e \setminus \{v_i\}} w_j w_e^{(m)} - \sum_{e' \in \mathcal{E}^{(m)}(i) \setminus \{e\}} w_i \sqrt{w_e^{(m)} w_{e'}^{(m)}}, \tag{16}$$

where $w_i = 1$ and $w_j = 1$ are uniform node weights (set to unity by default), and $w_e^{(m)}$ (resp. $w_{e'}^{(m)}$) denotes the weight of hyperedge e (resp. e'), corresponding to the diagonal entry $[\mathbf{W}^{(m)}]_{e,e}$ of the fixed hyperedge-weight matrix $\mathbf{W}^{(m)}$, with $w_e^{(m)} = 1$ for all hyperedges. This formula quantifies

local dispersion versus concentration in the hypergraph's combinatorial structure. The node-level curvature is then averaged over incident hyperedges:

$$\kappa^{(m)}(v_i) = \frac{1}{|\mathcal{E}^{(m)}(i)|} \sum_{e \in \mathcal{E}^{(m)}(i)} \kappa^{(m)}(v_i, e), \tag{17}$$

where $|\mathcal{E}^{(m)}(i)|$ is the hyperdegree of node v_i in view m. The number of hyperedges incident to v_i, which corresponds to the i-th diagonal entry of the degree matrix $\mathbf{D}_v^{(m)}$. This scalar summarizes the geometric properties of the node's neighborhood. Integrating sensitivity and geometry, we define the role index for node v_i in view m as:

$$\Psi^{(m)}(v_i) = \exp(-\beta \kappa^{(m)}(v_i)) \cdot F^{(m)}(v_i), \tag{18}$$

where $\beta > 0$ is a curvature scaling coefficient (set to $\beta = 1.0$ in our experiments). This index modulates the Fisher-inspired diffusion sensitivity by an exponential decay based on curvature, characterizing the node's functional role in the view. To align roles across views, we compute the pairwise discrepancy loss:

$$\mathcal{L}_{\text{HFRA}} = \frac{1}{3K} \sum_{(m,m') \in P} \sum_{i=1}^{K} \left| \Psi^{(m)}(v_i) - \Psi^{(m')}(v_i) \right|, \tag{19}$$

where $P = \{(1,2), (1,3), (2,3)\}$ denotes the set of all distinct view-index pairs. This loss averages the absolute differences over all nodes and view pairs, encouraging coherent multi-view node interpretations. The total training objective combines classification and alignment: $\mathcal{L} = \mathcal{L}_{\text{CE}} + \gamma \mathcal{L}_{\text{HFRA}}$, where $\gamma > 0$ is a balancing hyperparameter that controls the trade-off between task-specific learning and multi-view consistency (set to $\gamma = 0.1$ in our experiments).

3 Experiments

3.1 Experimental Settings

Baselines We compare HHRA with multiple baselines, including several SOTA brain time-series signal learning and brain graph learning approaches. The baselines include: (1) K-Nearest Neighbors (KNN) and Random Forest: Two traditional machine learning models. (2) GroupINN (Yan et al., 2019): A grouping-based interpretable neural network for functional subnetwork identification. (3) IGS (Li et al., 2023): An interpretable brain graph sparsification framework that learns a task-relevant joint edge mask. (4) EvoBrain (Kotoge et al., 2025): A dynamic graph neural network utilizing a time-then-graph strategy. (5) Medformer (Wang et al., 2024c): A multi-granularity patching Transformer with cross-channel attention. (6) DMSGL (Pan et al., 2025): A dynamic mixed spatiotemporal brain graph network for real-time analysis. (7) FSTA-EC (Xiong et al., 2025): A Fourier spatiotemporal attention model for robust brain causal dependency extraction.(8) TarDiff (Deng et al., 2025): A target-oriented diffusion framework for brain time-series signal synthetic data generation. Since its performance on synthetic data surpasses that on real data, we include it as a data augmentation method for comparison. (9) LaBraM (Jiang et al.): A large-scale EEG pre-training model based on neural tokenization.(10) EEGPT (Wang et al., 2024a): A self-supervised Transformer for brain spatiotemporal feature alignment.

Datasets. To evaluate the effectiveness of HHRA, we conduct binary classification experiments on both the EEG and the fNIRS datasets. Datasets include three publicly available electroencephalography datasets: ADFTD (Miltiadous et al., 2023), APAVA (Escudero et al., 2006), and TDBrain (Van Dijk et al., 2022), and one private proprietary functional near-infrared spectroscopy dataset, M-fNIRS. They vary in imaging modality, number of channels, and number of subjects, providing a comprehensive testbed for evaluating our method across diverse brain signal characteristics. Detailed descriptions of these datasets are provided in Appendix B.

Evaluation Metrics. All tasks are formulated as binary classification. Following previous brain time series learning research (Deng et al., 2025), we adopt two threshold-free metrics: the Area Under the Receiver Operating Characteristic curve (AUROC) and the Area Under the Precision–Recall curve (AUPRC). Detailed mathematical formulations and definitions of these metrics are provided in Appendix D.

Table 1: Comparison with state-of-the-art methods on four datasets. AUROC: Area Under the Receiver Operating Characteristic curve (↑), AUPRC: Area Under the Precision-Recall Curve (↑). The best result is in **bold** and the second best is underlined.

Dataset	ADFTD		APAVA		TDBrain		M-fNIRS	
Method	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
KNN	49.92 ± 1.86	65.05 ± 2.70	39.30 ± 5.64	46.97 ± 2.70	70.83 ± 7.22	88.33 ± 2.89	52.12 ± 3.12	37.39 ± 4.88
Random Forest	57.84 ± 3.48	71.37 ± 2.99	44.35 ± 3.76	44.82 ± 5.10	66.39 ± 2.75	77.50 ± 4.13	$\underline{75.82 \pm 1.41}$	59.59 ± 4.59
GroupINN[KDD19]	58.50 ± 3.57	72.58 ± 2.89	64.81 ± 2.49	67.70 ± 3.75	81.59 ± 4.08	45.47 ± 2.83	62.82 ± 3.19	39.66 ± 5.93
IGS[KDD23]	67.52 ± 4.72	$\underline{76.00 \pm 6.68}$	46.32 ± 5.46	65.30 ± 4.23	90.60 ± 3.48	93.57 ± 2.51	75.27 ± 1.49	61.57 ± 0.75
LaBraM[ICLR24]	51.62 ± 1.03	66.23 ± 2.02	51.20 ± 2.01	38.48 ± 5.44	58.47 ± 4.16	46.25 ± 4.50	50.05 ± 0.11	31.95 ± 4.67
EEGPT[NeurIPS24]	50.20 ± 0.65	64.47 ± 2.68	74.25 ± 4.52	75.36 ± 2.40	41.20 ± 8.68	55.75 ± 5.68	48.97 ± 3.37	40.12 ± 4.11
Medformer[NeurIPS24]	$\underline{76.42 \pm 3.82}$	72.06 ± 5.42	83.72 ± 5.22	81.84 ± 6.79	$\mathbf{97.54 \pm 2.96}$	$\underline{97.23 \pm 3.41}$	75.15 ± 4.25	$\underline{72.79 \pm 2.08}$
DMSGL[ACM MM25]	52.63 ± 2.58	68.14 ± 1.88	67.63 ± 3.96	69.16 ± 4.66	75.77 ± 3.42	88.66 ± 1.30	50.34 ± 1.87	35.44 ± 1.15
EvoBrain[NeurIPS25]	49.73 ± 0.53	66.15 ± 0.98	$\underline{87.32 \pm 3.51}$	$\underline{93.29 \pm 5.69}$	61.33 ± 1.50	85.64 ± 5.71	56.28 ± 3.97	40.26 ± 6.87
FSTA-EC[KDD25]	47.31 ± 2.15	66.11 ± 1.31	64.82 ± 3.03	64.04 ± 2.74	51.79 ± 4.72	52.38 ± 6.30	49.96 ± 0.87	29.36 ± 2.89
TarDiff[KDD25]	49.73 ± 0.23	66.33 ± 0.74	77.39 ± 4.61	76.84 ± 2.02	62.89 ± 2.98	65.36 ± 6.30	48.86 ± 1.94	30.09 ± 2.01
Ours	$\mathbf{80.65 \pm 4.32}$	$\mathbf{88.75 \pm 3.43}$	$\mathbf{92.22 \pm 0.72}$	$\mathbf{94.21 \pm 4.39}$	$\underline{97.29 \pm 4.39}$	$\mathbf{97.94 \pm 1.40}$	$\mathbf{88.04 \pm 1.26}$	$\mathbf{79.83 \pm 4.23}$

Implementation Details. All datasets use binary classification tasks. For ADFTD, healthy control subjects are labeled as class 0, while both subjects with Alzheimer's disease and subjects with frontotemporal dementia are merged into class 1. For APAVA, patients with Alzheimer's disease are labeled as class 1, and healthy controls are labeled as class 0. For TDBrain, healthy control subjects are labeled as class 0, and subjects with Parkinson's disease are labeled as class 1. For M-fNIRS, healthy controls are labeled class 0, and disorder categories are labeled class 1. For our method, HyperGCNN (Yadati et al., 2019) is employed as the backbone architecture. For dataset splitting, we employ a subject-independent setup across all datasets to prevent data leakage. Following prior work on the same public datasets (Deng et al., 2025), all subject IDs are shuffled and split into training, validation, and test sets with a ratio of 8:1:1, with all samples of a subject assigned to the same split; for ADFTD, the split is stratified within each diagnostic category for balanced class representation. Since M-fNIRS samples are not organized by subject, they are pooled, shuffled with a fixed seed, and split at the sample level in an 8:1:1 ratio. All experiments ran on four NVIDIA 4090 GPUs, and results are averaged over five random seeds.

3.2 Performance Comparison

Tab. 1 summarizes the performance comparison between the our framework and multiple state-of-the-art baselines across four evaluation datasets. The results show that our method achieves superior performance in terms of $AUROC$ and $AUPRC$ metrics in the majority of experimental settings.

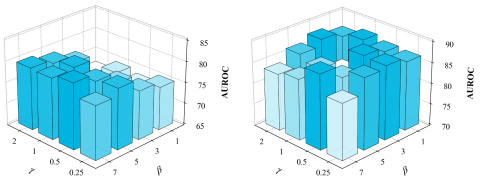
3.3 Ablation Study

Effects of Multi-Level Configuration in HHKC. To investigate the contribution of each level within HHKC, ablation studies are conducted on ADFTD and M-fNIRS, where the channel-level pairwise graph, the channel-level hypergraph, and the brain-region-level hypergraph correspond to the three anatomical levels. As illustrated in Tab. 2, any single level alone results in limited performance; combining two levels consistently improves both AUROC and AUPRC over the corresponding single-level configuration, and integrating all three levels achieves the best results among configurations that exclude the alignment loss. This confirms that complementary level-specific interaction knowledge is essential for effective brain disorder diagnosis.

Effects of HFRA. The role of HFRA is further examined by pairing it with each two-level and three-level configuration. As demonstrated in Tab. 2, the incorporation of HFRA improves performance across the majority of multi-level configurations, confirming that the alignment promotes functional role consistency across levels. In conclusion, Hierarchical High-order Knowledge Capturing and Hierarchical Functional Role Alignment are complementary: the former captures diverse structural patterns through three distinct interaction knowledge capturing levels, while the latter enforces cross-level functional consistency. Therefore, Hierarchical High-order Knowledge Capturing and HFRA effectively address the dual challenges posed by hierarchical anatomical heterogeneity.

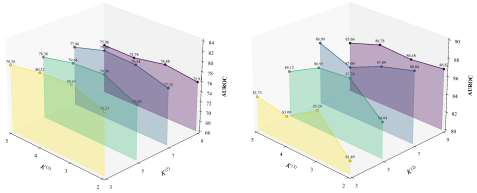
Table 2: Ablation study of key components of our method on ADFTD and M-fNIRS. Columns 1, 2, and 3 indicate whether level 1, level 2, and level 3 are enabled, respectively. Column HFRA indicates whether the functional role alignment loss is applied.

HHKC			HFRA	ADFTD		M-fNIRS	
Level 1	Level 2	Level 3		AUROC	AUPRC	AUROC	AUPRC
✓	✗	✗	✗	77.10 ± 2.45	84.75 ± 3.56	83.65 ± 0.63	73.42 ± 4.09
✗	✓	✗	✗	77.37 ± 2.16	85.14 ± 1.24	86.98 ± 2.47	76.56 ± 8.07
✗	✗	✓	✗	76.68 ± 2.85	84.35 ± 4.24	84.90 ± 2.05	74.08 ± 5.34
✓	✓	✗	✗	76.75 ± 5.56	85.22 ± 4.14	86.91 ± 1.39	76.80 ± 3.49
✓	✓	✗	✓	78.67 ± 2.19	86.47 ± 2.44	87.60 ± 2.29	77.63 ± 6.78
✓	✗	✓	✗	75.53 ± 3.20	81.13 ± 5.17	85.16 ± 1.57	74.73 ± 4.88
✓	✗	✓	✓	75.95 ± 3.03	83.95 ± 4.67	85.52 ± 1.94	75.24 ± 5.76
✗	✓	✓	✗	77.11 ± 4.04	84.59 ± 1.83	85.75 ± 2.19	73.43 ± 1.31
✗	✓	✓	✓	79.28 ± 3.56	87.39 ± 4.55	87.05 ± 2.05	78.53 ± 3.39
✓	✓	✓	✗	79.51 ± 0.64	88.13 ± 2.31	87.71 ± 2.08	$\mathbf{80.69 \pm 3.73}$
✓	✓	✓	✓	$\mathbf{80.65 \pm 4.32}$	$\mathbf{88.75 \pm 3.43}$	$\mathbf{88.04 \pm 1.26}$	79.83 ± 4.23



(a) ADFTD (b) M-fNIRS

Figure 3: Sensitivity of γ and β.



(a) ADFTD (b) M-fNIRS

Figure 4: Sensitivity of $k^{(2)}$ and $k^{(3)}$.

3.4 Hyper-parameter Study

Sensitivity to γ and β. We evaluate the joint influence of the alignment loss weight γ and the Ricci curvature coefficient β in the Geometric-Fisher role index, which control the intensity and geometric awareness of cross-level alignment. AUROC results on ADFTD and M-fNIRS (Fig. 3) show that HHRA maintains competitive performance across a broad range of γ and β values.

Sensitivity to Hypergraph Construction Parameters $k^{(2)}$ and $k^{(3)}$. We evaluate the sensitivity of HHRA to the KNN parameters $k^{(2)}$ and $k^{(3)}$, which govern the construction of channel- and region-induced hypergraphs and dictate connectivity density. AUROC results for ADFTD and M-fNIRS (Fig. 4) demonstrate that HHRA achieves consistently high performance across a wide range of $k^{(2)}$ and $k^{(3)}$ settings.

4 Conclusion

In this work, we address the largely overlooked challenge of hierarchical anatomical heterogeneity in GNN-based brain disorder diagnosis, which manifests as high-order interaction knowledge hierarchy heterogeneity and element functional role hierarchy heterogeneity. Existing methods fail to adequately capture level-specific high-order interaction knowledge and suffer from knowledge conflicts caused by functional role inconsistency across anatomical levels. To tackle these challenges, we propose HHRA, a hierarchical hypergraph learning framework that integrates Hierarchical High-order Knowledge Capturing (HHKC) to jointly model multi-level pairwise and high-order interactions, and Hierarchical Functional Role Alignment (HFRA) to align functional roles across levels for mitigating knowledge conflicts and improving cross-level generalizability. Extensive experiments on EEG and fNIRS datasets demonstrate that HHRA consistently outperforms state-of-the-art graph- and hypergraph-based approaches, validating its effectiveness.

AI USE STATEMENT

We used generative AI tools for polishing the language and improving the readability of this manuscript, without altering the intended meaning of the original text. All AI-assisted writing was reviewed by the authors, and we take responsibility for the final content of this work.

REFERENCES

Sergi Abadal, Pablo Galván, Alberto Mármol, Nadia Mammone, Cosimo Ieracitano, Michele Lo Giudice, Alessandro Salvini, and Francesco Carlo Morabito. Graph neural networks for electroencephalogram analysis: Alzheimer's disease and epilepsy use cases. *Neural Networks*, pp. 106792, 2025.

A Paul Alivisatos, Anne M Andrews, Edward S Boyden, Miyoung Chun, George M Church, Karl Deisseroth, John P Donoghue, Scott E Fraser, Jennifer Lippincott-Schwartz, Loren L Looger, et al. Nanotools for neuroscience and brain activity mapping. *ACS nano*, pp. 1850–1866, 2013.

Mohammad Amini, Zhanguang Zhang, Surya Penmetsa, Yingxue Zhang, Jianye Hao, and Wulong Liu. Generalizable floorplanner through corner block list representation and hypergraph embedding. In *SIGKDD*, pp. 2692–2702, 2022.

Alessia Antelmi, Gennaro Cordasco, Daniele De Vinco, Valerio Di Pasquale, Mirko Polato, and Carmine Spagnuolo. Hypergraph motif representation learning. In *SIGKDD*, pp. 13–24, 2025.

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

Danielle S Bassett and Olaf Sporns. Network neuroscience. *Nature neuroscience*, pp. 353–364, 2017.

Danielle S Bassett, Perry Zurn, and Joshua I Gold. On the nature and use of models in network neuroscience. *Nature Reviews Neuroscience*, pp. 566–578, 2018.

Richard F Betzel and Danielle S Bassett. Multi-scale brain networks. *Neuroimage*, pp. 73–83, 2017.

Deyu Bo, Chuan Shi, Lele Wang, and Renjie Liao. Specformer: Spectral graph neural networks meet transformers. *arXiv preprint arXiv:2303.01028*, 2023.

Samuel P Burns, Sabato Santaniello, Robert B Yaffe, Christophe C Jouny, Nathan E Crone, Gregory K Bergey, William S Anderson, and Sridevi V Sarma. Network dynamics of the brain and influence of the epileptic seizure onset zone. *PNAS*, pp. E5321–E5330, 2014.

Hongjie Chen, Ryan A Rossi, Sungchul Kim, Kanak Mahadik, and Hoda Eldardiry. Probabilistic hypergraph recurrent neural networks for time-series forecasting. In *SIGKDD*, pp. 82–93, 2025a.

Jiawen Chen, Qi Shao, Duxin Chen, and Wenwu Yu. Decoupling spatio-temporal prediction: When lightweight large models meet adaptive hypergraphs. In *SIGKDD*, pp. 167–178, 2025b.

Zhaoliang Chen, Zhihao Wu, William K Cheung, Hong-Ning Dai, Byron Choi, and Jiming Liu. Mshtrans: Multi-scale hypergraph transformer with time-series decomposition for temporal anomaly detection. In *SIGKDD*, pp. 274–285, 2025c.

Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive universal generalized pagerank graph neural network. In *ICLR*, 2021.

Minyoung Choe, Sunwoo Kim, Jaemin Yoo, and Kijung Shin. Classification of edge-dependent labels of nodes in hypergraphs. In *SIGKDD*, pp. 298–309, 2023.

Bowen Deng, Chang Xu, Hao Li, Yu-hao Huang, Min Hou, and Jiang Bian. Tardiff: Target-oriented diffusion guidance for synthetic electronic health record time series generation. In *SIGKDD*, pp. 474–485, 2025.

Haipeng Ding, Zhewei Wei, and Yuhang Ye. Large-scale spectral graph neural networks via laplacian sparsification: Technical report. In *SIGKDD*, 2025.

Manh Tuan Do and Kijung Shin. Improving the core resilience of real-world hypergraphs. *Data Mining and Knowledge Discovery*, pp. 2438–2493, 2023.

Manh Tuan Do and Kijung Shin. Unsupervised alignment of hypergraphs with different scales. In *SIGKDD*, pp. 609–620, 2024.

Yihe Dong, Will Sawin, and Yoshua Bengio. Hnhn: Hypergraph networks with hyperedge neurons. *arXiv preprint arXiv:2006.12278*, 2020.

Wenying Duan, Shujun Guo, Zimu Zhou, Wei Huang, Hong Rao, and Xiaoxi He. Dynamic localisation of spatial-temporal graph neural network. In *SIGKDD*, pp. 283–294, 2025.

Oana M Dumitrascu, Xin Li, Wenhui Zhu, Bryan K Woodruff, Simona Nikolova, Jacob Sobczak, Amal Youssef, Siddhant Saxena, Janine Andreev, Richard J Caselli, et al. Color fundus photography and deep learning applications in alzheimer disease. *Mayo Clinic Proceedings: Digital Health*, pp. 548–558, 2024.

J Escudero, Daniel Abásolo, Roberto Hornero, Pedro Espino, and Miguel López. Analysis of electroencephalograms in alzheimer's disease patients with multiscale entropy. *Physiological measurement*, pp. 1091, 2006.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *AAAI*, pp. 3558–3565, 2019.

Yue Gao, Xiangling Fu, Tianxiong Ouyang, and Yi Wang. Eeg-gcn: Spatio-temporal and self-adaptive graph convolutional networks for single and multi-view eeg-based emotion recognition. *IEEE SPL*, pp. 1574–1578, 2022.

Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized pagerank. *arXiv preprint arXiv:1810.05997*, 2018.

Manuel Graña and Igone Morais-Quilez. A review of graph neural networks for electroencephalography data analysis. *Neurocomputing*, pp. 126901, 2023.

Xiangmin Han, Rundong Xue, Jingxi Feng, Yifan Feng, Shaoyi Du, Jun Shi, and Yue Gao. Hypergraph foundation model for brain disease diagnosis. *IEEE TNNLS*, 2025.

Mingguo He, Zhewei Wei, Hongteng Xu, et al. Bernnet: Learning arbitrary graph spectral filters via bernstein approximation. In *NeurIPS*, pp. 14239–14251, 2021.

Mingguo He, Zhewei Wei, and Ji-Rong Wen. Convolutional neural networks on graphs with chebyshev approximation, revisited. In *NeurIPS*, pp. 7264–7276, 2022.

Yimin Hou, Shuyue Jia, Xiangmin Lun, Ziqian Hao, Yan Shi, Yang Li, Rui Zeng, and Jinglei Lv. Gcns-net: a graph convolutional neural network approach for decoding time-resolved eeg motor imagery signals. *IEEE TNNLS*, pp. 7312–7323, 2022.

Zheng Huang, Qihui Yang, Dawei Zhou, and Yujun Yan. Enhancing size generalization in graph neural networks through disentangled representation learning. *arXiv preprint arXiv:2406.04601*, 2024.

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

Shuyi Ji, Yifan Feng, Rongrong Ji, Xibin Zhao, Wanwan Tang, and Yue Gao. Dual channel hypergraph collaborative filtering. In *SIGKDD*, pp. 2020–2029, 2020.

Weibang Jiang, Liming Zhao, and Bao-liang Lu. Large brain model for learning generic representations with tremendous eeg data in bci. In *ICLR*.

Rikuto Kotoge, Zheng Chen, Tasuku Kimura, Yasuko Matsubara, Takufumi Yanagisawa, Haruhiko Kishima, and Yasushi Sakurai. Evobrain: Dynamic multi-channel eeg graph modeling for time-evolving brain networks. In *NeurIPS*, 2025.

Gaotang Li, Marlena Duda, Xiang Zhang, Danai Koutra, and Yujun Yan. Interpretable sparsification of brain graphs: Better practices and effective designs for graph neural networks. In *SIGKDD*, pp. 1223–1234, 2023.

Yonghui Li, Yong Liu, Jun Li, Wen Qin, Kuncheng Li, Chunshui Yu, and Tianzi Jiang. Brain anatomical network and intelligence. *PLoS computational biology*, pp. e1000395, 2009.

Renjie Liao, Zhizhen Zhao, Raquel Urtasun, and Richard S Zemel. Lanczosnet: Multi-scale deep graph convolutional networks. In *ICLR*, 2019.

Chenyu Liu, Xinliang Zhou, Jiaping Xiao, Zhengri Zhu, Liming Zhai, Ziyu Jia, and Yang Liu. Vsgt: variational spatial and gaussian temporal graph models for eeg-based emotion recognition. In *IJCAI*, pp. 3078–3086, 2024a.

Jin Liu, Min Li, Yi Pan, Wei Lan, Ruiqing Zheng, Fang-Xiang Wu, and Jianxin Wang. Complex brain network analysis and its applications to brain disorders: a survey. *Complexity*, pp. 8362741, 2017.

Xuyuan Liu, Yinghao Cai, Qihui Yang, and Yujun Yan. Exploring consistency in graph representations: from graph kernels to graph neural networks. *NeurIPS*, pp. 136393–136423, 2024b.

Jielong Lu, Zhihao Wu, Luying Zhong, Zhaoliang Chen, Hong Zhao, and Shiping Wang. Generative essential graph convolutional network for multi-view semi-supervised classification. *IEEE TMM*, pp. 7987–7999, 2024.

Yuankai Luo, Lei Shi, and Xiao-Ming Wu. Classic gnns are strong baselines: Reassessing gnns for node classification. In *NeurIPS*, pp. 97650–97669, 2024.

Zihan Luo, Hong Huang, Jianxun Lian, Xiran Song, and Hai Jin. Towards controllable hybrid fairness in graph neural networks. In *SIGKDD*, pp. 950–961, 2025.

Andreas Miltiadous, Katerina D Tzimourta, Theodora Afrantou, Panagiotis Ioannidis, Nikolaos Grigoriadis, Dimitrios G Tsalikakis, Pantelis Angelidis, Markos G Tsipouras, Euripidis Glavas, Nikolaos Giannakeas, et al. A dataset of scalp eeg recordings of alzheimer's disease, frontotemporal dementia and healthy subjects from routine eeg. *Data*, pp. 95, 2023.

Yue Pan, Cunbo Li, Peiyang Li, Fali Li, Feng Wan, Dezhong Yao, Zehong Cao, and Peng Xu. Real-time eeg emotion recognition from dynamic mixed spatiotemporal graph learning. In *ACMMM*, pp. 5697–5706, 2025.

Marios Papachristou and Jon Kleinberg. Core-periphery models for hypergraphs. In *SIGKDD*, pp. 1337–1347, 2022.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. In *ICLR*, 2021.

QingGuo Qi, Hongyang Chen, Minhao Cheng, and Han Liu. Input snapshots fusion for scalable discrete-time dynamic graph neural networks. In *SIGKDD*, pp. 1138–1149, 2025.

Xinmei Qiu, Fan Wang, Yongheng Sun, Chunfeng Lian, and Jianhua Ma. Towards graph neural networks with domain-generalizable explainability for fmri-based brain disorder diagnosis. In *MICCAI*, pp. 454–464, 2024.

Siddhant Saxena, Shounak Ghatak, Raghu Kolla, Debashis Mukherjee, and Tanmoy Chakraborty. Dphgnn: A dual perspective hypergraph neural networks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2548–2559, 2024.

Junhao Shen, Hong Qian, Shuo Liu, Wei Zhang, Bo Jiang, and Aimin Zhou. Capturing homogeneous influence among students: Hypergraph cognitive diagnosis for intelligent education systems. In *SIGKDD*, pp. 2628–2639, 2024.

Wei Tong, Yong-Xia Li, Xiao-Yan Zhao, Qi-Qi Chen, Yu-Bing Gao, Ping Li, and Edmond Q Wu. fmri-based brain disease diagnosis: A graph network approach. *IEEE TMRB*, pp. 312–322, 2023.

Hanneke Van Dijk, Guido Van Wingen, Damiaan Denys, Sebastian Olbrich, Rosalinde Van Ruth, and Martijn Arns. The two decades brainclinics research archive for insights in neurophysiology (tdbrain) database. *Scientific data*, pp. 333, 2022.

Guangyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. Eegpt: Pretrained transformer for universal and reliable representation of eeg signals. *NeurIPS*, pp. 39249–39280, 2024a.

Shiping Wang, Zhaoliang Chen, Shide Du, and Zhouchen Lin. Learning deep sparse regularizers with applications to multi-view clustering and semi-supervised classification. *IEEE TPAMI*, pp. 5042–5055, 2021.

Xiyuan Wang and Muhan Zhang. How powerful are spectral graph neural networks. In *ICML*, pp. 23341–23362, 2023.

Xuan Wang, Xiaotong Zhang, Yang Chen, and Xiaopeng Yang. Ifc-gnn: Combining interactions of functional connectivity with multimodal graph neural networks for asd brain disorder analysis. *Alexandria Engineering Journal*, 2024b.

Yihe Wang, Nan Huang, Taida Li, Yujun Yan, and Xiang Zhang. Medformer: A multi-granularity patching transformer for medical time-series classification. *NeurIPS*, pp. 36314–36341, 2024c.

Joanna M Wardlaw, Helene Benveniste, Maiken Nedergaard, Berislav V Zlokovic, Humberto Mestre, Hedok Lee, Fergus N Doubal, Rosalind Brown, Joel Ramirez, Bradley J MacIntosh, et al. Perivascular spaces in the brain: anatomy, physiology and pathology. *Nature Reviews Neurology*, pp. 137–153, 2020.

Hanrui Wu, Yuguang Yan, and Michael Kwok-Po Ng. Hypergraph collaborative network on vertices and hyperedges. *IEEE TPAMI*, pp. 3245–3258, 2022.

Yushun Xiao, Wenming Zheng, and Guoying Zhao. Dynamical causal graph neural network for eeg emotion recognition. *IEEE TAC*, 2025.

Wen Xiong, Jinduo Liu, Junzhong Ji, and Fenglong Ma. Brain effective connectivity estimation via fourier spatiotemporal attention. In *SIGKDD*, pp. 1657–1668, 2025.

Yuanyuan Xu, Wenjie Zhang, Ying Zhang, Xiwei Xu, and Xuemin Lin. Fast and accurate temporal hypergraph representation for hyperedge prediction. In *SIGKDD*, pp. 1727–1738, 2025.

Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. Hypergcn: A new method for training graph convolutional networks on hypergraphs. In *NeurIPS*, 2019.

Yujun Yan, Jiong Zhu, Marlena Duda, Eric Solarz, Chandra Sripada, and Danai Koutra. Groupinn: Grouping-based interpretable neural network for classification of limited, noisy brain data. In *SIGKDD*, pp. 772–782, 2019.

Yujun Yan, Milad Hashemi, Kevin Swersky, Yaoqing Yang, and Danai Koutra. Two sides of the same coin: Heterophily and oversmoothing in graph convolutional neural networks. In *ICDM*, pp. 1287–1292, 2022.

Yanwu Yang, Chenfei Ye, Xutao Guo, Tao Wu, Yang Xiang, and Ting Ma. Mapping multi-modal brain connectome for brain disorder diagnosis via cross-modal mutual learning. *IEEE Transactions on Medical Imaging*, pp. 108–121, 2023.

Gongzhu Yin, Hongli Zhang, Yuchen Yang, and Yi Luo. Inductive link prediction on n-ary relational facts via semantic hypergraph reasoning. In *SIGKDD*, pp. 1821–1832, 2025.

Qiao Yu, Rui Wang, Jia Liu, Long Hu, Min Chen, and Zhongchun Liu. Gnn-based depression recognition using spatio-temporal information: a fnirs study. *IEEE JBHI*, pp. 4925–4935, 2022.

Hao Zhang, Ran Song, Liping Wang, Lin Zhang, Dawei Wang, Cong Wang, and Wei Zhang. Classification of brain disorders in rs-fmri via local-to-global graph neural networks. *IEEE TMI*, pp. 444–455, 2022.

Zhen Zhang, Jiajun Bu, Martin Ester, Zhao Li, Chengwei Yao, Zhi Yu, and Can Wang. H2mn: Graph similarity learning with hierarchical hypergraph matching networks. In *SIGKDD*, pp. 2274–2284, 2021.

Tao Zhao, Yunlong Xue, Jie Zhu, and Wenming Zheng. Heterogeneous graph convolutional neural networks for eeg-fnirs bimodal emotion recognition. In *ICASSP*, 2025.

Tianyi Zhao and Gaoyan Zhang. Enhancing major depressive disorder diagnosis with dynamic-static fusion graph neural networks. *IEEE JBHI*, pp. 4701–4710, 2024.

Jiajun Zhou, Shengbo Gong, Xuanze Chen, Chenxuan Xie, Shanqing Yu, Qi Xuan, and Xiaoniu Yang. Clarify confused nodes via separated learning. *IEEE TPAMI*, 2025.

Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. *NeurIPS*, pp. 7793–7804, 2020.

Jiong Zhu, Yujun Yan, Mark Heimann, Lingxiao Zhao, Leman Akoglu, and Danai Koutra. Heterophily and graph neural networks: Past, present and future. *IEEE Data Eng. Bull.*, 2023.

Meiqi Zhu, Xiao Wang, Chuan Shi, Houye Ji, and Peng Cui. Interpreting and unifying graph neural networks with an optimization framework. In *WWW*, pp. 1215–1226, 2021.

Yanqiao Zhu, Hejie Cui, Lifang He, Lichao Sun, and Carl Yang. Joint embedding of structural and functional brain networks with graph neural networks for mental illness diagnosis. In *IEEE EMBC*, pp. 272–276. IEEE, 2022.

APPENDIX

A Related Work

A.1 Graph Neural Network

Graph neural networks (GNNs) have attracted increasing attention for learning on structured data due to their strong ability to capture complex dependencies (Zhu et al., 2020; Yan et al., 2022; Zhu et al., 2023). In general, GNNs can be categorized into spatial-based and spectral-based approaches. Spatial GNNs operate by directly aggregating information from neighborhoods. Among them, Geom-GCN (Pei et al., 2021) performs bi-level aggregation over both structural and semantic neighborhoods while GPR-GNN (Chien et al., 2021) integrates the generalized PageRank to address challenges of heterophily and over-smoothing. TunedGNN (Luo et al., 2024) improves conventional spatial GNNs through hyperparameter tuning. NCGCN (Zhou et al., 2025) proposes neighborhood confusion-guided separated learning. Moreover, spectral GNNs typically perform graph convolution through the spectral filters. Specifically, APPNP (Gasteiger et al., 2018) and GNN-LF/HF (Zhu et al., 2021) employs fixed-filters. Bernstein (He et al., 2021), Chebyshev (He et al., 2022), and Jacobi (Wang & Zhang, 2023) approximate flexible filters using orthogonal polynomials. Other methods parameterize filters with neural networks, including LanczosNet (Liao et al., 2019) and Specformer (Bo et al., 2023). Although existing GNN studies have established a foundation for brain network modeling, their effectiveness in direct applications remains limited due to the lack of neuroscience-informed design.

A.2 Hypergraph Neural Network

Hypergraphs have attracted increasing attention due to their strong capability to model high-order relationships (Antelmi et al., 2025; Yin et al., 2025; Xu et al., 2025; Chen et al., 2025b). Therefore, HyperGraph Neural Networks (HGNN) have emerged (Chen et al., 2025c), greatly advancing the development of hypergraph learning (Chen et al., 2025a; Zhang et al., 2021; Choe et al., 2023; Do & Shin, 2024). Early on, HGNN (Feng et al., 2019) introduced the concept of hypergraph convolution. HyperGCN (Yadati et al., 2019) approximates each hyperedge using a linear number of representative pairwise edges. Subsequently, DHCF (Ji et al., 2020) leverages dual-channel collaborative filtering through jump hypergraph convolution. HNHN (Dong et al., 2020) flexibly adjusts the importance of high-cardinality hyperedges and high-degree vertices depending on the dataset. HCHA (Bai et al., 2021) learns a dynamic incidence matrix via an attention module to obtain more discriminative node embeddings. HCoN (Wu et al., 2022) learns informative latent representations and introduces the hypergraph reconstruction error as a regularizer. ED-HDD provably represents any continuous equivariant hypergraph diffusion operators. DPHGNN (Saxena et al., 2024) introduces equivariant operator learning to capture lower-order semantics. Although HGNNs provide a foundation for capturing high-order brain interactions (Papachristou & Kleinberg, 2022; Amini et al., 2022), directly applying them often overlooks the anatomically multi-scale nature of brain activity, making it difficult to achieve effective learning and reliable brain disorder diagnosis.

A.3 Brain Graph Learning

Graph learning, owing to its capability to capture relationships and dependencies, has shown great promise in brain network modeling for scenarios such as emotion recognition and disease diagnosis, and has achieved notable progress. Among existing methods, GroupINN (Yan et al., 2019) combines node grouping layer with a random walk with restart for efficient and interpretable learning EEG-GCN (Gao et al., 2022) accounts for the temporal sparsity of emotions and the spatial variability of brain topology for emotion recognition. VSGT (Liu et al., 2024a) integrates a Conditional Gaussian Graph Transform with a recurrent spatio-temporal module to model dynamic brain interactions. Another work (Xiao et al., 2025) models causal connectivity among EEG channels to improve interpretability and cross-subject generalization. HGFM (Han et al., 2025) adopts a two-stage strategy that combines self-supervised pretraining and fine-tuning from both individual-level and population-level perspectives. EvoBrain (Kotoge et al., 2025) proposes a time-then-graph framework that incorporates two-stream Mamba-based temporal modeling and Laplacian positional encoding to capture temporal dynamics and graph structures. However, they fail to adequately capture anatomically hierarchical

high-order interaction knowledge and are limited by functional role hierarchy heterogeneity. To address these challenges, HHRA investigates and develops effective solutions.

B Dataset Details

ADFTD. A public EEG time series dataset containing three categories in total: 36 patients with Alzheimer's disease, 23 patients with frontotemporal dementia, and 29 healthy control subjects, totaling 88 subjects recorded with 19 channels at a raw sampling rate of 500 Hz. Each subject has a single trial, with trial durations on average averaging approximately 13.5 minutes for Alzheimer's disease subjects, 12 minutes for frontotemporal dementia subjects, and 13.8 minutes for healthy subjects. A bandpass filter between 0.5 and 45 Hz is applied to each trial. Each trial is then downsampled to 256 Hz and segmented into non-overlapping 1-second samples of 256 timestamps, discarding any samples shorter than 1 second. It yields a total of 69,752 samples. The 19 channels are grouped into 5 brain regions covering the frontopolar-frontal, lateral frontal-temporal, central, posterior temporal, and parieto-occipital areas.

APAVA. A public EEG time series dataset comprising 23 subjects with 16 channels, including 12 patients with Alzheimer's disease and 11 healthy control subjects. On average, each subject has 30.0 ± 12.5 trials, where each trial is a 5-second time sequence of 1,280 timestamps across 16 channels. Before further preprocessing, each trial is scaled using the standard scaler. Subsequently, each trial is segmented into 9 half-overlapping samples, where each sample is a 1-second time sequence of 256 timestamps. This process yields a total of 5,967 samples. Each sample is assigned a subject ID to indicate the originating subject. The 16 channels are grouped into 5 brain regions covering the prefrontal, frontal-temporal, temporal, central, and parieto-occipital areas.

TDBrain. A large permission-accessible EEG time series dataset recording brain activities of 1,274 subjects with 33 channels. Each subject has two trials: one under an eye-open condition and one under an eye-closed condition. It includes 60 diagnostic labels, with each subject potentially having multiple labels indicating co-occurring conditions. We utilize a subset containing 25 subjects with Parkinson's disease and 25 healthy controls, all under the eye-closed task condition. Each eye-closed trial is segmented into non-overlapping 1-second samples of 256 timestamps, and any samples shorter than 1 second are discarded. This process yields a total of 6,240 samples. The 33 channels are grouped into 11 regions covering the frontopolar, frontal, frontocentral, temporal, central, centroparietal, parietal, and occipital cortical areas, along with auxiliary channels for ocular, cardiac, and muscular artifact monitoring.

M-fNIRS. M-fNIRS is a proprietary functional near-infrared spectroscopy dataset collected for mental disorder diagnosis, which will be made publicly available upon acceptance. The dataset comprises 1,090 subjects recorded with 53 channels, covering five diagnostic categories: healthy controls, schizophrenia, bipolar I disorder, bipolar II disorder, and major depressive disorder. The detailed distribution of diagnostic categories is provided in Appendix C. fNIRS captures cortical hemodynamic activity by measuring changes in oxygenated hemoglobin concentration, offering superior spatial localization over EEG, particularly in the prefrontal cortex. Each recording consists of a time series with 3,400 timestamps measured across 53 channels. The input data can therefore be represented as a matrix with 3,400 temporal samples and 53 measurement channels. The 53 channels are grouped in to 10 brain regions based on anatomical correspondence, with 5 regions per hemisphere covering the premotor and supplementary motor area, Broca area, frontal eye field, dorsolateral prefrontal cortex, and frontopolar area.

Table 3: Summary of brain regions and corresponding channels across datasets.

Dataset	Brain Region	Channels / Electrodes
ADFTD	Frontopolar-Frontal	Fp1, Fp2, F3, F4, Fz
	Lateral Frontal-Temporal	F7, F8, T3, T4
	Central	C3, Cz, C4
	Posterior Temporal	T5, T6
	Parieto-Occipital	P3, Pz, P4, O1, O2
APAVA	Prefrontal	Fp1, Fp2, F3, F4
	Frontal-Temporal	F7, F8, T3, T4
	Temporal	T5, T6
	Central	C3, C4
	Parieto-Occipital	P3, P4, O1, O2
TDBrain	Frontopolar	Fp1, Fp2
	Frontal	F7, F3, Fz, F4, F8
	Frontocentral	FC3, FCz, FC4
	Temporal	T7, T8
	Central	C3, Cz, C4
	Centroparietal	CP3, CPz, CP4
	Parietal	P7, P3, Pz, P4, P8
	Occipital	O1, Oz, O2
	EOG	VPVA, VNVB, HPHL, HNHR, OrbOcc
	ECG/Cervical	Erbs
	EMG/Masseter	Mass
M-fNIRS	L-PreM&SMA	1, 4, 10
	L-BROCA	2, 3, 5, 7, 8, 13
	L-FEF	12, 24
	L-DLPFC	6, 11, 14, 17, 18, 20
	L-FPA	9, 15, 16, 19, 21, 22, 23, 27
	R-PreM&SMA	30, 33, 35, 36, 37, 41, 43, 48
	R-BROCA	31, 32, 34, 39, 42, 45
	R-FEF	26, 38
	R-DLPFC	44, 46, 49, 50, 51, 53
	R-FPA	40, 47, 52

C Diagnostic Category Distribution of M-fNIRS

Table 4: Distribution of diagnostic categories in the M-fNIRS dataset.

Label	Diagnostic Category	Number of Subjects
0	Healthy Controls	739
1	Schizophrenia	36
2	Bipolar I Disorder	99
3	Bipolar II Disorder	100
4	Major Depressive Disorder	116
Total		**1,090**

D Calculation of Evaluation Metrics

The Receiver Operating Characteristic curve plots the true positive rate $\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$ against the false positive rate $\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP+TN}}$ across decision thresholds, with AUROC defined as

$$\mathrm{AUROC} = \int_0^1 \mathrm{TPR}\big(\mathrm{FPR}^{-1}(t)\big)\,\mathrm{d}t. \tag{20}$$

AUROC reflects the overall discriminative capability and is robust to moderate class imbalance. The Precision–Recall curve plots precision $\mathrm{Prec} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$ against recall $\mathrm{Rec} = \mathrm{TPR}$, and AUPRC is:

$$\mathrm{AUPRC} = \int_0^1 \mathrm{Prec}\big(\mathrm{Rec}^{-1}(t)\big)\,\mathrm{d}t. \tag{21}$$

AUPRC is particularly informative under class imbalance, as it emphasizes performance on the positive class.