DYNAMIC BRAIN TRANSFORMER WITH MULTI-LEVEL ATTENTION FOR FUNCTIONAL BRAIN NETWORK ANALYSIS

Xuan Kan[†]   *Antonio Aodong Chen Gu*[‡]   *Hejie Cui*[†]   *Ying Guo*[†]   *Carl Yang*[†*]

[†] Emory University, [‡] Georgia Institute of Technology

ABSTRACT

Recent neuroimaging studies have highlighted the importance of network-centric brain analysis, particularly with functional magnetic resonance imaging. The emergence of Deep Neural Networks has fostered a substantial interest in predicting clinical outcomes and categorizing individuals based on brain networks. However, the conventional approach involving static brain network analysis offers limited potential in capturing the dynamism of brain function. Although recent studies have attempted to harness dynamic brain networks, their high dimensionality and complexity present substantial challenges. This paper proposes a novel methodology, Dynamic bRAin Transformer (DART), which combines static and dynamic brain networks for more effective and nuanced brain function analysis. Our model uses the static brain network as a baseline, integrating dynamic brain networks to enhance performance against traditional methods. We innovatively employ attention mechanisms, enhancing model explainability and exploiting the dynamic brain network's temporal variations. The proposed approach offers a robust solution to the low signal-to-noise ratio of blood-oxygen-level-dependent signals, a recurring issue in direct DNN modeling. It also provides valuable insights into which brain circuits or dynamic networks contribute more to final predictions. As such, DRAT shows a promising direction in neuroimaging studies, contributing to the comprehensive understanding of brain organization and the role of neural circuits.

Index Terms— Dynamic Brain Networks, Deep Learning

1. INTRODUCTION

Network-centric analysis on brain imaging has gained substantial attention in neuroimaging studies recently, contributing profoundly to our understanding of brain organization in healthy individuals and those with brain disorders (1). Neuroscience research has consistently demonstrated that insights into neural circuits are pivotal for distinguishing brain function across diverse populations, with disruptions in these circuits often instigating and delineating brain disorders (2). Functional magnetic resonance imaging (fMRI) has emerged as a widely employed imaging modality for exploring brain
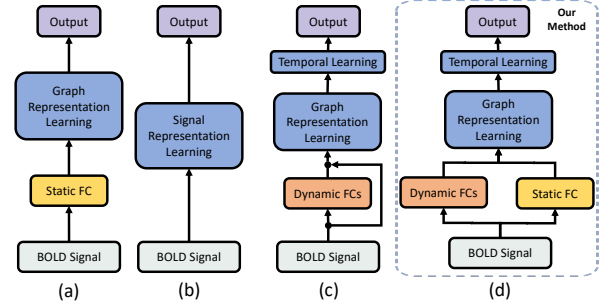
---

*Corresponding author: Carl Yang <j.carlyang@emory.edu>

Fig. 1. Four distinct schemas when employing Deep Learning for brain network analysis. From fMRI imaging, three types of input data can be acquired: (1) raw time-series data (BOLD signals), (2) static functional connectivity (FC), and (3) dynamic FCs, which capture temporal changes. Both kinds of FC are derived from the BOLD signal. Our method is the first attempt to combine Static FCs and dynamic FCs.

function and organization (3). Predicting clinical outcomes or categorizing individuals based on brain networks extracted from fMRI images with deep neural networks is a topic of significant interest in the neuroimaging community (4; 5; 6; 7).

Figure 1 succinctly summarizes various schemas used for analyzing brain networks with neural networks. The classic approach to network analyses primarily relies on using individual fMRI data to construct functional brain networks (8). This established process involves selecting a brain atlas or regions of interest (ROI), extracting fMRI blood-oxygen-level-dependent (BOLD) signal series from each node or region, and computing pairwise connectivity measures. Once static brain networks are obtained, various neural network models can be applied for downstream analyses, as demonstrated in Figure 1 (a). There have been attempts to model BOLD signals directly using deep neural networks (DNNs), as seen in Figure 1 (b), but these have generally yielded unsatisfactory results due to the low signal-to-noise ratio of BOLD signals (9; 10). However, recent works have attempted to use dynamic brain networks to replace static ones for downstream analyses (11; 12). These dynamic networks are created by segmenting BOLD signals into several overlapping or non-overlapping windows, each contributing to a unique connectivity matrix. This strategy shown in Figure 1 (c) allows for exploring temporal variations and state transitions in functional connectivity over time, providing crucial insights
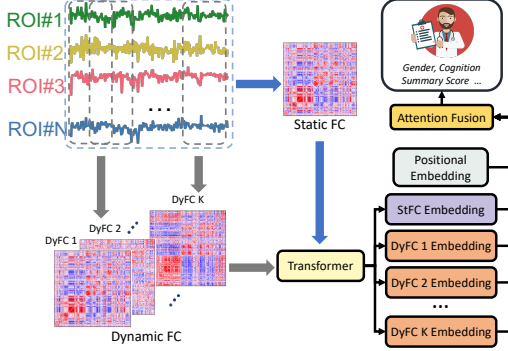
Fig. 2. Diagram illustrating the comprehensive workflow of the proposed methodology, DRAT.

into brain function. However, there is significant space for improvement due to the high dimensionality and complexity of dynamic brain networks. In response to these challenges, this paper proposes a novel methodology, Dynamic bRAin Transformer (DRAT), depicted in Figure 1 (d). DRAT exploits static brain networks as a foundation measurement to integrate dynamic brain networks, thereby improving performance against benchmark methods. Additionally, we incorporate specific attention mechanisms to enhance model explainability, aiming to capitalize on dynamic brain networks' switching in neuroimaging studies.

2. METHOD

Table 1. Performance comparison with baselines. The $\uparrow$ indicates a higher metric value is better, while $\downarrow$ is opposite.

Type	Method	PNC		ABCD
		AUROC$\uparrow$	Accuracy$\uparrow$	MSE$\downarrow$
Dynamic	STAGIN	63.5±4.0	54.2±1.4	102.4±6.1
	ST-GCN	64.7±3.5	57.3±3.2	89.2±11.2
Static	BrainGNN	62.4±3.5	59.4±2.3	80.8±4.7
	BrainGB	69.7±3.3	63.6±1.9	78.1±4.3
	BrainNetCNN	74.9±2.4	67.8±2.7	77.1±4.5
	BNT	78.2±1.9	70.6±2.1	60.2±1.5
Dynamic & Static	DRAT	**80.7±3.1**	**72.5±2.3**	**58.3±3.5**

In this section, we elaborate on the design of DRAT and its four main components as shown in Figure 2. Specifically, the input $\boldsymbol{X} \in \mathbb{R}^{v \times T}$ denotes the BOLD time-series for regions of interest (ROIs) represents a sample (individual), v is the number of ROIs, and T is the length of time-series. We set L as the window size, S as the stride size. Given a sample $\boldsymbol{X}$, we can obtain k dynamic brain networks, where $k = \lfloor \frac{T-L+S}{S} \rfloor$. For the classification task, the target output is the prediction label $\boldsymbol{Y} \in \mathbb{R}^{|\mathcal{C}|}$, where $\mathcal{C}$ is the class set of Y and $|\mathcal{C}|$ is the number of classes. For the regression task, the target output is the prediction label $\boldsymbol{Y} \in \mathbb{R}$.

Static FC Generation. We begin by generating a static functional connectivity (FC) matrix, which provides a summary of the overall functional connections in the brain during the entire scan period. This static FC, $\boldsymbol{A} \in \mathbb{R}^{v \times v}$, represents the connectivity matrix between all pairs of ROIs for each individual. Specifically, we use the Pearson Correlation as a measure of statistical dependence between the time series of different ROIs. Each element of the static FC, $\boldsymbol{A}_{ij}$, is computed as $\mathrm{Corr}(\boldsymbol{X}_i, \boldsymbol{X}_j)$, which denotes the correlation between the time-series of ROI i and ROI j. This matrix captures the overall brain functional organization and serves as an anchor for the subsequent steps.

Dynamic FC Generation. To generate dynamic brain networks, we partition the BOLD signal into a series of overlapping or non-overlapping windows of length T, with a stride of size S. We calculate a dynamic functional connectivity matrix for the time window t, $\boldsymbol{D}^t \in \mathbb{R}^{v \times v}$. Finally, we can obtain k dynamic functional connectivity matrix, where k is the total number of dynamic networks given by $k = \lfloor \frac{T-L+S}{S} \rfloor$. Each of these matrices represents a snapshot of brain connectivity at a specific time point. Similar to the static FC generation, we use Pearson Correlation for this computation. Each element in the dynamic connectivity matrix, $\boldsymbol{D}^t_{ij}$, is then calculated as $\mathrm{Corr}(\boldsymbol{X}^t_i, \boldsymbol{X}^t_j)$, where $\boldsymbol{X}^t_i$ and $\boldsymbol{X}^t_j$ are the BOLD signals for ROIs i and j at the time window t.

Edge-level Attention and Transformer Projection. After generating both static and dynamic FCs, we utilize the graph transformer proposed by (5) for processing these matrices. This transformer comprises a Multi-Head Self-Attention Module, which is adept at capturing complex dependencies between different nodes in the network, thus enabling a rich representation of the FCs. A learnable clustering readout function is applied to compress the matrix into a graph-level embedding. In the case of the static brain network, the hidden representation, $\boldsymbol{h}_A = \boldsymbol{f}_{\mathrm{TF}}(A)$, is obtained. In contrast, for each dynamic brain network, the hidden representation is equipped with an attention layer $\boldsymbol{\alpha}$ and added by a positional embedding as delineated in (13), resulting in $\boldsymbol{h}^t_D = \boldsymbol{f}_{\mathrm{TF}}(\boldsymbol{\alpha} \circ \boldsymbol{D}^t) + \boldsymbol{P}^t$. Here, $\circ$ denotes the Hadamard product, $\boldsymbol{\alpha} \in \mathbb{R}^{v \times v}$ represents a learnable attention matrix (initialized at 1) shared across all dynamic networks, and $\boldsymbol{P}^t$ refers to the positional embedding at time window t. The attention mechanism based on $\boldsymbol{\alpha}$ allows the model to focus on the most informative connections in the brain networks.

Temporal-level Attention and Fusion. Given the hidden embedding of static FC $\boldsymbol{h}_A$ and the sequence of dynamic FCs' hidden embedding $\boldsymbol{h}^t_D$, we utilize an attention mechanism to fuse these networks. The attention scores are computed based on the similarity between $\boldsymbol{h}_A$ and each $\boldsymbol{h}^t_D$. Specifically, the attention score $\boldsymbol{\beta}^t$ for each dynamic FC $\boldsymbol{h}^t_D$ is calculated as $\boldsymbol{\beta}^t = \mathrm{softmax}(\mathrm{sim}(\boldsymbol{h}_A, \boldsymbol{h}^t_D))$, where $\mathrm{sim}(.)$ is a similarity function, such as dot product. Then, the fused FC $\boldsymbol{F}$ is generated as a weighted sum of dynamic FCs, i.e., $\boldsymbol{F} = \sum_{t=1}^{k} \boldsymbol{\beta}^t \cdot \boldsymbol{h}^t_D$. The final FC $\boldsymbol{F}$ is then fed into the multi-layer perception module for final prediction. The fusion mechanism thus enables the model to direct its focus towards dynamic FCs bearing higher similarity to the static FC, effectively leveraging the static FC's stable functional information to guide the dynamic FC fusion.

3. EXPERIMENTS

3.1. Experimental Settings

Dataset. This study utilizes two public neuroimaging datasets. The first is the Adolescent Brain Cognitive Development Study (ABCD), one of the largest publicly available fMRI datasets with stringent access control (14). We employ fully anonymized brain networks based on the HCP 360 ROI atlas (15) and define a task, the Cognition Summary Score Prediction, a regression problem focused on five cognitive sub-domains. Considering the variability in sequence lengths within the ABCD dataset, we included only those samples with a sequence length exceeding 1024, truncating them to form a unified length dataset, yielding 4613 samples for regression analysis. The second dataset is the Philadelphia Neuroimaging Cohort (PNC), whose individuals aged 8–21 years provided by the Children's Hospital of Philadelphia (16). After quality control, the dataset includes 503 subjects, each providing 120 timesteps of data from 264 nodes (17).

Metric. To evaluate performance in binary classification tasks, we employ two widely accepted metrics: Area Under the Receiver Operating Characteristic (AUROC) and accuracy. We set the classification threshold at 0.5 to determine the final class labels. In the case of regression tasks, we utilize the Mean Square Error (MSE) as a comprehensive measure of model performance. Please note, all the results presented in this study are the mean values derived from five independent runs, each initiated with a different random seed, to ensure the robustness and reproducibility of our findings.

Implementation. We configure the window size L and stride size as 24 to ensure each window encapsulates a one-minute BOLD signal. The architecture of our Transformer is built according to the design described in (5), setting the number of transformer layers to 2, matching the hidden dimension for each transformer layer with the number of nodes v, and employing 4 heads. We divide our datasets such that 70% is utilized for training, 10% for validation, and the remainder for testing. We leverage the Adam optimizer throughout the training process with a learning rate and weight decay set at 10^{-4}. Our batch size is 16, and all models undergo 200 training epochs. The epoch displaying optimal performance on the validation set is chosen for the final report.

3.2. Performance and Analysis

Our model's performance is benchmarked against several state-of-the-art methodologies in brain network analysis, and the result can be found in Table 1. We consider methods that leverage both static and dynamic brain networks for comparison. The baseline models include STAGIN (12), which constructs dynamic brain networks and fuses them using an attention mechanism without considering static brain networks; ST-GCN (18), an improved version of GCN that takes into account not only the current graph but also the adjacency

of prior and future graphs; BrainGNN (19) and BrainGB (20), two Graph Neural Networks designed explicitly for static brain networks; BrainNetCNN (4), a convolutional neural network model designed for static brain networks; and finally BNT (5), a graph transformer model also designed for static brain networks, which is the same transformer that we employ in our model to project brain networks into an embedding.

The comparative analysis with these methods demonstrates several vital insights. Models that rely exclusively on dynamic brain networks perform the poorest, underlining the critical role of global information provided by static brain networks in predictions. Static brain networks were found to encompass the most significant predictive signals, which illustrates our strategy of using the hidden representations of these static networks as anchor points (or query embeddings) to fuse dynamic brain representations. Furthermore, integrating dynamic brain networks is observed to enhance model performance because it exploits the fine-grained variations in brain states. As a result, our proposed methodology, denoted as DRAT, consistently outperforms all baseline methods, exhibiting the highest performance across various metrics on two datasets with both regression and classification tasks.

3.3. Attention Visualization and Analysis

Due to the special attention design in DRAT, the proposed method enables two-level attention-driven interpretability - edge-level attention (α), representing the significance of each edge, and temporal-level attention (β), indicating the importance of each dynamic brain network. As evidenced in Figure 3, during the first 100 epochs, attention weights progressively evolve, stabilizing over the subsequent 100 epochs. The edge-level attention, initially uniform, progressively concentrates on the Visual and Default Mode sub-networks, aligning with the scores evaluated from heavy visual tasks like the Picture Vocabulary and Picture Sequence Memory Tests. Meanwhile, the temporal-level attention highlights dynamic brain networks recorded during the middle collection. Given the fact that the ABCD dataset is collected from children, it is plausible that resting-state brain activity is most prominent for children during the middle of the data collection process, corroborating our temporal attention insights.

4. CONCLUSION

Our proposed method, DRAT, addresses the challenges presented by dynamic brain networks' complexity and high dimensionality, thereby advancing brain network analysis in neuroimaging. By leveraging static brain networks as a foundational measurement, we successfully integrate dynamic brain networks and improve performance compared to standard methods. The specific attention mechanisms incorporated within DRAT further enhance model explainability.
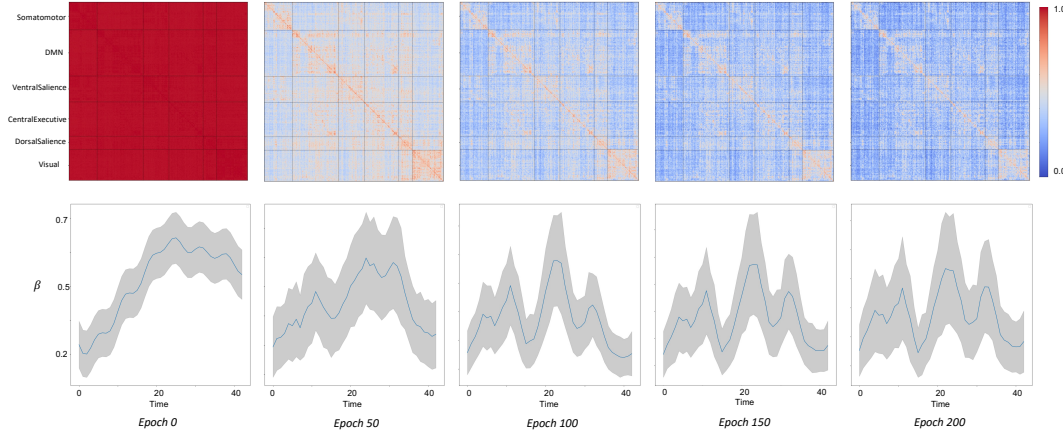
Fig. 3. Evolution of two-level attention during the training on the ABCD dataset. The first row displays the progression of edge-level attention (α) across epochs, while the second row shows the changes in temporal-level attention (β) across epochs.

References

[1] Ed Bullmore and Olaf Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, pp. 186–198, 2009.

[2] Thomas R Insel and Bruce N Cuthbert, "Brain disorders? precisely," *Science*, vol. 348, pp. 499–500, 2015.

[3] Stephen M Smith, "The future of fmri connectivity," *Neuroimage*, vol. 62, pp. 1257–1266, 2012.

[4] Jeremy Kawahara et al., "BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, 2017.

[5] Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang, "Brain network transformer," in *NeurIPS*, 2022.

[6] Hejie Cui, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, and Carl Yang, "Interpretable graph neural networks for connectome-based brain disorder analysis," in *MICCAI*, 2022.

[7] 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 *EMBC*, 2022.

[8] Stephen M. Smith, Karla L. Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F. Beckmann, Thomas E. Nichols, Joseph D. Ramsey, and Mark W. Woolrich, "Network modelling methods for FMRI," *NeuroImage*, vol. 54, 2011.

[9] Xuan Kan, Hejie Cui, Joshua Lukemire, Ying Guo, and Carl Yang, "Fbnetgen: Task-aware gnn-based fmri analysis via functional brain network generation," in *MIDL*. PMLR, 2022, pp. 618–637.

[10] Yue Yu, Xuan Kan, Hejie Cui, Ran Xu, Yujia Zheng, Xiangchen Song, Yanqiao Zhu, Kun Zhang, Razieh Nabi, Ying Guo, Chao Zhang, and Carl Yang, "Deep dag learning of effective brain connectivity for fmri analysis," in *ISBI*, 2023.

[11] Lingwen Liu, Guangqi Wen, Peng Cao, Tianshun Hong, Jinzhu Yang, Xizhe Zhang, and Osmar R. Zaiane, "Braintgl: A dynamic graph representation learning model for brain network analysis," *Computers in Biology and Medicine*, vol. 153, pp. 106521, 2023.

[12] Byung-Hoon Kim, Jong Chul Ye, and Jae-Jin Kim, "Learning dynamic graph representation of brain connectome with spatio-temporal attention," *NeurIPS*, vol. 34, pp. 4314–4327, 2021.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *NeurIPS*, vol. 30, 2017.

[14] B.J. Casey, Tariq Cannonier, and May I. Conley et al., "The adolescent brain cognitive development (abcd) study: Imaging acquisition across 21 sites," *Developmental Cognitive Neuroscience*, pp. 43–54, 2018.

[15] Matthew F. Glasser et al., "The minimal preprocessing pipelines for the human connectome project," *NeuroImage*, vol. 80, pp. 105–124, 2013.

[16] Theodore D. Satterthwaite et al., "Neuroimaging of the Philadelphia Neurodevelopmental Cohort," *NeuroImage*, vol. 86, pp. 544–553, 2014.

[17] Jonathan D Power et al., "Functional Network Organization of the Human Brain," *Neuron*, vol. 72, pp. 665–678, 2011.

[18] Soham Gadgil et al., "Spatio-temporal graph convolution for resting-state fmri analysis," in *MICCAI*. Springer, 2020, pp. 528–538.

[19] Xiaoxiao Li et al., "Braingnn: Interpretable brain graph neural network for fmri analysis," *Medical Image Analysis*, 2021.

[20] Hejie Cui et al., "BrainGB: A Benchmark for Brain Network Analysis with Graph Neural Networks," *IEEE Transactions on Medical Imaging (TMI)*, 2022.