dFCExpert: Learning Dynamic Functional Connectivity Patterns with Modularity and State Experts

Tingting Chen, Hongming Li, Hao Zheng, and Yong Fan, *Senior Member, IEEE*

***Abstract*— Characterizing brain dynamic functional connectivity (dFC) patterns from functional Magnetic Resonance Imaging (fMRI) data is of paramount importance in imaging neuroscience and medicine. Recently, many graph neural network (GNN) models, combined with transformers or recurrent neural networks (RNNs), have shown great potential for modeling the dFC patterns. However, these methods face challenges in effectively characterizing the modularity organization of brain networks and capturing varying dFC state patterns. To address these limitations, we propose dFCExpert, a novel method designed to learn robust representations of dFC patterns in fMRI data with modularity experts and state experts. Specifically, the modularity experts optimize multiple experts to characterize the brain modularity organization during graph feature learning process by combining GNN and mixture of experts (MoE), with each expert focusing on brain nodes within the same functional network module. The state experts aggregate temporal dFC features into a set of distinctive connectivity states using a soft prototype clustering method, providing insight into how these states support diverse brain functions and how they vary across brain conditions. Experiments on three large-scale fMRI datasets have demonstrated the superiority of our method over existing alternatives. The learned dFC representations not only enhance interpretability but also hold promise for advancing our understanding of brain function across a range of conditions, including development, sex difference, and Autism Spectrum Disorder. Our code is publicly available at MLDataAnalytics/dFCExpert [1].**

***Index Terms*— fMRI, dynamic functional connectivity, brain modularity organization, dynamic FC states, mixture of experts.**

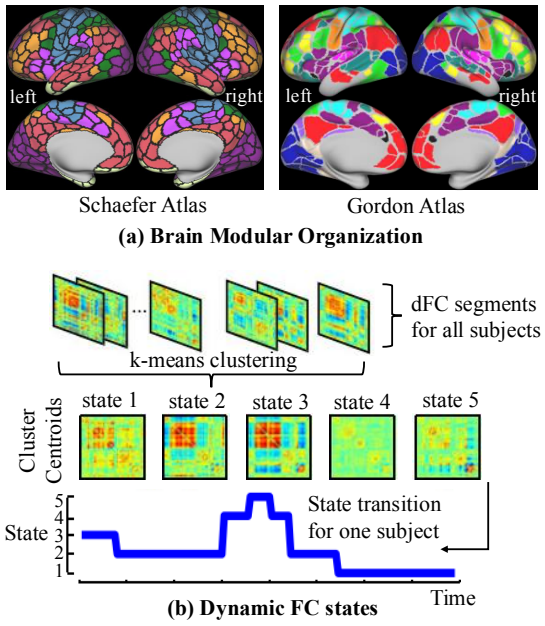

Fig. 1. (a) The Schaefer [7] and Gordon atlases [8] with different functional networks/modules marked by different colors. Each black/gray-line-bordered region represents a brain node, while regions of the same color belong to a single functional module, collectively associated with specific brain functions. (b) dFC states represent a set of recurring patterns of dFC measures, capturing the evolving FC dynamics of the brain.

I. Introduction

The human brain is a dynamic network system that generates complex spatiotemporal dynamics of brain activity. Analyzing such dynamics holds promise to provide insights into the brain's functional organization and its relationship with human cognition, behaviors, and brain disorders [1]–[5]. Among many brain imaging techniques, functional Magnetic Resonance Imaging (fMRI) is a particularly powerful tool that models the spatiotemporal patterns of brain activity by measuring fluctuations in blood-oxygen level-dependent (BOLD) signals [6]. Since brain activity often exhibits strong spatial correlations, BOLD signals can be parcellated into a set of pre-defined brain regions of interest (ROIs). The pairwise correlations between those ROIs define functional connectivity (FC), which has emerged as a key tool for understanding brain function.

Based on the FC, the brain can be modeled as a functional network with graph theory approaches, where brain ROIs serve as network nodes and the FC strengths between them act as edges. Leveraging this graph-structured nature of the brain, many fMRI data analytic methods have employed graph neural networks (GNNs) to learn brain network representations,

This work was supported in part by NIH grants: R01EB022573 and R01AG06665.

Tingting Chen, Hongming Li, and Yong Fan are with the Center for Biomedical Image Computing and Analytics, and Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA (e-mail: tingting.chen/hongming.li/yong.fan@pennmedicine.upenn.edu).

Hao Zheng is with the School of Computing and Informatics, University of Louisiana at Lafayette, Lafayette, LA 70503, USA (e-mail: hao.zheng@louisiana.edu).

[1]This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

enabling tasks such as decoding human traits or diagnosing diseases [9]–[11]. Generally, these methods fall into two categories: *static FC* and *dynamic FC methods*. The *static-FC methods* characterize the FC between nodes based on the entire fMRI scan, i.e., the full time series [12], [13]. However, these methods are not equipped to characterize the dynamic properties of FCs (fluctuate over time), which are crucial for capturing the brain's evolving states. Differently, the *dynamic-FC methods* split the whole fMRI time series into temporal segments and quantify the FC based on each segment so that time-varying FC measures can be derived. A typical pipeline for dynamic FC methods involves extracting brain network representations using GNNs for each temporal FC segment, followed by the applications of RNNs or transformers for temporal dynamics learning [4], [5], [14], [15]. Although promising, these methods still face challenges, primarily due to the unique characteristics of brain functional networks and dFC measures.

Firstly, **most GNN brain functional network analysis methods overlook the intrinsic brain modularity organization**, leading to suboptimal graph representation learning [9], [14], [16]–[19]. Both theoretical and empirical studies highlight that the brain functions as a modular system, comprising specialized cognitive and topological modules (also referred as functional networks in this paper). Each module or functional network comprises tightly connected brain network nodes that collaboratively perform specific functions [20], [21], as illustrated in Fig. 1(a). However, existing GNN-based methods typically treat all brain nodes uniformly, employing the same aggregation mechanism regardless of variations in node features. This limitation prevents these methods from effectively capturing modularity-specific features. *Secondly*, **most known approaches fail to capture distinctive dFC state patterns.** Evidence suggests that dFC measures often correspond to distinctive dynamic states (Fig. 1(b)), which can be identified by clustering dFC measures from temporal segments of fMRI scans [22]. Studies on brain disorders have revealed that disease-specific alterations are confined to certain dynamic states [22], underscoring that capturing dFC states can potentially improve the detection of functional brain changes associated with brain disorders.

To address these challenges, we propose a novel GNN-based dFC learning framework, called dFCExpert, aiming to enhance the representation learning of dynamic functional connectivity of fMRI data. The framework consists of two key components: modularity experts and state experts, specifically designed to characterize the brain modularity and the dFC state patterns. The modularity experts are built upon GNN and Mixture of Experts (MoE) for learning brain graph features of each FC segment, which characterize the brain modularity organization in the graph learning process by optimizing multiple experts at each graph layer. Each of the experts focuses on specific subsets of nodes that exhibit similar behaviors and interactions, enabling fine-grained modeling of the brain's modular structure. On top of the graph representations learned by the modularity experts, the state experts aggregate the temporal features of dFCs into a compact set of distinctive states based on a soft prototype clustering method, where each state has similar FC patterns and reflects different activities of the dFCs related to human behaviors or brain conditions. Finally, the clustering-derived state features are used for task prediction. With the modularity and state experts, our dFCExpert introduces two novel strategies for learning-based fMRI analysis: modeling the brain functional network modularity and dynamic FC states. This pioneering approach presents a first-of-its-kind method for learning informative and explainable features of dFC patterns from fMRI data, offering significant potential for advancing imaging neuroscience research and clinical applications.

We have evaluated the performance of dFCExpert on three large-scale fMRI datasets, including the Human Connectome Project (HCP) [23], the Adolescent Brain Cognitive Development (ABCD) [24], and the Autism Brain Imaging Data Exchange (ABIDE) [25]. Across pattern recognition tasks such as classification and regression, dFCExpert consistently outperformed known methods, achieving state-of-the-art performance. To validate the framework's design, we conducted comprehensive ablation studies to analyze the contribution of individual components of dFCExpert. Visualization results revealed that the modularity experts effectively target distinct brain modules and the state experts identify unique dynamic brain states, indicating that the method was capable of learning interpretable representations of dFC measures for prediction tasks. Moreover, we demonstrated that the brain modules and dynamic states identified by dFCExpert are biologically meaningful, bridging the gap between deep learning-based functional network modeling and traditional functional network analytic methods, highlighting the potential of dFCExpert to deliver both robust predictive performance and valuable insights into the brain's functional organization.

II. Related Works

A. Dynamic Functional Connectivity Learning

To capture the time-varying FC patterns of the brain, a spatio-temporal graph convolutional network (ST-GCN) is developed to incorporate both spatial and temporal convolutions to model the non-stationary nature of dFC measures [18]. Many methods have adopted a GNN-RNN pipeline to learn graph-level features from each FC segment with the GNNs, followed by RNNs or transformers to capture temporal FC patterns [14]–[17]. Particularly, STAGIN utilizes GNNs with spatial attention and transformers with temporal attention to model the dynamics of brain networks [14]. DynDepNet introduces a dynamic graph structure learning method to capture the time-varying structures of fMRI data [17]. NeuroGraph systematically evaluates the effectiveness of different GNN designs for modeling dynamic functional networks [16]. While these methods have shown promise in modeling the dynamics of brain functional connectivity, they often overlook the brain's intrinsic modular organization and fail to capture distinctive dFC states [14], [16]–[18]. This limitation hinders their ability to provide deeper insights into the brain's functional organization and its evolving connectivity states.

B. Brain Modularity Organization

The human brain functions as a modular system, composed of distinct functional and topological modules. Brain regions within the same modules are tightly connected and tend to perform similar functions [15]. For instance, salience network (SN) and default mode network (DMN) are two crucial neurocognitive modules in the brain, where the SN mainly detects internal or external stimuli and coordinates the brain's response to those stimuli, and the DMN is responsible for self-referential thinking, mind-wandering, and introspection [20], [21]. A few existing methods have incorporated the brain modular organization into graph representation learning. BRAINNETTF introduces a novel graph readout function that leverages modular-level similarities between brain nodes to pool graph-level embeddings from clusters of functionally similar nodes [11]. MSGNN [15] develops a modularity-constrained GNN that enforces node embeddings to align with three functional network modules (i.e., central executive network, SN, DMN), by applying a modularity constraint loss after the graph layers to encourage similarity between node-level embeddings within the same module. However, since these methods incorporate the brain modularity after the graph model, all brain network nodes are still processed uniformly within each graph layer. This may be suboptimal for effectively learning representations of nodes belonging to distinct brain modules. To address this, we propose modularity experts, a combination of GNN and MoE, to mimic the brain's modular organization, which employs multiple experts at each graph layer to guide brain network nodes with similar functions towards the same experts.

C. Dynamic FC States

The dynamic brain states are a compact set of unique patterns derived from dFC measures, which represent connectivity patterns that repetitively occur during the fMRI acquisition [26]. Existing studies typically identify the dynamic brain states by clustering the dFC measures into different states with methods, such as k-means, and then conduct statistic analyses to investigate the relationships between these states and biological or behavior measures of interest [22]. However, this two-step strategy has significant limitation in that the features used for clustering are not explicitly optimized for the downstream statistical analyses, leading to a potential mismatch between the identified states and their informativeness for specific tasks. In contrast, our state experts overcome this limitation by adopting an end-to-end learning framework, simultaneously learning distinctive dFC states and optimizing representations of dFC graphs for specific tasks. By integrating state discovery and task optimization, our method ensures that the resulting states are both highly informative and tailored to the requirements of the analysis, enhancing their interpretability and utility in studying dFC patterns.

D. Expert Models

The concept of Mixture of Experts (MoE) is a well-established machine learning technique that employs a divide-and-conquer strategy, where a system is composed of multiple specialized experts, each responsible for solving a specific subtask or learning a specific substructure. Over the years, MoE has demonstrated remarkable success across various domains, including multi-modal learning [27], computer vision [28], and machine translation [29]. In the context of graph classification, for example, TopExpert applies MoE to extracted graph features, leveraging topology-specific experts for molecule property prediction [30]. More recently, a graph MoE (GMoE) model incorporates multiple experts within each graph layer to enhance the capacity of GNNs [31]. Building upon MoE, we propose modularity experts, a novel approach designed to capture the brain's modular organization with mutilple experts. Each expert specializes in learning representations for a specific brain module, enabling our model to function in a brain-inspired manner and improving the representation learning of dFCs.

III. Method

We begin by introducing the preliminaries of the proposed dFCExpert framework, including the concepts of GNNs and the construction of dFC graphs from fMRI data (Section III-A), followed by a detailed description of the framework, including an overview, modularity experts, state experts, and the overall loss function (Section III-B).

A. Preliminaries

1) Graph Neural Networks: To glean useful information from graph-structured data, GNNs iteratively compute node features by aggregating information of neighbor nodes and updating the node features with non-linear functions in a layer-wise manner. The propagation mechanism for a node i at the (l)-th GNN layer is formulated as:

$$h_i^{(l)} = F_u^{(l)}\left(h_i^{(l-1)}, F_a^{(l)}\left(\{(h_i^{(l-1)}, h_j^{(l-1)}, e_{ij}) : j \in N_i\}\right)\right), \quad (1)$$

where $h_i^{(l-1)}$ denotes the features of node i at the $(l-1)$ layer, N_i is the set of neighbors of node i, e_{ij} denotes the edge between nodes i and j, and F_a and F_u are differentiable functions for aggregating information and updating node features, respectively. Different choices of these two functions lead to various GNN architectures. For instance, Graph Isomorphism Network (GIN) [32], a variant of GNN, uses summation as F_a and a multi-layer perceptron (MLP) as F_u:

$$\begin{aligned} h_i^{(l)} &= M^{(l)}(h_i^{(l-1)}, e_{ij}, W^{(l)}) \\ &= \mathrm{MLP}^{(l)}\left((1+\epsilon^{(l)}) \cdot h_i^{(l-1)} + \sum_{j \in N_i} h_j^{(l-1)}\right). \end{aligned} \quad (2)$$

We simplify the two functions described above into a single operation $M^{(l)}$, where $W^{(l)}$ represents the trainable weight of the multi-layer perceptron (MLP), and ϵ is a learnable parameter initialized to zero. Given its strong ability for graph representation learning, we adopt GIN as our dFC graph feature extractor, following the approach in [14].

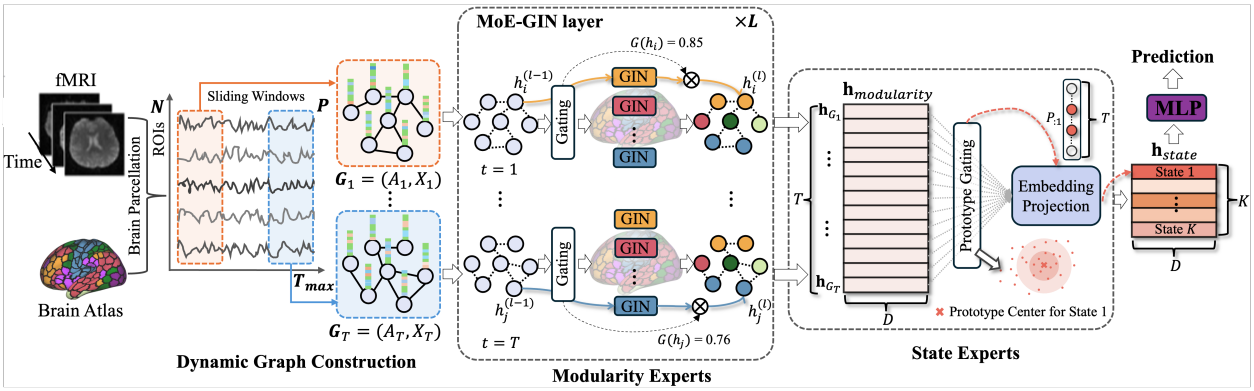

Fig. 2. The dFCExpert framework consists of modularity experts and state experts. The modularity experts include L layers of MoE-GIN, which route node features h_i to a specific GIN expert with a gating score $G(h_i)$ for learning graph-level features $\mathbf{h}_{G_t}$ for each temporal segment t, yielding $\mathbf{h}_{modularity} = (\mathbf{h}_{G_1}, \ldots, \mathbf{h}_{G_T})$ across all T segments. The state experts aggregate the learned temporal graph features into dynamic states using prototype gating and embedding projection. Finally, the state features $\mathbf{h}_{state}$ are processed through an MLP layer for task prediction.

2) Dynamic Graph Definition: To construct dFC graphs from fMRI scans, we first use a brain atlas to transform the 4D fMRI data into a time-course matrix $\mathbf{P} \in \mathbb{R}^{N \times T_{\max}}$. This matrix is created by averaging the fMRI signals within each brain region defined by the atlas, yielding fMRI signals of N brain network nodes at each time point. Next, using a sliding-window approach, the time-course matrix is divided into $T = \lfloor T_{\max} - \Gamma / S \rfloor$ temporal segments, where Γ is the window length and S is the stride size. For each segment t, an FC matrix $X_t \in \mathbb{R}^{N \times N} (t = 1, \ldots, T)$ is computed by calculating Pearson correlation coefficients between the time series of all brain region pairs, yielding T FC matrices. Following [16], the correlation matrices can be informative node features, where the features for the i-th node in segment t correspond to the elements of the i-th row in X_t. Additionally, a binary adjacent matrix $A_t \in \{0, 1\}^{N \times N}$ is derived from X_t by retaining the top 30-percentile correlation values as connected and setting the remaining values as unconnected, as described in [13]. Thus, the input dFC graphs for each subject are represented as $\mathbf{G}_t = \{A_t, X_t\} (t = 1, \ldots, T)$ (Fig. 2).

B. dFCExpert

1) Overview: As illustrated in Fig. 2, dFCExpert consists of modularity experts and state experts. Taking the dFC graphs as input, the modularity experts leverage a combination of GIN and MoE to learn brain graph features for each FC segment, aiming to capture the brain modularity mechanism effectively. Building upon the outputs of the modularity experts, the state experts adaptively group the temporal graph features into distinctive states using a soft prototype clustering method. This approach allows the model to learn expressive state features by assigning soft clusters to the temporal graph features. Finally, the learned state features are passed through an MLP layer to predict a specific task, such as predicting sex in a classification setting or predicting an intelligence measure in a regression setting. Formally, the objective of dFCExpert is to train a neural network $f : (\mathbf{G}_1, \ldots, \mathbf{G}_T) \rightarrow \mathbf{h}_{state}$, where $\mathbf{G}_t = \{A_t, X_t\} (t = 1, \ldots, T)$ represents the sequence of constructed dFC graphs with T segments, and $\mathbf{h}_{state} \in \mathbb{R}^{K \times D}$ is the output features from the state experts. We formulate $f = s \circ m$ as a composition of modularity experts m and state experts s, where m learns dFC graph representations $\mathbf{h}_{modularity} = (\mathbf{h}_{G_1}, \ldots, \mathbf{h}_{G_T})$ for each of T temporal segments, and s aggregates these learned temporal graph features into the final state features $\mathbf{h}_{state}$:

$$\begin{aligned} m &: (\mathbf{G}_1, \ldots, \mathbf{G}_T) \rightarrow (\mathbf{h}_{G_1}, \ldots, \mathbf{h}_{G_T}), \\ s &: (\mathbf{h}_{G_1}, \ldots, \mathbf{h}_{G_T}) \rightarrow \mathbf{h}_{state}. \end{aligned} \tag{3}$$

2) Modularity Experts: The modularity experts are designed to emulate the human brain's modular organization, enabling effective learning of brain graph representations. The modularity experts are implemented using a combination of GIN and MoE, referred to as MoE-GIN, consisting of L layers of MoE-GIN. Specifically, each MoE-GIN layer includes one gating function and multiple GIN experts, and all T segments share the same MoE-GIN layer, as illustrated in Fig. 2. The gating function determines which experts are most suitable for processing a given node, while the multiple experts are independent GIN propagation functions (Eq. (2)), each with its own trainable parameters. This design of MoE-GIN enables nodes that are tightly connected or behaviorally similar to be routed to the same expert. Consequently, each GIN expert specializes in learning representations for a specific brain module to effectively capture distinct brain activities. By organizing the brain nodes into specialized groups, the modularity experts mimic the way the brain operates, where tightly connected regions interact to perform specific functions or activities. This design allows the model to reflect the brain's modularity, improving its ability to learn meaningful and interpretable representations of brain networks. Formally, the operation of a single MoE-GIN layer is described as:

$$h_i^{(l)} = \sum_{c=1}^{C} G_c^{(l)}(h_i^{(l-1)}) M_c^{(l)}(h_i^{(l-1)}, e_{ij}, W^{(l)}) \tag{4}$$

where C is the number of experts, and M denotes the node propagation function of GIN as defined in Eq. (2). G is a gating function to generate assignment scores based on the input of node features $h_i^{(l-1)}$, and $G_c(h_i^{(l-1)})$ denotes the score for the c-th expert. We implement a simple gating function by multiplying the input with a trainable weight matrix W_g, followed by the application of the $\mathrm{softmax}$ function [33]:

$G(h_i) = \text{softmax}(h_i W_g)$. Given that each node typically belongs to a single brain module, we adopt a simplified top-1 gating scheme as in [34], where each node is routed to only one expert, corresponding to the expert with the highest gating score. This simplification reduces computational overhead of routing while maintaining model quality and interpretability. Additionally, each of the T segments will go through the L layers of MoE-GIN, we omit segment notation t in Eq. (4) for brevity, whenever it is not of contextual importance.

The graph-level representation $h_{G_t}^{(l)}$ for segment t at layer l is computed by averaging the updated node features $\{h_i^{(l)}, i \in N\}$, and the final graph representation $\mathbf{h}_{G_t}$ is then obtained by concatenating the graph representations from all L layers [35] followed by an MLP layer for dimension reduction: $\mathbf{h}_{G_t} = \text{MLP}(\text{concatenate}(h_{G_t}^{(l)} | l \in \{1, \cdots, L\})) \in \mathbb{R}^D$. Finally, we obtain $\mathbf{h}_{modularity} = (\mathbf{h}_{G_1}, \ldots, \mathbf{h}_{G_T}) \in \mathbb{R}^{T \times D}$.

Auxiliary Loss Functions for Training MoE-GIN. If the model is trained solely using the task-specific prediction loss, the gating network may converge to trivial solutions where only a few experts are consistently selected [36]. This imbalance in expert selection becomes self-reinforcing, as the favored experts dominate the learning process, further increasing their frequency of selection. Another trivial situation is that the gating network may produce similar assignment scores across all experts for a given node, resulting in a lack of specialization in experts. In our scenario, we expect the gating network to be capable of effectively distinguishing nodes of different brain functional modules and enable sparse gating so that the selected expert receives a significantly higher score than the others. To achieve this, we introduce two auxiliary loss functions to promote balanced loading and enforce sparse gating, respectively. Given a batch B, each containing T graphs, the loading balance loss is computed as:

$$\begin{aligned} L_{balance} &= \frac{1}{BTC} \sum^{B} \sum^{T} \sum_{c=1}^{C} p_c \log(p_c), \\ p_c &= \frac{1}{N} \sum_{i=1}^{N} G_c(h_i), \end{aligned} \tag{5}$$

where N is the number of nodes, and p_c represents the fraction of nodes routed to expert c. By minimizing the negative entropy of the p_c distribution, the $L_{balance}$ encourages a uniform distribution of nodes across all experts. To promote sparse gating in $G(h_i)$, we introduce a sparse gating loss:

$$L_{sparse} = \frac{1}{BTN} \sum^{B} \sum^{T} \sum^{N} \frac{1}{C} \sum_{c=1, g \in G(h_i)}^{C} -g_c log(g_c), \tag{6}$$

where g_c represents the gating score for assigning a given node to expert c. Minimizing the entropy of the gating scores across all experts encourages sparsity, ensuring each node is strongly associated with a specific expert. We implement these two loss functions similar to [37], as they are both fundamentally deep clustering-based and exert opposing influences during training.

3) State Experts: The state experts follow the traditional dFC analysis in imaging neuroscience to adaptively group the temporal graph features $\mathbf{h}_{modularity}$ into a small number of states, each reflecting distinct dynamic states of dFC measures associated with different brain conditions. From another perspective, representing dFC measures with a smaller set of states simplifies the model, reduces data complexity, and makes it easier to interpret the FC dynamics. Specifically, we design the state experts using a soft prototype clustering method, where temporal FC graph features are softly assigned to clusters in an unsupervised manner. As illustrated in Fig. 2, distinctive dFC states are characterized through prototype gating and embedding projection. The prototype gating introduces K trainable prototype centroids $U = [u_1, \ldots, u_K]$, each with D dimensions ($U \in \mathbb{R}^{K \times D}$), to adaptively learn the feature centers of each state. Given the learned graph features $\mathbf{h}_{modularity} = (\mathbf{h}_{G_1}, \ldots, \mathbf{h}_{G_T})$, the prototype gating calculates the assignment score P_{tk} for assigning segment t to state k using a Softmax projection:

$$P_{tk} = \frac{e^{\langle \mathbf{h}_{G_t}, u_k \rangle}}{\sum_{k'}^{K} e^{\langle \mathbf{h}_{G_t}, u_{k'} \rangle}}, \tag{7}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. The embedding projection aggregates the temporal graph features $\mathbf{h}_{modularity} \in \mathbb{R}^{T \times D}$ into state features $\mathbf{h}_{state}$ under the guidance of the obtained soft assignment score $P \in \mathbb{R}^{T \times K}$. Specifically, to compute features for a single state (e.g., State 1 as in Fig. 2), the embedding projection integrates features of the temporal segments assigned to that state using soft gating, performed as $P_{:1}^{\top} \mathbf{h}_{modularity}$. This process ensures that the state features are derived by aggregating the temporal FC features with high probabilities assigned to a specific state, thus enabling the model to effectively capture the dynamic nature of FC patterns and associate them with interpretable states.

Since the clustering lacks explicit guidance labels, we optimize the temporal FC features and cluster centroids in a self-supervised manner by leveraging orthonormal initialization of U [11] and optimizing the cluster distribution for cohesive clustering. Specifically, we define a target distribution Q_{tk} for each $\mathbf{h}_{G_t}$ based on the current cluster assignment distribution P_{tk}, which therefore enhances high-confident assignments [38] through a squaring operation, defined as:

$$Q_{tk} = \frac{P_{tk}^2 / \sum_t P_{tk}}{\sum_{k'} (P_{tk'}^2 / \sum_t P_{tk'})}. \tag{8}$$

This target distribution strengthens the cluster cohesion, refining temporal FC representations to become more discriminative with respect to their state patterns. Finally, we minimize a KL divergence loss to align P and Q, boosting the cohesion of the clusters:

$$L_{state} = \mathbf{KL}(Q || P) = \frac{1}{T} \sum_{t=1}^{T} \sum_{k=1}^{K} Q_{tk} \log \frac{Q_{tk}}{P_{tk}}. \tag{9}$$

4) Overall Loss Function: The final loss combines a task-specific loss (L_{task}), two auxiliary losses for the modularity experts ($L_{balance}$ and L_{sparse}), and a loss for state experts (L_{state}), yielding the overall optimization objective. Two scaling factors, α and β, are used to balance the contributions of these components:

$$L_{overall} = L_{task} + \alpha(L_{balance} + L_{sparse}) + \beta L_{state}. \tag{10}$$

After experimenting with various scaling ratios, we set $\alpha = 1$ and $\beta = 10$ for optimal performance. Note that the auxiliary losses for modularity experts are computed as the summation across all MoE-GIN layers.

TABLE I
PERFORMANCE COMPARISON WITH ALTERNATIVES AND BASELINES. ("M"= MODULARITY, "S" = STATE)

Method	HCP				ABCD				ABIDE		FC Type
	Sex		Intelligence		Sex		Cognition		ASD		
	ACC(%)	AUC(%)	MSE(↓)	CORR(↑)	ACC(%)	AUC(%)	MSE(↓)	CORR(↑)	ACC(%)	AUC(%)	
BrainNetCNN [10]	84.56±1.87	91.97±1.33	1.130±0.039	0.200±0.071	85.38±0.07	92.67±0.33	0.474±0.031	0.442±0.011	64.29±1.73	72.66±1.93	Static
BRAINNETTF [11]	85.42±2.36	92.51±2.01	1.015±0.052	0.205±0.074	84.55±0.64	91.38±0.21	0.462±0.022	0.465±0.024	65.57±2.06	70.13±1.61	Static
NeuroGraph [16]	84.38±2.60	92.03±2.39	1.321±0.043	0.198±0.065	85.71±0.60	93.75±0.61	0.512±0.013	0.423±0.015	64.98±3.19	70.94±2.53	Static
STAGIN [14]	87.74±1.75	92.09±0.93	1.002±0.042	0.215±0.078	86.93±0.45	91.57±0.57	0.452±0.012	0.477±0.008	60.86±4.30	67.24±4.28	Dynamic
MSGNN [15]	86.25±2.03	92.37±1.42	1.073±0.057	0.206±0.070	87.13±0.63	92.67±0.42	0.453±0.024	0.483±0.017	65.36±3.16	70.46±3.42	Dynamic
NeuroGraph [16]	84.66±3.30	93.14±2.23	1.017±0.032	0.207±0.069	87.37±1.08	94.64±0.69	0.441±0.031	0.482±0.005	65.18±3.04	70.28±3.36	Dynamic
Baseline	88.11±4.26	96.68±1.71	0.978±0.079	0.238±0.086	87.25±0.70	94.86±0.70	0.442±0.021	0.484±0.013	65.85±3.12	72.46±2.96	Dynamic
Baseline + M	89.61±4.22	96.91±1.68	0.946±0.041	0.272±0.059	88.09±0.76	95.16±0.54	0.425±0.026	0.499±0.018	67.29±3.05	**74.87±3.30**	Dynamic
Baseline + S	89.98±2.83	97.23±1.32	0.967±0.074	0.248±0.083	88.56±1.25	95.81±0.53	0.419±0.014	0.496±0.013	66.17±3.05	74.31±2.82	Dynamic
dFCExpert	**91.03±1.71**	**97.40±1.30**	**0.932±0.036**	**0.296±0.049**	**89.28±0.83**	**95.99±0.52**	**0.412±0.012**	**0.513±0.009**	**68.93±2.89**	74.62±3.02	Dynamic

IV. EXPERIMENTS

A. Experimental Settings

1) Datasets: We conducted experiments on three fMRI datasets. (a) *Human Connectome Project (HCP)*: We used data from the HCP S1200 release [23], which includes preprocessed and ICA de-noised [39] resting-state fMRI scans of 1071 individuals with 578 females and 493 males. Following [14], we selected data from the first run and excluded scans with fewer than 1200 time points. All the resting-state fMRIs used in this study contain exactly 1200 time points. To build functional networks, we parcellated the brain into 400 regions using the Schaefer atlas [7], which is organized into seven intrinsic connectivity networks. (b) *Adolescent Brain Cognitive Development Study (ABCD)*: The study is a multi-site investigation of brain development and its relationship with behavioral outcomes in children aged 9-10 years [24]. In our study, we used preprocessed baseline resting-state fMRI data from the ABCD BIDS Community Collection (ABCC) [40]. As in [41], participants with incomplete data, excessive head motion, or fewer than 600 remaining time points after motion censoring were excluded. After quality control, our final dataset included resting-state fMRI scans of 6,195 children (3,080 females and 3,115 males), with the number of time points ranging from 626 to 3,516. The Gordon atlas [8] with 352 regions ($N = 352$) was used for brain parcellation, as defined by the ABCC. (c) *Autism Brain Imaging Data Exchange (ABIDE)*: The ABIDE initiative aggregates functional brain imaging data from multiple sites to support research on Autism Spectrum Disorder (ASD) [25]. We used data from [42], comprising fMRI scans from 1025 subjects (537 typical controls and 488 individuals with ASD), with time points ranging from 120 to 300. Similar to HCP, the Scheafer atlas with 400 regions was used for brain parcellation.

All datasets are publicly available, with all identification information anonymized. The image acquisition parameters for the HCP, ABCD and ABIDE datasets can be found in [39], [24], and [43], respectively. We created 5 stratified splits for all datasets with a train-validation-test ratio of 7:1:2. Average results and standard deviations across the 5 splits were reported.

2) Targets and Metrics: We selected three evaluation tasks: sex and ASD classification, and cognitive intelligence prediction ("fluid intelligence" for HCP and "general cognition" for ABCD). The sex/ASD classification task is a binary classification problem, and the cognitive intelligence prediction is a regression problem. These tasks were chosen to explore brain-biology and brain-cognition associations. For the sex/ASD classification task, we used accuracy (ACC) and AUC (Area Under ROC Curve) as evaluation metrics. For the regression task, the regression targets were z-normalized and the performance was evaluated using Mean Squared Error (MSE) and correlation coefficient (CORR) between the measured and predicted values.

3) Implementation Details: Our method was implemented using PyTorch [44] and trained on an NVIDIA A100 GPU with 80 GB memory. We set the number of MoE-GIN layers to $L = 3$ and the embedding dimension to $D = 256$. For dFC graph construction, we used a window length of $\Gamma = 50$ and a window stride of $S = 3$, following common protocols for sliding-window-based dFC analyses [26], [45]. Consistent with [14], we randomly selected 600 time points at each training step for the dFC graph computation. This procedure reduced computational and memory overhead while augmenting the training data. For testing, we utilized the full time-course matrix $\mathbf{P} \in \mathbb{R}^{N \times T_{\max}}$ to construct dFC graphs and evaluate the model. The sex/ASD classification and cognitive intelligence regression tasks were trained using cross-entropy and MSE losses, respectively. To ensure a fair comparison, we trained both dFCExpert and known methods under comparison in an end-to-end supervised fashion using a similar configuration:

- *Optimizer*: Adam.
- *Learning rate*: A learning rate of $5e^{-4}$ for training the classification task, and a learning rate of $1e^{-3}$ for training the regression task.
- *Mini-batch size*: mini-batch size of 8 for the HCP/ABIDE dataset and mini-batch size of 16 for the ABCD dataset.
- *Epochs*: 50 epochs for the HCP/ABIDE dataset, and 30 epochs for the ABCD dataset.
- *Number of experts*: $C = 7$ for HCP/ABIDE dataset and $C = 5$ for ABCD dataset.
- *Number of states*: $K = 7$ for HCP/ABIDE dataset and $K = 5$ for ABCD dataset.

Based on our experiments, C can be set to 5 or 7, and K can range from 5 to 7. These values for C and K fall within a reasonable and effective range for our model's performance.

B. Sex Classification and Cognitive Intelligence Regression Results

1) Comparison with Existing Methods: We evaluated the performance of dFCExpert against alternative *static-* and

TABLE II

IMPACT OF THE NUMBER OF EXPERTS IN THE MODULARITY EXPERTS ON PERFORMANCE.

#Experts	HCP				ABCD				ABIDE	
	Sex		Intelligence		Sex		Cognition		ASD	
	ACC(%)	AUC(%)	MSE(↓)	CORR(↑)	ACC(%)	AUC(%)	MSE(↓)	CORR(↑)	ACC(%)	AUC(%)
1 (baseline)	88.11±4.25	96.68±1.71	0.978±0.079	0.238±0.086	87.25±0.70	94.86±0.70	0.442±0.021	0.484±0.013	65.85±3.12	72.46±2.96
3	88.68±2.57	96.46±1.58	0.961±0.067	0.266±0.061	87.57±0.94	95.16±0.70	0.435±0.019	0.481±0.005	66.02±3.11	73.65±3.70
5	89.25±3.46	96.41±1.27	0.953±0.054	**0.279±0.042**	**88.09±0.76**	**95.16±0.54**	**0.425±0.026**	**0.499±0.018**	66.90±3.98	73.18±2.94
7	**89.61±4.22**	**96.91±1.68**	**0.946±0.041**	0.272±0.059	87.80±1.06	94.85±0.65	0.429±0.023	0.481±0.016	**67.29±3.05**	**74.87±3.30**
9	88.73±3.48	95.26±2.10	0.963±0.036	0.224±0.051	87.65±0.87	94.98±0.69	**0.425±0.020**	0.479±0.010	66.61±3.52	73.41±3.61
17	87.83±3.83	95.87±1.47	0.976±0.032	0.209±0.069	87.19±1.30	94.84±0.53	0.431±0.024	0.477±0.020	65.58±3.56	72.25±3.70

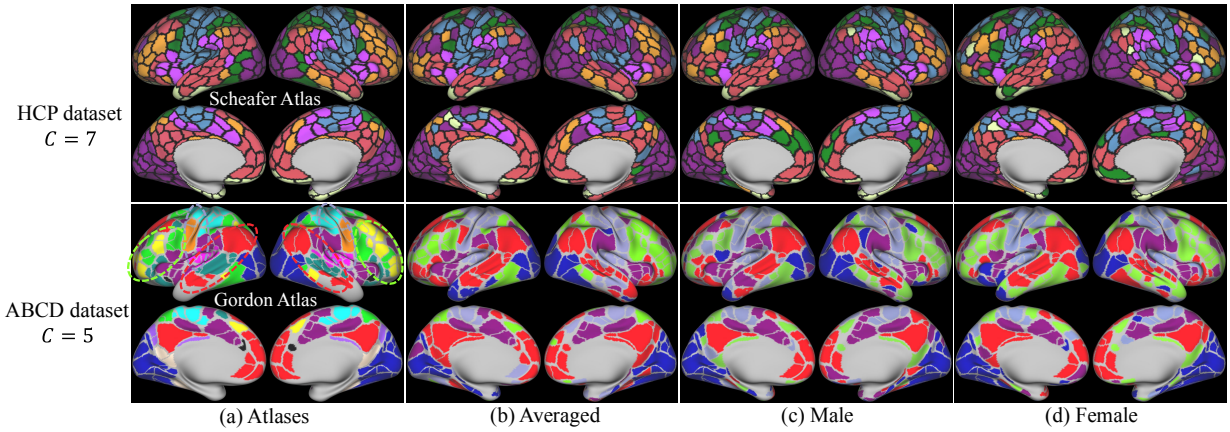

Fig. 3. Visualization of the learned brain modules (node assignment results) averaged across all subjects (b), and for two randomly selected individuals (c) and (d) from the test datasets of HCP (first row) and ABCD (second row). These are compared to the functional networks defined by the Schaefer and Gordon atlases (a). In the first row, the seven functional networks from the Schaefer atlas are color-matched with our learned modules ($C = 7$). In the second row, since the number of learned modules ($C = 5$) does not match the 12 networks in the gordon atlas, we highlight key networks using consistent colors: blue for the vision network, red for the default network, and purple for the cingulo opercular network in (b)(c)(d). Other combined modules are represented using lilac and bright green.

dynamic-FC methods on all the sex/ASD classification and cognitive intelligence regression tasks. The results of these alternative methods are summarized in the top two blocks of Table I. For the sex classification task, dFCExpert outperformed all the alternative methods under comparison on both the HCP and ABCD datasets. On the regression task, our method demonstrated significant performance improvements, particularly on the HCP dataset, with gains of nearly 7 points. For ASD classification on a disease dataset, dFCExpert also achieved superior performance. These quantitative results highlight the importance and effectiveness of incorporating brain modularity and learning state patterns to model brain dynamics in fMRI data, underscoring the advantages of our approach.

2) Ablation Studies: We conducted ablation experiments to evaluate the contributions of individual components of dFCExpert, specifically the modularity experts and state experts. For fair comparison, the baseline model was a 3-layer GIN followed by an MLP layer. The GIN was used to learn graph-level features, and the MLP made predictions for each segment t, which were then averaged to generate the final result. The baseline model was chosen for the following reasons: 1) Generating predictions for each temporal segment provides stronger supervision, potentially improving prediction compared to the alternatives that use RNNs or transformers on top of the GIN layer (e.g., STAGIN [14] and NeuroGraph [16], as shown in Table I); 2) Since the state experts aggregate the temporal graph features into K state features, the use of RNNs or transformers for final predictions is unnecessary. Based on this baseline, we introduced modularity experts ("Baseline + M") by replacing GIN with MoE-GIN, and state experts ("Baseline + S") by aggregating temporal features into several connectivity states before applying the MLP layer, respectively. The third block of Table I summarizes the ablation study results, from which several key observations can be made: 1) Compared with the baseline model, the modularity experts achieved significantly improved performance, particularly for the regression task on the HCP dataset, highlighting the effectiveness of characterizing brain modularity for learning more informative brain graph features; 2) By grouping the temporal FC measures into several states, the state experts achieved better performance across both tasks and both datasets, underscoring the benefit of using this new approach to explore the dynamics of FC measures; 3) Incorporating both the modularity and state experts, dFCExpert outperformed the baseline model, validating the overall effectiveness of our method. These results demonstrate that both components play a crucial role in improving the model's ability to capture dynamic brain connectivity and its association with cognitive and biological outcomes.

C. Modularity Experts Analysis

We evaluated the effectiveness of our modularity experts from the following perspectives: 1) the impact of the number of experts in the MoE-GIN layer; 2) the ability of modularity experts to learn specific brain modules; 3) the performance of

TABLE III
PERFORMANCE OF MODULARITY EXPERTS (MOE-GIN) APPLIED TO **STATIC-FC** MEASURES.

Method	GNN Type	HCP				ABCD			
		Sex		Intelligence		Sex		Cognition	
		ACC(%)	AUC(%)	MSE(↓)	CORR(↑)	ACC(%)	AUC(%)	MSE(↓)	CORR(↑)
NeuroGraph [16]	GIN	84.38±2.60	92.03±2.39	1.321±0.043	0.198±0.065	85.71±0.60	93.75±0.61	0.512±0.013	0.423±0.015
NeuroGraph [16]	MoE-GIN	**87.30±1.33**	**95.56±1.68**	**0.989±0.035**	**0.243±0.052**	**87.08±0.74**	**93.12±0.52**	**0.486±0.021**	**0.477±0.017**

TABLE IV
PERFORMANCE OF MODULARITY EXPERTS WITH DIFFERENT SCALING RATIOS α (α_b AND α_s REPRESENT THE SCALING RATIOS FOR $L_{balance}$ AND L_{sparse}, RESPECTIVELY).

α	α_b, α_s	Sex		Intelligence	
		ACC(%)	AUC(%)	MSE(↓)	CORR(↑)
0	$\alpha_b = 0, \alpha_s = 0$	88.56±3.45	94.67±1.68	0.962±0.022	0.256±0.048
0.1	$\alpha_b = 0.1, \alpha_s = 0.1$	89.16±4.12	95.52±2.01	0.948±0.032	0.270±0.036
1	$\alpha_b = 1, \alpha_s = 1$	**89.61±4.22**	**96.91±1.68**	**0.946±0.041**	**0.272±0.059**
-	$\alpha_b = 1, \alpha_s = 0$	87.83±4.05	94.84±2.19	0.965±0.028	0.252±0.050
-	$\alpha_b = 0, \alpha_s = 1$	88.65±3.52	95.13±1.53	0.956±0.023	0.261±0.043

modularity experts when applied to *static-FC* measures; and 4) the effect of different scaling ratios for the auxiliary losses.

1) Different Number of Experts: To evaluate the effect of modularity experts on model performance, we investigated the impact of the number of experts (C) in the MoE-GIN layer. Specifically, we tested C values of 3, 5, 7, 9, 17, where $C = 7$ and $C = 17$ are widely used settings in large-scale functional network studies [46]. The results are summarized in Table II and reveal the following insights: 1) On the HCP dataset, the model's performance improved as the number of experts increased from 3 to 7 but began to decline when the number of experts increased further from 7 to 17, suggesting that a smaller number of experts is not only optimal for performance but also beneficial for reducing computational cost and model complexity. 2) On the ABCD dataset, the modularity experts outperformed the baseline across all settings. While the number of experts had a smaller effect on the classification task, the performance on the regression task improved significantly as C increased, peaking at $C = 9$. 3) On the ABIDE dataset, we observed similar trends as in the HCP dataset, with $C = 7$ yielding the best performance for ASD identification. These results suggest that $C = 5$ or $C = 7$ are reasonable choices, consistent with the typical number of functional brain modules. We selected $C = 7$ for the HCP and ABIDE datasets, and $C = 5$ for the ABCD dataset. The variation in optimal number of experts likely reflects differences in brain parcellation strategies and the intrinsic characteristics of each dataset, as HCP (young adults), ABCD (young kids) and ABIDE (disease) represent different age groups, developmental stages, and health conditions.

2) Visualization of the Learned Brain Modules: To further evaluate the effectiveness of the modularity experts, we visualized the node assignment results produced by the modularity experts with $C = 7$ on HCP dataset and $C = 5$ on ABCD dataset, and compared them with the Schaefer Atlas (7 networks) and Gordon Atlas (12 networks) (Fig. 3 (a)), respectively. Specifically, given the gating score of last MoE-GIN layer of modularity experts, we first average the scores for all T segments and then average through all subjects (Fig. 3 (b)) or randomly choose one male and female subjects (Fig. 3 (c)(d)). Finally, we perform $argmax$ operation on the resulting scores, and we can obtain 7 and 5 learned modules from modularity experts for HCP and ABCD datasets, respectively, where each module consists of nodes assigned to a specific expert. For the Gordon atlas, we simply visualize the 5 learned modules, as shown the second row of Fig. 3. For the Schaefer atlas, we perform a color match. Since the correspondence between these expert-learned modules and the 7 atlas networks was unknown, we matched them based on the maximum overlap of their respective nodes using the Hungarian method [47], and the matched results are visualized in the first row of Fig. 3. We can observe that the visualization demonstrates a strong overlap between the atlas networks and the modules learned by the modularity experts, with a Dice coefficient of 0.71 and $p < 0.001$, as determined by a spin-based spatial permutation test [48]). This result indicates that our modularity experts effectively grouped tightly connected nodes into coherent modules, enabling each expert to specialize in capturing the structure of a specific brain functional module. Moreover, because the modularity experts were optimized for specific tasks, they could potentially learn node-expert (or node-network) assignments that go beyond the prior knowledge encoded in the Atlases. From Fig. 3 (c)(d), we observed individual differences in resulting learned modules across subjects, indicating the ability of our modularity experts to learn unique FC patterns of each individual.

3) Modularity Experts in Static-FC Method: To investigate the effectiveness of modularity experts in static FC analysis, we integrated the MoE-GIN layer into Neurograph framework [16] that employs a GNN architecture with residual connections and combines hidden representations from message passing at each layer by concatenation. These combined representations are then processed by an MLP layer for predictions. In Table III, incorporating the modularity experts into a *static-FC* method enhanced brain graph representations and improved performance, particularly on the HCP dataset. This highlights the versatility of modularity experts in various FC analyses.

4) Scaling Ratio α Analysis for Auxiliary Losses: We analyzed the impact of different scaling ratios α for the two auxiliary losses in the modularity experts on the HCP dataset, experimenting with $\alpha \in \{0.1, 1\}$. The results are summarized in Table IV. When $\alpha = 0.1$, performance improved on both tasks, highlighting the importance of the auxiliary losses for balance expert loading and sparse gating. With $\alpha = 1$, classification performance further improved, although the enhancement on the regression task was less significant. For consistency, we used $\alpha = 1$ in all subsequent experiments. To further validate the effectiveness of the proposed loading balance loss $L_{balance}$ and sparse gating loss L_{sparse}, we conducted experiments by removing either $L_{balance}$ or L_{sparse}. The results showed that

TABLE V
RESULTS FOR STATE EXPERTS WITH DIFFERENT SCALING RATIOS β.

β	Sex		Intelligence	
	ACC(%)	AUC(%)	MSE(↓)	CORR(↑)
0	86.75±2.43	93.52±1.12	1.023±0.065	0.210±0.072
1	88.23±2.12	96.28±1.85	0.974±0.044	0.268±0.063
10	**89.98±2.83**	**97.23±1.32**	**0.967±0.074**	**0.248±0.083**

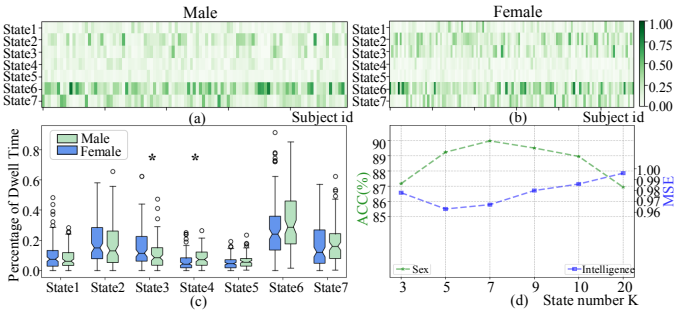

Fig. 4. **a)** and **b)** Probabilities of temporal segments of individual males and females assigned to specific states, with the *x*-axis representing individuals in the testing dataset; **c)** Boxplots summarizing **a)** and **b)**, where each box corresponds to a row in **a)** or **b)**, and ⋆ indicates statistically significant difference; **d)** Model performance with different numbers of the dFC states.

the model's performance decreased, underscoring their role in enabling balanced selection and specialization of experts.

D. State Experts Analysis

To assess the impact of prototype learning for identifying dFC states, we conducted experiments on HCP dataset to explore the performance of different numbers of states, visualize how these learned states provide insights for interpretability, and analyze the effect of the scaling factor β for L_{state}.

1) Different Numbers of States: We examined the effect of varying the number of states K within the range of 3, 5, 7, 9, 10, and 20. Fig. 4 (d) shows how the classification and prediction performance changed with the number of states. Consistent with observations from the modularity analysis, performance initially improved as K increased but began to degrade beyond a certain number of states. The best performance was achieved with $K = 7$ for the classification task and $K = 5$ for the regression task. This also aligns with the findings in existing dFC studies, which suggest that the typical number of dynamic states ranges from 5 to 7 [22].

2) Visualization of Learned dFC States: We visualized the state assignment results in Fig. 4 and the averaged FCs of temporal segments in Fig. 5 to demonstrate whether the learned states provide interpretable evidence for distinguishing males from females. Specifically, we averaged the soft assignment results P across the temporal dimension T, such that each resulting value represented the proportion of a subject's temporal segments assigned to a specific state (e.g., the k-th state), where each state captures a representative, summary patterns of the dFC measures. These proportions were computed for all subjects in the HCP test dataset and visualized separately for males and females. The results shown in Fig. 4 (a) and (b), along with the boxplot in Fig. 4 (c) reveal distinct patterns. From Fig. 4 (a) and (b), we observe that female dFC graphs are more frequently assigned to States 2 and 3 compared to male dFC graphs. Although both groups show frequent assignments to State 6, males exhibit a higher possibility. In contast, States 1, 5, and 7 show no significant differences in assignment between males and females. These findings are further supported by the boxplot in Fig. 4 (c), where notable differences are observed in States 3, 4 and 6, particularly States 3 and 4. Wilcoxon rank-sum tests indicate statistically significant differences between males and females in these states, with p-values less than 0.05. This indicates that sex differences are most pronounced in States 3 and 4, with State 3 being more characteristic of females and State 4 of males. Overall, the different state assignment patterns between females and males suggest that the learned dFC states capture meaningful and interpretable sex-related differences. These results also highlight the potential clinical relevance of clustering-derived dFC states, demonstrating that male and female groups are differentially associated with specific dFC states.

To examine how brain regions are specifically connected within the learned states and how they differ by sex, we visualized the averaged FCs of temporal segments assigned to each state for males and females. As shown in the first two rows of Fig. 5, each state exhibits distinct connectivity patterns, adaptively summarized from the temporal FC measures. In these visualizations, red indicates strong positive connectivity (correlation), while blue indicates strong negative connectivity. For example, in States 1 and 4, the DAN shows strong positive connectivity with the SN, SMN and VN. In contrast, the LN shows strong negative connectivity with all other networks in State 6. These distinct patterns across states validate the effectiveness of our state experts in aggregating similar FC patterns and capturing meaningful brain dynamics. From the female-male difference, we find that sex-related differences in FC are particularly evident in the SN and SMN, where significant differences between male and female groups are observed across nearly all states. These findings are consistent with prior studies on sex differences in functional networks [49]. Specifically, in States 1, 2, and 3, significant sex differences are observed in the connectivity between SN and DMN, SN and FPN, and SN and DAN, respectively. In State 4, significant connectivity differences appear in DMN, SMN and VN. In State 7, almost all networks exhibit significant sex-related differences, especially in the connectivity between DAN and FPN, SMN and FPN, SN and DAN, and SMN and DAN. Consistent with the state assignment results in Fig. 4, the FC patterns in States 3 and 4 also show pronounced sex differences, reinforcing the interpretability of these states. These findings demonstrate that the learned dFC states capture explainable and biologically meaningful sex differences in dFC. Moreover, they highlight the ability of our state experts to identify distinctive dFC states that align with findings from conventional dFC studies [22], [26], despite employing a different learning strategy.

Overall, the observed sex differences in both state assignment and FC strength underscore the interpretability and clinical relevance of our state experts, emphasizing their capacity to capture meaningful variations in dFC.

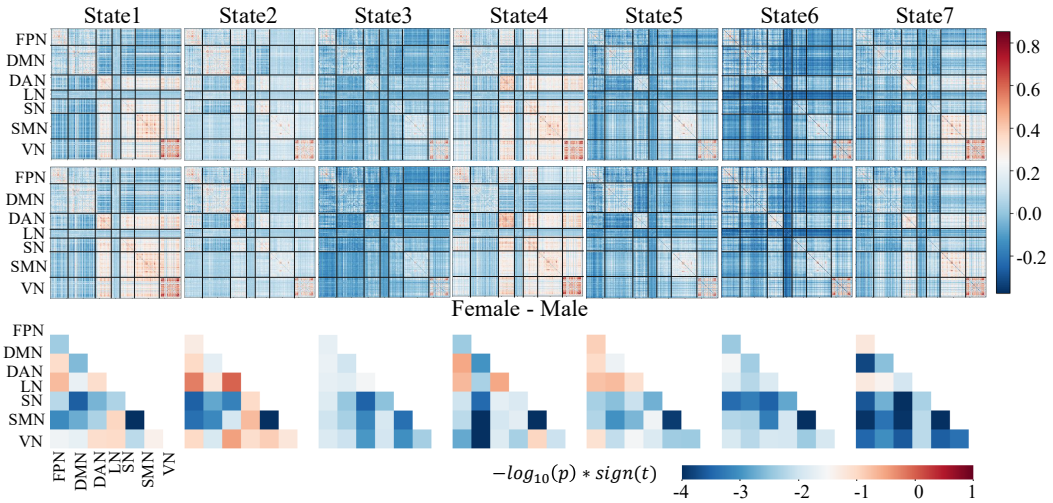

Fig. 5. Visualization of the averaged FC measures across temporal segments for males and females assigned to each specific state. The first row shows results for the female group, the second row displays results for the male group, and the third row highlights sex differences in FC. These differences were assessed using a two-sample t-test comparing the averaged FCs between female and male groups for each state. Brain network nodes were grouped into seven functional networks. The visualized differences are represented as $-log_{10}(p) \times sign(t)$, where p and t are the p-value and t-statistic from the two-sample t-test. For reference, $log_{10}(0.05) \approx -1.30$. The y-axis corresponds to the seven networks: Frontoparietal Network (FPN), Default Mode Network (DMN), Dorsal Attention Network (DAN), Limbic Network (LN), Salience Network (SN), Somatosensory Motor Network (SMN), and Visual Network (VN).

3) Scaling Factor β Analysis for L_{state}: We investigated the impact of the scaling factor β for the clustering loss L_{state} on the HCP dataset. As shown in Table V, $\beta = 1$ resulted in improved performance compared to $\beta = 0$, demonstrating the critical role of the clustering loss L_{state} in learning meaningful state patterns. Since L_{state} is relatively small in magnitude, we also tested a larger scaling factor, $\beta = 10$, to amplify its influence. The resulting performance showed further improvement, indicating enhanced state assignment.

V. Conclusions and Discussions

We developed dFCExpert, a novel method for learning effective representations of dFC measures, which consists of modularity experts and state experts, designed to reflect the brain's modular organization and dFC states that have been extensively studied in fMRI research on functional neuroanatomy, brain development and aging, and brain disorders, but underexplored in the machine learning community for functional brain networks. As demonstrated by extensive experiments on three large-scale fMRI datasets, the modularity experts automatically routed network nodes with similar brain functions to the same expert, enabling specialization of experts and promoting effective representation learning of dFC measures. Meanwhile, the state experts grouped the learned temporal graph features into distinctive states, each characterized by similar dFC patterns. This facilitated the effective modeling of temporal dynamics in brain functional networks, revealing insights into different brain states and biological characteristics. These results highlight not only the superior performance of dFCExpert compared to state-of-the-art alternative methods but also its enhanced interpretability, offering a promising tool for understanding and analyzing functional brain networks.

While dFCExpert demonstrates strong performance and interpretability, several limitations offer opportunities for future improvements. First, the model has been evaluated on three large-scale fMRI datasets (HCP, ABCD, and ABIDE), but its generalizability to other datasets with different acquisition protocols, parcellation schemes, or population characteristics remains unverified. Future studies could evaluate the model on diverse datasets, including task-based fMRI, multimodal imaging (e.g., combining fMRI with EEG), or smaller datasets to assess robustness and adaptability. Second, while the modularity and state experts improve interpretability, connecting these findings to specific clinical or neurological outcomes remains a challenge. Integrating the model with clinical datasets and leveraging the learned features for diagnostic or prognostic tasks would enhance its clinical relevance. Third, the current framework relies on predefined atlases, which may not effectively capture individual variability in brain networks [50], [51]. Future research could focus on adapting dFCExpert to support personalized functional network mapping, enabling more individualized investigations of brain function. Fourth, while the model captures dFC states effectively, it does not explicitly model state transitions or the temporal relationships between states. Incorporating temporal sequence models, such as Markov chains or advanced sequence-learning architectures (e.g., transformers or RNNs), could provide deeper insights into the dynamics of state transitions. Lastly, the current implementation for sliding window adopts a window length of 50, and window stride of 3. Although it's a standard setting for dFC analysis used in many methods [14], [26], different settings of sliding windows need to be further explored, since these settings may yield different dynamics and fluctuations for the brain activities or FCs and have different performance. Addressing these limitations through future research will enhance the robustness, efficiency, and applicability of

dFCExpert, paving the way for broader adoption in both scientific and clinical settings.

REFERENCES

[1] M. Greicius, "Resting-state functional connectivity in neuropsychiatric disorders," *Current opinion in neurology*, vol. 21, no. 4, pp. 424–430, 2008.

[2] K. Wang, M. Liang, L. Wang, L. Tian, X. Zhang, K. Li, and T. Jiang, "Altered functional connectivity in early alzheimer's disease: A resting-state fmri study," *Human brain mapping*, vol. 28, no. 10, pp. 967–978, 2007.

[3] S. M. Smith, T. E. Nichols, D. Vidaurre, A. M. Winkler, T. E. Behrens, M. F. Glasser, K. Ugurbil, D. M. Barch, D. C. Van Essen, and K. L. Miller, "A positive-negative mode of population covariation links brain connectivity, demographics and behavior," *Nature neuroscience*, vol. 18, no. 11, pp. 1565–1567, 2015.

[4] R. Jing, X. Lin, Z. Ding, S. Chang, L. Shi, L. Liu, Q. Wang, J. Si, M. Yu, C. Zhuo *et al.*, "Heterogeneous brain dynamic functional connectivity patterns in first-episode drug-naive patients with major depressive disorder," *Human Brain Mapping*, vol. 44, no. 8, pp. 3112–3122, 2023.

[5] X. Lin, R. Jing, S. Chang, L. Liu, Q. Wang, C. Zhuo, J. Shi, Y. Fan, L. Lu, and P. Li, "Understanding the heterogeneity of dynamic functional connectivity patterns in first-episode drug naïve depression using normative models," *Journal of Affective Disorders*, vol. 327, pp. 217–225, 2023.

[6] P. M. Matthews and P. Jezzard, "Functional magnetic resonance imaging," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 75, no. 1, pp. 6–12, 2004.

[7] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. Yeo, "Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri," *Cerebral cortex*, vol. 28, no. 9, pp. 3095–3114, 2018.

[8] E. M. Gordon, T. O. Laumann, B. Adeyemo, J. F. Huckins, W. M. Kelley, and S. E. Petersen, "Generation and evaluation of a cortical area parcellation from resting-state correlations," *Cerebral cortex*, vol. 26, no. 1, pp. 288–303, 2016.

[9] X. Li, Y. Zhou, N. Dvornek, M. Zhang, S. Gao, J. Zhuang, D. Scheinost, L. H. Staib, P. Ventola, and J. S. Duncan, "Braingnn: Interpretable brain graph neural network for fmri analysis," *Medical Image Analysis*, vol. 74, p. 102233, 2021.

[10] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, "Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, 2017.

[11] X. Kan, W. Dai, H. Cui, Z. Zhang, Y. Guo, and C. Yang, "Brain network transformer," *Advances in Neural Information Processing Systems*, vol. 35, pp. 25 586–25 599, 2022.

[12] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert, "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431–442, 2018.

[13] B.-H. Kim and J. C. Ye, "Understanding graph isomorphism network for rs-fmri functional connectivity analysis," *Frontiers in neuroscience*, vol. 14, p. 545464, 2020.

[14] B.-H. Kim, J. C. Ye, and J.-J. Kim, "Learning dynamic graph representation of brain connectome with spatio-temporal attention," *Advances in Neural Information Processing Systems*, vol. 34, pp. 4314–4327, 2021.

[15] Q. Wang, M. Wu, Y. Fang, W. Wang, L. Qiao, and M. Liu, "Modularity-constrained dynamic representation learning for interpretable brain disorder analysis with functional mri," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 46–56.

[16] A. Said, R. Bayrak, T. Derr, M. Shabbir, D. Moyer, C. Chang, and X. Koutsoukos, "Neurograph: Benchmarks for graph machine learning in brain connectomics," *Advances in Neural Information Processing Systems*, vol. 36, pp. 6509–6531, 2023.

[17] A. Campbell, A. G. Zippo, L. Passamonti, N. Toschi, and P. Lio, "Dyndepnet: Learning time-varying dependency structures from fmri data via dynamic graph structure learning," *arXiv preprint arXiv:2209.13513*, 2022.

[18] S. Gadgil, Q. Zhao, A. Pfefferbaum, E. V. Sullivan, E. Adeli, and K. M. Pohl, "Spatio-temporal graph convolution for resting-state fmri analysis," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23*. Springer, 2020, pp. 528–538.

[19] H. Cui, W. Dai, Y. Zhu, X. Kan, A. A. C. Gu, J. Lukemire, L. Zhan, L. He, Y. Guo, and C. Yang, "Braingb: a benchmark for brain network analysis with graph neural networks," *IEEE transactions on medical imaging*, vol. 42, no. 2, pp. 493–506, 2022.

[20] O. Sporns and R. F. Betzel, "Modular brain networks," *Annual review of psychology*, vol. 67, pp. 613–640, 2016.

[21] M. A. Bertolero, B. T. Yeo, and M. D'Esposito, "The modular and integrative functional architecture of the human brain," *Proceedings of the National Academy of Sciences*, vol. 112, no. 49, pp. E6798–E6807, 2015.

[22] E. Damaraju, E. A. Allen, A. Belger, J. M. Ford, S. McEwen, D. Mathalon, B. Mueller, G. Pearlson, S. Potkin, A. Preda *et al.*, "Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia," *NeuroImage: Clinical*, vol. 5, pp. 298–308, 2014.

[23] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium *et al.*, "The wu-minn human connectome project: an overview," *Neuroimage*, vol. 80, pp. 62–79, 2013.

[24] B. J. Casey, T. Cannonier, M. I. Conley, A. O. Cohen, D. M. Barch, M. M. Heitzeg, M. E. Soules, T. Teslovich, D. V. Dellarco, H. Garavan *et al.*, "The adolescent brain cognitive development (abcd) study: imaging acquisition across 21 sites," *Developmental cognitive neuroscience*, vol. 32, pp. 43–54, 2018.

[25] A. Di Martino, C. G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, B. Deen, S. Delmonte, I. Dinstein, B. Ertl-Wagner, D. A. Fair, L. Gallagher, D. P. Kennedy, C. L. Keown, C. Keysers, J. E. Lainhart, C. Lord, B. Luna, V. Menon, N. J. Minshew, C. S. Monk, S. Mueller, R. A. Müller, M. B. Nebel, J. T. Nigg, K. O'Hearn, K. A. Pelphrey, S. J. Peltier, J. D. Rudie, S. Sunaert, M. Thioux, J. M. Tyszka, L. Q. Uddin, J. S. Verhoeven, N. Wenderoth, J. L. Wiggins, S. H. Mostofsky, and M. P. Milham, "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism," *Mol Psychiatry*, vol. 19, no. 6, pp. 659–67, 2014.

[26] M. G. Preti, T. A. Bolton, and D. Van De Ville, "The dynamic functional connectome: State-of-the-art and perspectives," *Neuroimage*, vol. 160, pp. 41–54, 2017.

[27] H. Akbari, D. Kondratyuk, Y. Cui, R. Hornung, H. Wang, and H. Adam, "Alternating gradient descent and mixture-of-experts for integrated multimodal perception," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[28] Y. Jain, H. Behl, Z. Kira, and V. Vineet, "Damex: Dataset-aware mixture-of-experts for visual understanding of mixture-of-datasets," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[29] T. Shen, M. Ott, M. Auli, and M. Ranzato, "Mixture models for diverse machine translation: Tricks of the trade," in *International conference on machine learning*. PMLR, 2019, pp. 5719–5728.

[30] S. Kim, D. Lee, S. Kang, S. Lee, and H. Yu, "Learning topology-specific experts for molecular property prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, 2023, pp. 8291–8299.

[31] H. Wang, Z. Jiang, Y. You, Y. Han, G. Liu, J. Srinivasa, R. Kompella, Z. Wang *et al.*, "Graph mixture of experts: Learning on large-scale graphs with explicit diversity modeling," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[32] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.

[33] M. I. Jordan and R. A. Jacobs, "Hierarchical mixtures of experts and the em algorithm," *Neural computation*, vol. 6, no. 2, pp. 181–214, 1994.

[34] W. Fedus, B. Zoph, and N. Shazeer, "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity," *Journal of Machine Learning Research*, vol. 23, no. 120, pp. 1–39, 2022.

[35] K. Xu, C. Li, Y. Tian, T. Sonobe, K.-i. Kawarabayashi, and S. Jegelka, "Representation learning on graphs with jumping knowledge networks," in *International conference on machine learning*. PMLR, 2018, pp. 5453–5462.

[36] N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. Le, G. Hinton, and J. Dean, "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer," *arXiv preprint arXiv:1701.06538*, 2017.

[37] J. Huang, S. Gong, and X. Zhu, "Deep semantic clustering by partition confidence maximisation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8849–8858.

[38] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *International conference on machine learning*. PMLR, 2016, pp. 478–487.

[39] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni *et al.*, "The minimal preprocessing pipelines for the human connectome project," *Neuroimage*, vol. 80, pp. 105–124, 2013.

[40] E. Feczko, G. Conan, S. Marek, B. Tervo-Clemmens, M. Cordova, O. Doyle, E. Earl, A. Perrone, D. Sturgeon, R. Klein *et al.*, "Adolescent brain cognitive development (abcd) community mri collection and utilities," *BioRxiv*, pp. 2021–07, 2021.

[41] A. S. Keller, A. R. Pines, S. Shanmugan, V. J. Sydnor, Z. Cui, M. A. Bertolero, R. Barzilay, A. F. Alexander-Bloch, N. Byington, A. Chen *et al.*, "Personalized functional brain network topography is associated with individual differences in youth cognition," *Nature communications*, vol. 14, no. 1, p. 8411, 2023.

[42] J. Xu, Y. Yang, D. Huang, S. S. Gururajapathy, Y. Ke, M. Qiao, A. Wang, H. Kumar, J. McGeown, and E. Kwon, "Data-driven network neuroscience: On data collection and benchmark," in *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36. Curran Associates, Inc., 2023, pp. 21 841–21 856. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/44e3a3115ca26e5127851acd0cedd0d9-Paper-Datasets_and_Benchmarks.pdf

[43] A. Di Martino, D. O'connor, B. Chen, K. Alaerts, J. S. Anderson, M. Assaf, J. H. Balsters, L. Baxter, A. Beggiato, S. Bernaerts *et al.*, "Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii," *Scientific data*, vol. 4, no. 1, pp. 1–15, 2017.

[44] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[45] A. Zalesky and M. Breakspear, "Towards a statistical test for functional connectivity dynamics," *Neuroimage*, vol. 114, pp. 466–470, 2015.

[46] B. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni *et al.*, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of neurophysiology*, 2011.

[47] G. Carpaneto and P. Toth, "Algorithm 548: Solution of the assignment problem [h]," *ACM Transactions on Mathematical Software (TOMS)*, vol. 6, no. 1, pp. 104–111, 1980.

[48] A. F. Alexander-Bloch, H. Shou, S. Liu, T. D. Satterthwaite, D. C. Glahn, R. T. Shinohara, S. N. Vandekar, and A. Raznahan, "On testing for spatial correspondence between maps of human brain structure and function," *Neuroimage*, vol. 178, pp. 540–551, 2018.

[49] E. Dhamala, D. S. Bassett, B. T. Yeo, and A. J. Holmes, "Functional brain networks are associated with both sex and gender in children," *Science Advances*, vol. 10, no. 28, p. eadn4202, 2024.

[50] H. Li, T. D. Satterthwaite, and Y. Fan, "Large-scale sparse functional networks from resting state fmri," *NeuroImage*, vol. 156, pp. 1–13, 2017.

[51] H. Li, D. Srinivasan, C. Zhuo, Z. Cui, R. E. Gur, R. C. Gur, D. J. Oathes, C. Davatzikos, T. D. Satterthwaite, and Y. Fan, "Computing personalized brain functional networks from fmri using self-supervised deep learning," *Medical Image Analysis*, vol. 85, p. 102756, 2023.