Brain Network State Transformer: Leveraging State Functional Connectivity for Enhanced Brain Network Analysis

Jiawei Nie[1], Keqi Han[1], Tianyi Zhang[1], Chenyu You[2], Sanne van Rooij[3], Jennifer Stevens[3], Boadie Dunlop[3], Charles Gillespie[3], Carl Yang[1*]

***Abstract*— Recent advancements in functional Magnetic Resonance Imaging (fMRI) have highlighted the importance of capturing the dynamic nature of brain activities, prompting a shift from static Functional Connectivity (FC) to dynamic FC (DFC). However, existing DFC approaches often struggle to balance temporal granularity with interpretability, leading to challenges in disentangling meaningful connectivity patterns. In this work, we introduce the Brain Network State Transformer (BNST), a novel framework that leverages State FC to enhance brain network analysis. Our approach integrates three key steps: (1) Deep Clustering to identify recurring brain states from high-dimensional DFC matrices, (2) State-Based Rechunking to reorganize BOLD time series according to these states, and (3) a Transformer-Based Feature Extraction mechanism that models intra-state and inter-state relationships for downstream prediction tasks. We demonstrate the effectiveness of BNST on two publicly available fMRI datasets—ABCD and HCP—across both classification and regression tasks. By capturing structured temporal dynamics, BNST not only boosts prediction performance but also improves interpretability by identifying distinct brain states and their functional significance, providing a structured representation that aligns with meaningful cognitive and neural processes.**

I. INTRODUCTION

Brain network analysis has emerged as a pivotal field for understanding the organization of the human brain, facilitating the identification of neurological biomarkers and the development of enhanced diagnostic and therapeutic strategies [1]–[3]. Central to this endeavor are functional brain networks, which are constructed from blood-oxygen-level-dependent (BOLD) signals. In these networks, nodes represent predefined regions of interest (ROIs), while edges capture the correlations between ROIs' signals [4], [5]. A major focus in neuroimaging is using these brain networks to predict clinical outcomes or classify individuals, which has driven the development of deep learning approaches that extract meaningful patterns from this complex data [6], [7].

A longstanding approach to constructing these networks relies on static functional connectivity (FC), where correlations between ROIs are calculated over the full duration of the scan, forming a single static FC per subject [8]. This static representation has proven effective for many prediction tasks due to its simplicity and stability. However, while this time-averaged approach has been instrumental in advancing our understanding of brain connectivity, collapsing the entire time series into a single FC inherently overlooks transient or evolving patterns that characterize the dynamic nature of neural activity, leaving substantial room for improvement [9], [10].

To capture this temporal information, DFC-based approaches have been developed more recently [11], [12]. A common method involves using the sliding window technique [13], which divides the BOLD time series into fixed-length segments and computes a separate connectivity matrix for each segment. By producing a sequence of time-resolved connectivity matrices, the sliding window technique enables neuroscientists to capture the evolving patterns of FC. However, the high dimensionality and sheer volume of these temporal snapshots can introduce additional noise, heighten computational complexity, and complicate interpretation. [14], [15].

In this work, we propose a novel framework called Brain Network State Transformer (BNST) that addresses the limitations of both static and dynamic FC by leveraging the concept of brain states [16]. Brain states are distinct and recurring patterns of DFC that emerge as the brain transitions through various modes of activity over time. Different states reflect different stable configurations of interactions between ROIs. [17]. These states are typically derived by clustering DFC matrices, with each state characterized by a cluster centroid that encapsulates a representative connectivity pattern [18]. BNST utilizes brain states as a guiding mechanism to reorganize the BOLD time series based on each segment's state correspondence. It begins with a deep clustering approach to identify recurring brain states from DFC matrices generated by the sliding windows. Using these states, a state-based rechunking mechanism reorganizes the BOLD time series data by grouping time segments corresponding to the same brain state. These time series segments belonging to the same state are concatenated and converted into state-specific FC matrices by calculating the correlation within the concatenated BOLD time series for each state, providing a structured representation of brain dynamics. These matrices are then processed by Brain Network Transformer (BNT) [19], a model that is chosen for its proven performance and widespread adoption in the field, to extract meaningful features as embeddings. The embeddings are fed into a multi-

*Carl Yang is the corresponding author.

[1]Jiawei Nie, Keqi Han, Tianyi Zhang, and Carl Yang are with the Department of Computer Science, Emory University, Atlanta, USA {jiawei.nie,keqi.han,tianyi.zhang2,j.carlyang}@emory.edu

[2]Chenyu You is with the Department of Computer Science , Stony Brook University , Stony Brook, USA chenyu.you@stonybrook.edu

[3]Sanne van Rooij , Jennifer Stevens , Boadie Dunlop , and Charles Gillespie are with the Department of Psychiatry and Behavioral Sciences , Emory University, Atlanta, USA {sanne.van.rooij,jennifer.stevens,bdunlop,cgilles}@emory.edu

layer perceptron (MLP) for downstream tasks like prediction of gender and cognitive score.

Experiments on two large-scale neuroimaging datasets, Adolescent Brain Cognitive Development (ABCD) [20] and Human Connectome Project (HCP) [21], demonstrate that BNST achieves significant improvements in predictive performance compared to both static FC-based and DFC-based approaches. These results highlight the strength of State FC, which combines the stability of static FC with the adaptability of DFC to enhance predictions by capturing long-term stable interactions and transient changes in connectivity, ensuring that the model emphasizes the most informative aspects of brain connectivity while reducing noise. By identifying recurring and distinct patterns of connectivity, State FC isolates stable configurations of brain activity that reflect meaningful functional states. This structured representation allows the model to leverage state-specific connectivity features, which provide richer and more interpretable insights into the temporal and spatial organization of brain networks. Additionally, the brain states also offer valuable information about how the brain shifts between functional modes, which can be critical for understanding variability in cognitive or clinical outcomes. As a result, this balanced and comprehensive approach enables superior predictive performance and enhanced interpretability into the brain's functional organization.

II. The Proposed Approach: BNST

A. Problem Definition

Given BOLD time-series data $\{\mathbf{X}_i \in \mathbb{R}^{T\times N}, y_i\}_{i=1}^{S}$, where T is the length of the time series, N is the number of brain regions, S is the number of subjects, and y_i represents the label corresponding to the i-th subject such as gender, cognitive scores, or neurological conditions, the goal is to accurately predict these subject-level outcomes.

B. The Overall Framework

As illustrated in Figure 1, the BNST framework begins by constructing DFC matrices using a sliding window approach applied to the BOLD time series. It then identifies recurring brain states by clustering these DFC matrices, grouping similar patterns into distinct states that reflect stable configurations of brain network activity over time. These recurring brain states form the foundation of the BNST framework and guide subsequent stages of the pipeline. Based on the identified brain states, the BOLD time series is reorganized for each sample to produce state-specific FC matrices, capturing structured temporal dynamics. Finally, a BNT is used to analyze both intra-state and inter-state relationships, generating embeddings that enable accurate predictions for downstream tasks, such as classification or regression.

C. Clusetering to obtain Brain States

Data Preprocessing. To discover recurring brain states from BOLD time series, we designed a robust preprocessing pipeline with the following steps:

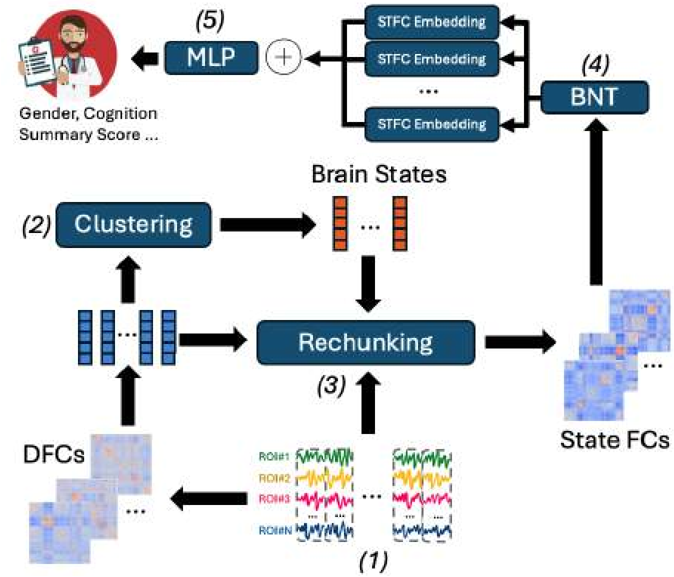

Fig. 1. The overall architecture of the BNST framework, consisting of five steps: *(1)* Sliding Window Method to generate DFCs, *(2)* Clustering to obtain Brain States, *(3)* State-Based Rechunking, *(4)* feature extraction with the BNT, and *(5)* MLP for downstream task prediction.

- **Sliding Window Segmentation:** To construct dynamic brain networks, the BOLD signal for each sample is segmented into a sequence of overlapping or non-overlapping windows of size L, with a stride length of D. For each sample i, the total number of DFC matrices is
$$W = \left\lfloor \frac{T-L}{D} + 1 \right\rfloor.$$
Each window produces a DFC matrix $\mathbf{F}_{i,t} \in \mathbb{R}^{N\times N}$, where $t = 1, \ldots, W$. These matrices provide snapshots of the brain's connectivity at different time points. Pearson correlation is used to compute the connectivity strength, such that $F_{xy}^t = \mathrm{Corr}(X_x^t, X_y^t)$, where X_x^t and X_y^t represent the BOLD signals for regions x and y within the time window t.
- **Fisher-z Normalization:**To stabilize variance and improve comparability across subjects, we applied Fisher-z transformation [22] to all DFC matrices:
$$z_{xy} = \frac{1}{2}\ln\left(\frac{1+F_{xy}}{1-F_{xy}}\right),$$
- **Vectorization:** Each Fisher-z transformed FC matrix $\mathbf{Z}_{i,t} \in \mathbb{R}^{N\times N}$ is vectorized by flattening its upper triangular elements to form $\mathbf{v}_{i,t} \in \mathbb{R}^d$, where $d = \frac{N(N-1)}{2}$. This process produces a feature matrix $\mathbf{V}_i \in \mathbb{R}^{W\times d}$ for each sample i, suitable for subsequent dimensionality reduction and clustering.

Deep Clustering. As illustrated in Fig. 2, we employed a deep clustering framework inspired by [18] to identify brain states. Before clustering, a fully connected autoencoder is trained to project the high-dimensional vectors $\mathbf{v}_{i,t}$ into a lower-dimensional space $\mathbf{l}_{i,t} \in \mathbb{R}^q$, where $q \ll d$. The

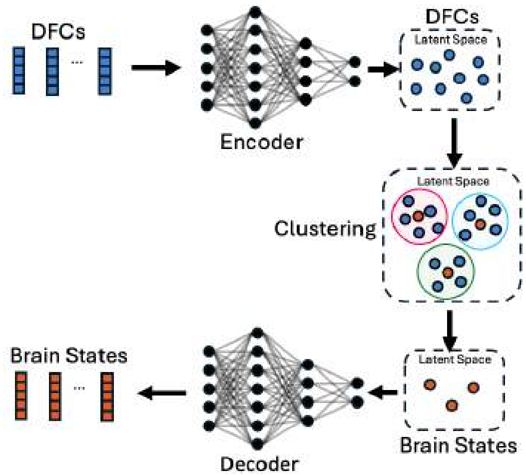

Fig. 2. Deep clustering framework for identifying brain states from DFCs. The autoencoder projects high-dimensional DFCs into a latent space, where k-means clustering extracts recurring brain states.

architecture consists of an encoder, which is a series of fully connected layers that compress $\mathbf{v}_{i,t}$ into the latent representation $\mathbf{l}_{i,t}$, and a decoder, which is a symmetric set of layers that reconstruct $\mathbf{v}_{i,t}$ from $\mathbf{l}_{i,t}$ to ensure the compressed representation retains critical features. The autoencoder is trained to minimize the mean squared error (MSE) between the input $\mathbf{v}_{i,t}$ and the reconstructed output $\hat{\mathbf{v}}_{i,t}$:

$$\mathcal{L} = \frac{1}{S \cdot W} \sum_{i=1}^{S} \sum_{t=1}^{W} \|\mathbf{v}_{i,t} - \hat{\mathbf{v}}_{i,t}\|^2$$

As shown in Fig. 2, after the training process, the encoder is employed to generate the latent representations $\{\mathbf{l}_{i,t}\}_{t=1}^{W}$ for each sample i, forming a lower-dimensional representation of the DFCs $\mathbf{L}_i \in \mathbb{R}^{W \times q}$.

The latent representations $\mathbf{L}_i$ from all samples are aggregated into a single matrix $\mathbf{L} \in \mathbb{R}^{S \times W \times q}$. This matrix is then clustered into K states using the k-means algorithm in the latent space, resulting in K clusters. Each cluster is represented by its centroid μ_k in the latent space.

To interpret these clusters in the original feature space, the centroids μ_k are projected back to the original space using the decoder. The resulting vectors represent the centroids of the clusters in the original feature space and are referred to as the brain states $\mathbf{c}_k$.

D. State-Based Rechunking

Figure 3 summarizes the State-Based Rechunking process. It reorganizes the BOLD time-series signals $\mathbf{X}_i$ based on the clustering assignments of the DFC matrices. For each sample i, let $\mathcal{S}_{i,k}$ denote the set of BOLD time-series chunks assigned to state k, and the process produces an ordered set of State FC matrices $\{\mathbf{F}_i^{(k)}\}_{k=1}^{K}$, where K is the total number of states identified during clustering. The steps are as follows:

- **State Assignment:** Each DFC vector $\mathbf{v}_{i,t}$ is assigned to the state k with the closest brain states $\mathbf{c}_k$ (from the clustering stage) based on Euclidean distance:

$$k_{i,t} = \arg\min_k \|\mathbf{v}_{i,t} - \mathbf{c}_k\|,$$

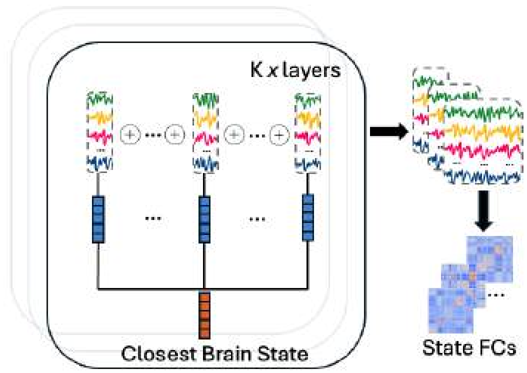

Fig. 3. Each BOLD time-series segment is assigned to its closest brain state, grouping similar patterns together. The reorganized time-series are then concatenated to compute State FC matrices, capturing structured brain dynamics.

For each sample i, the set of time-series chunks assigned to state k is then defined as:

$$\mathcal{S}_{i,k} = \{\mathbf{X}_{i,t} \mid k_{i,t} = k\}.$$

- **Chunk Extraction and Concatenation:** For each state k, the BOLD time-series segments corresponding to $\mathcal{S}_{i,k}$ are extracted from $\mathbf{X}_i$. These segments are concatenated together to form a continuous, state-specific time-series block:

$$\mathbf{X}_i^{(k)} = \text{Concat}(\{\mathbf{X}_{i,t} \mid t \in \mathcal{S}_{i,k}\}),$$

where $\mathbf{X}_i^{(k)} \in \mathbb{R}^{T_{i,k} \times N}$, and $T_{i,k}$ is the total length of time-series assigned to state k for sample i.
- **State FC Computation:** For each state k, if $\mathcal{S}_{i,k}$ is non-empty ($T_{i,k} > 0$), a State FC matrix $\mathbf{F}_i^{(k)} \in \mathbb{R}^{N \times N}$ is computed for the entire concatenated block $\mathbf{X}_i^{(k)}$ using Pearson correlation:

$$F_{xy}^{(k)} = \text{Corr}(\mathbf{X}_i^{(k)}[x], \mathbf{X}_i^{(k)}[y]),$$

where $\mathbf{X}_i^{(k)}[x]$ represents the concatenated BOLD time-series for ROI x in state k. If $\mathcal{S}_{i,k}$ is empty ($T_{i,k} = 0$), $\mathbf{F}_i^{(k)}$ is replaced by the brain state FC matrix for the specific state k, which is the FC matrix reconstructed from the brain state vector $\mathbf{c}_k$.
- **Ordered State FC Representation:** To ensure consistency across samples, the resulting State FC matrices $\{\mathbf{F}_i^{(k)}\}_{k=1}^{K}$ are ordered based on state indices k, forming a structured representation:

$$[\mathbf{F}_i^{(1)}, \mathbf{F}_i^{(2)}, \ldots, \mathbf{F}_i^{(K)}].$$

This process aligns the State FC matrices across samples, enabling the model to better capture population-level patterns and inter-state relationships while preserving consistency across individual samples.

E. Feature Extraction with the BNT

After rechunking, the State FC matrices $\{\mathbf{F}_i^{(k)}\}_{k=1}^{K}$ for each sample i are passed through the BNT [19].

For each State FC matrix $\mathbf{F}_i^{(k)}$, the BNT generates a corresponding embedding $\mathbf{h}_i^{(k)}$ for state k:

$$\mathbf{h}_i^{(k)} = \text{BNT}(\mathbf{F}_i^{(k)}),$$

where it captures state-level patterns in the connectivity. The embeddings $\{\mathbf{h}_i^{(k)}\}_{k=1}^{K}$, representing all states for sample i, are concatenated to form a comprehensive representation:

$$\mathbf{H}_i = \text{Concat}(\mathbf{h}_i^{(1)}, \mathbf{h}_i^{(2)}, \ldots, \mathbf{h}_i^{(K)}).$$

The final concatenated representation $\mathbf{H}_i$ is then input to a MLP for downstream prediction tasks, such as gender classification or cognitive score regression.

By leveraging embeddings from all state-specific FC matrices, this process ensures that the model effectively captures both the distinct connectivity patterns of individual brain states and their inter-state relationships.

III. Experiments

A. Datasets

We utilize two publicly available fMRI datasets: The first is the ABCD [20] dataset, which provides fully anonymized brain networks constructed using the HCP 360 ROI atlas [23]. To ensure consistency across subjects, we truncated all time series to 512 time points. Our analysis focuses on two tasks: (1) a binary classification task for predicting gender, including 7,901 subjects after quality control (50.1% female), and (2) a regression task for predicting the Cognition Summary Score.

The second dataset is the HCP [21], which provides resting-state fMRI data processed using the CONN toolbox [24]. We use 982 samples with 132 ROI defined by the Harvard-Oxford atlas [25]. The BOLD time series are bandpass filtered (0.01–0.1 Hz) and truncated to 2400 time points. Our analysis focuses on two tasks: (1) a binary classification task for predicting gender, and (2) a regression task for predicting fluid intelligence scores with a mean of 17.03 and standard deviation of 4.70.

We divide our datasets such that 70% is utilized for training, 10% for validation, and the remainder for testing.

B. Experimental Protocols

Baselines. We evaluate the performance of our proposed BNST framework against several state-of-the-art models in brain network analysis, as summarized in Table I. To ensure a comprehensive comparison, the selected baselines include methods designed for both static and dynamic brain networks.

For static FC methods, we include BrainGB [1] and BNT [19]. BrainGB is specifically designed to work with static brain networks. BNT, a transformer-based model, projects static brain networks into embedding spaces and serves as a benchmark for transformer architectures, which is also utilized in our framework.

For DFC methods, we benchmark against STGCN [26] and STAGIN [27]. STGCN models spatiotemporal graph structures, incorporating not only current but also past and future connectivity patterns in its analysis. STAGIN builds on this approach by incorporating an attention mechanism, enabling it to effectively capture and integrate temporal dynamics across sequences of dynamic brain networks without relying on static representations.

Metrics. To assess the performance of the proposed model, we employ standard evaluation metrics for both classification and regression tasks. For classification, we use Accuracy (ACC), Area Under the Curve (AUC), and Sensitivity (SEN), which provide a comprehensive view of model performance across various thresholds. For regression, we use Mean Squared Error (MSE) to measure prediction accuracy and the Pearson Correlation Coefficient (PC) to quantify the linear correlation between predicted and actual values. Higher values for ACC, AUC, SEN, and PC, alongside lower values for MSE, reflect superior performance.

C. Implementation Details

For the clustering process, we configure the window size and stride according to the properties of the datasets. Specifically, for the ABCD dataset, we use a window size of 24 with a stride of 24 to ensure each window encapsulates a one-minute BOLD signal, while for the HCP dataset, we use a window size of 96 with a stride of 96. From each dataset, we randomly select 500 samples to perform clustering, balancing computational efficiency with representativeness [16]. Using a deep clustering framework, we identify 4 clusters for the ABCD dataset and 6 clusters for the HCP dataset, with the centroids representing brain state patterns. The difference in the number of clusters reflects the temporal variability and length of the datasets: the shorter BOLD time series in the ABCD dataset offers fewer distinct connectivity patterns, while the longer HCP series captures greater variability, justifying more clusters [17], [18].

BNT consisted of 2 layers, with the hidden dimension matching the number of nodes v in the brain network and 4 attention heads per layer. The model was optimized using the Adam optimizer with a learning rate of 10^{-4} and a weight decay of 10^{-4}. We trained the model for 100 epochs with a batch size of 16. The epoch achieving the best performance on the validation set was selected for the final evaluation. To ensure reliability and reproducibility, all results are averaged over 5 independent runs with different random seeds.

D. Results and Analysis

The overall performance of the proposed BNST model is summarized in Table I. We can observe that BNST achieves the best results compared to all baseline methods across both classification and regression tasks. By leveraging state-based FC, BNST demonstrates superior performance, achieving significant improvements in accuracy, sensitivity, and other key metrics across both datasets.

The results highlight the advantages of State FC in capturing meaningful and discriminative patterns in brain connectivity. Unlike static methods, which fail to incorporate temporal variations, and purely dynamic methods, which lack state-specific organization, BNST effectively encodes

TABLE I

COMPARISON OF PERFORMANCE METRICS FOR CLASSIFICATION AND REGRESSION TASKS ACROSS ABCD AND HCP DATASETS.

Method	ABCD					HCP				
	Classification			Regression		Classification			Regression	
	ACC ↑	AUC ↑	SEN ↑	MSE ↓	PC ↑	ACC ↑	AUC ↑	SEN ↑	MSE ↓	PC ↑
STAGIN	58.3 ± 2.4	61.5 ± 4.0	70.7 ± 2.2	90.5 ± 8.1	7.6 ± 3.0	55.6 ± 2.8	58.5 ± 4.3	68.1 ± 2.6	65.2 ± 6.6	7.1 ± 2.9
STGCN	60.4 ± 4.3	63.1 ± 4.9	74.1 ± 3.4	86.9 ± 7.6	9.7 ± 2.6	58.3 ± 4.1	60.8 ± 5.2	71.3 ± 3.8	61.1 ± 5.5	7.8 ± 3.2
BrainGB	79.7 ± 1.5	91.1 ± 0.3	79.9 ± 2.1	78.2 ± 4.4	34.3 ± 1.5	76.9 ± 1.8	88.7 ± 1.2	75.6 ± 2.4	45.2 ± 4.7	13.5 ± 3.9
BNT	86.2 ± 1.1	93.5 ± 0.5	86.5 ± 1.8	63.5 ± 0.8	44.1 ± 1.6	**85.0 ± 2.4**	93.2 ± 1.5	79.5 ± 2.1	28.0 ± 1.5	24.8 ± 6.1
BNST	**87.6 ± 0.6**	**93.9 ± 0.2**	**88.4 ± 1.3**	**61.6 ± 1.4**	**45.4 ± 0.8**	84.8 ± 3.3	**93.8 ± 1.5**	**82.1 ± 3.0**	**23.3 ± 1.1**	**27.5 ± 3.4**

both intra-state and inter-state dynamics. This enhanced representation enables the model to achieve more precise predictions across diverse tasks.

Additionally, the integration of State FC ensures a structured and consistent representation of brain connectivity across samples, further enhancing the robustness of the model. These findings validate the effectiveness of State FC and the proposed BNST framework in advancing brain network analysis.

E. Ablation Studies

We conducted a series of ablation studies to demonstrate the effectiveness of the key components in the BNST framework. The results, summarized in Tables II, III, and IV, provide a holistic validation of the proposed methodologies.

TABLE II

PERFORMANCE COMPARISON OF DYNAMIC FC AND STATE FC ACROSS DIFFERENT MODELS ON THE ABCD DATASET.

Base Model	Method	MSE ↓	PC ↑
BNST	DFC	64.8 ± 1.7	43.2 ± 1.1
	State FC	**61.6 ± 1.4**	**45.4 ± 0.8**
STGCN	DFC	86.9 ± 7.6	9.7 ± 2.6
	State FC	**84.6 ± 6.8**	**11.1 ± 2.1**
STAGIN	DFC	90.5 ± 8.1	7.6 ± 3.0
	State FC	**88.9 ± 7.5**	**8.8 ± 2.6**

TABLE III

COMPARISON OF K-MEANS AND DEEP K-MEANS CLUSTERING ACROSS MODELS ON THE ABCD DATASET.

Base Model	Method	MSE ↓	PC ↑
BNST	Kmeans	64.4 ± 3.4	41.1 ± 1.6
	Deep Kmeans	**61.6 ± 1.4**	**45.4 ± 0.8**
STGCN	Kmeans	86.1 ± 9.3	8.9 ± 3.3
	Deep Kmeans	**84.6 ± 6.8**	**11.1 ± 2.1**
STAGIN	Kmeans	92.3 ± 10.8	7.2 ± 3.9
	Deep Kmeans	**88.9 ± 7.5**	**8.8 ± 2.6**

First, to evaluate the effectiveness of State FC, we conducted a comparative study where both State FC and DFC were used as input features for the same set of models. Here, the models that directly utilize the raw sequence of DFC matrices generated through the sliding window method serve as the baseline. To compare, we also applied State FC as input features to these models and compared their performance. As shown in Table II, models using State FC

TABLE IV

PERFORMANCE COMPARISON OF AVERAGING AND STATE-GUIDED RECHUNKING ACROSS MODELS ON THE ABCD DATASET.

Base Model	Method	MSE ↓	PC ↑
BNST	Averaging	66.5 ± 2.8	39.8 ± 2.5
	Rechunking	**61.6 ± 1.4**	**45.4 ± 0.8**
STGCN	Averaging	91.4 ± 7.9	8.2 ± 3.6
	Rechunking	**84.6 ± 6.8**	**11.1 ± 2.1**
STAGIN	Averaging	96.0 ± 8.2	4.7 ± 4.2
	Rechunking	**88.9 ± 7.5**	**8.8 ± 2.6**

consistently outperform those using DFC. By focusing on state-specific connectivity patterns, State FC provides a more interpretable and structured representation of brain dynamics, enabling models like BNST to better capture meaningful features of brain activity and achieve superior predictive performance.

Next, a critical step in BNST is identifying meaningful brain states through clustering. By comparing the performance of several models utilizing State FC as input features, Table III shows that Deep K-Means significantly outperforms traditional K-Means in generating state representations. This is primarily because Deep K-Means incorporates an autoencoder for dimensionality reduction, effectively addressing the challenges of high-dimensional DFC data. The autoencoder compresses the high-dimensional input into a compact latent space while preserving salient features, ensuring that noise and irrelevant variations are minimized. This enhanced representation facilitates more robust clustering by enabling the subsequent k-means step to yield high-quality brain state representation, enabling BNST to accurately capture structured variations in brain activity.

Finally, State-guided rechunking is also essential for structuring BOLD time series data in a way that enhances state-specific representation. Table IV compares rechunking with a simpler averaging approach, where DFC matrices of the same state are averaged into a single representative matrix. While averaging simplifies the computational process, it fails to retain the distinct connectivity patterns within states. Rechunking addresses this limitation by reordering and concatenating BOLD time series segments assigned to the same state, ensuring that state-specific connectivity patterns are fully preserved. This allows BNST to model intra-state and inter-state relationships with greater precision, leading to substantial performance improvements.

Together, these components—robust state identification with Deep K-Means, structured representation through State

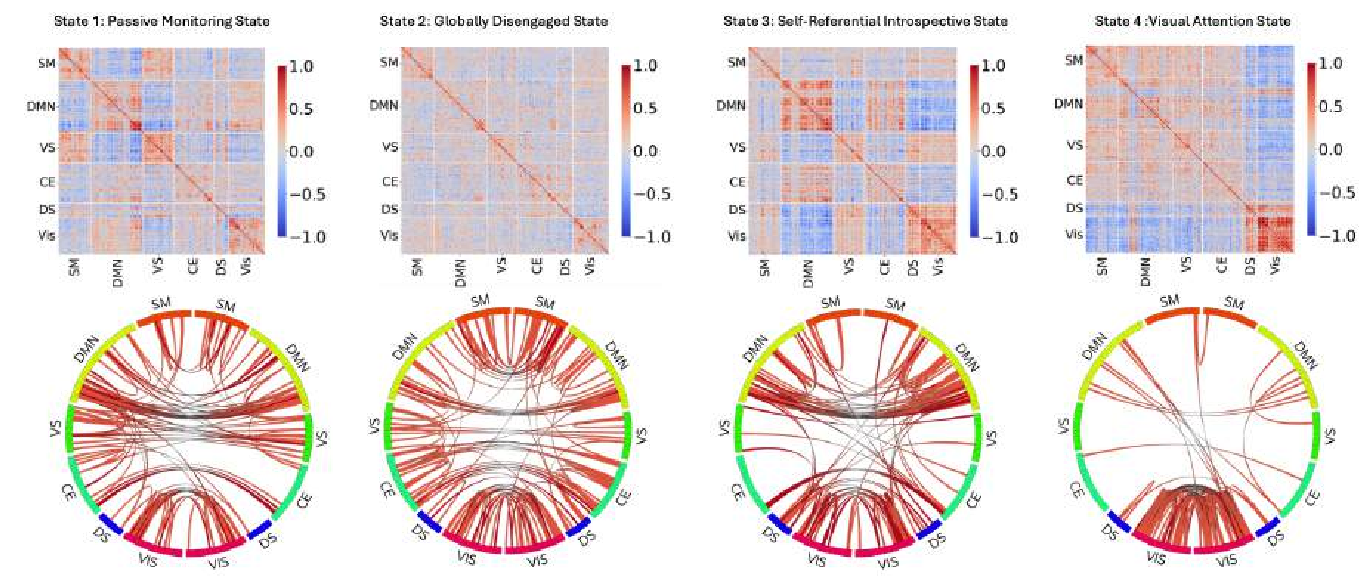

Fig. 4. Visualization of state-based functional connectivity patterns. The top row shows the four distinct brain states from ABCD dataset. The bottom row presents chord plots constructed from the top 0.2% of connections in terms of absolute value, highlighting the strongest functional interactions between major brain networks with the color intensity reflects connection strength.

FC, and rechunking to enhance state-specific organization—explain why BNST outperforms other approaches in modeling dynamic brain networks.

F. Case Studies

To qualitatively assess the effectiveness of the proposed BNST model and highlight its unique state-based interpretability, we present and analyze four distinct brain states identified through the clustering stage using the ABCD dataset in Figure 4: *Passive Monitoring State*, *Globally Disengaged State*, *Self-Referential Introspective State*, and *Visual Attention State*. These states represent unique DFC patterns that the BNST leverages for improved brain network analysis.

a) Passive Monitoring State: This state is characterized by connectivity across multiple brain networks, particularly moderate interactions between the somatomotor (SM), default mode network (DMN), and visual (Vis) networks, which together suggest a state of general environmental awareness and low-level sensory processing [28], [29]. The balanced connectivity pattern in this state may reflect an optimal baseline mode, allowing for efficient shifts between internally and externally directed cognitive processes. Smooth transitions to and from this state are indicative of effective neural resource allocation, further linking this state to the prediction of broader cognitive and behavioral traits.

b) Globally Disengaged State : This state reflects a transient period of reduced functional connectivity across large-scale brain networks, suggesting a disengagement from both externally driven and internally focused cognitive processes [28], [29]. Unlike other states that exhibit strong intra-network organization or task-relevant connectivity, this state is marked by a widespread weakening of network interactions, particularly within the DMN and cognitive executive (CE) networks. The absence of dominant connectivity suggests a transitional or low-engagement phase rather than a stable resting state. Rather than serving as a classical energy-conserving mode, this state may indicate a temporary lapse in coordinated neural processing or a resetting phase between more functionally active brain states. The duration and frequency of this state may provide insights into cognitive flexibility, with individuals who transition out of this state more quickly potentially exhibiting greater neural responsiveness and adaptive processing efficiency.

c) Self-Referential Introspective State: This state is characterized by heightened connectivity within the DMN, particularly in its interactions with the CE network. This pattern reflects engagement in self-referential thinking, memory retrieval, and introspective cognitive processes. This state shows reduced cross-network connectivity with sensory and motor systems, suggesting a shift away from external stimuli toward internally focused cognition. The stability and strength of DMN connectivity have been strongly linked to individual differences in higher-order cognitive functions, such as reasoning ability and abstract thinking [30]. Individuals with more pronounced activation of this state are likely to demonstrate enhanced internal cognitive efficiency, making it a critical feature for analyzing cognitive variability.

d) Visual Attention State: This state exhibits pronounced connectivity within the visual cortex and frontoparietal networks, particularly strengthening interactions between the Vis and dorsal stream (DS) networks. The enhanced intra-network connectivity within visual processing regions suggests heightened attentional engagement, while increased frontoparietal connectivity reflects the integration of visuospatial information with executive control processes. Variations in the occurrence or intensity of this state are linked to differences in attentional control and visuospatial processing capabilities, which are known to exhibit both individual and group-level differences [30]. Strong connectivity in this state may reflect an individual's ability to sustain focus and efficiently process visual information, providing insights into cognitive engagement and task performance.

The connectivity patterns and transition dynamics of these states provide a structured and interpretable representation of brain activity that enhances the predictive power of BNST. Unlike static or unstructured dynamic FC methods, the state-based segmentation isolates meaningful modes of the brain that are directly tied to cognitive and behavioral traits. By leveraging these state-specific features, BNST aligns neural dynamics with measurable outcomes, offering both superior predictive accuracy and neuroscientific interpretability.

IV. Conclusion

Our work introduced BNST, a novel framework leveraging State FC for enhanced brain network analysis. By integrating deep clustering for state identification, state-based rechunking of BOLD time series, and an advanced transformer architecture, BNST achieved superior performance across multiple tasks and datasets compared to baseline methods. The proposed use of brain states and State FC as a guiding principle in dynamic brain network analysis not only improved prediction accuracy but also provided a structured and interpretable representation of brain dynamics.

V. Acknowledgements

This research was partially supported by the US National Science Foundation under Award Number 2319449 and Award Number 2312502.

References

[1] 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," *arXiv preprint arXiv:2206.03645*, 2022.

[2] X. Li, Y. Zhou, N. C. 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*, 2021.

[3] F. V. Farahani, W. Karwowski, and N. R. Lighthall, "Application of graph theory for identifying connectivity patterns in human brain networks: a systematic review," *Frontiers in Neuroscience*, 2019.

[4] S. L. Simpson, F. D. Bowman, and P. J. Laurienti, "Analyzing complex functional brain networks: fusing statistics and network science to understand the brain," *Statistics Surveys*, 2013.

[5] L. Novelli and A. Razi, "A mathematical perspective on edge-centric brain functional connectivity," *Nature Communications*, 2022.

[6] 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*, 2017.

[7] Y. Zhu, H. Cui, L. He, L. Sun, and C. Yang, "Joint embedding of structural and functional brain networks with graph neural networks for mental illness diagnosis," in *Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2022.

[8] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, "Network modelling methods for fmri," *NeuroImage*, 2011.

[9] A. Li, Y. Yang, H. Cui, and C. Yang, "Brainsteam: A practical pipeline for connectome-based fmri analysis towards subject classification," in *Proceedings of a conference*, 2022.

[10] X. Kan, A. A. Chen Gu, H. Cui, Y. Guo, and C. Yang, "Dynamic brain transformer with multi-level attention for functional brain network analysis," in *Proceedings of a conference*, 2022.

[11] U. Pervaiz, D. Vidaurre, C. Gohil, S. Smith, and M. Woolrich, "Multi-dynamic modelling reveals strongly time-varying resting fmri correlations," *Medical Image Analysis*, vol. 77, pp. 102–116, 2022.

[12] L. Liu, G. Wen, P. Cao, T. Hong, J. Yang, X. Zhang, and O. R. Zaiane, "Braintgl: A dynamic graph representation learning model for brain network analysis," *Computers in Biology and Medicine*, 2023.

[13] A. D. Savva, G. D. Mitsis, and G. K. Matsopoulos, "Assessment of dynamic functional connectivity in resting-state fmri using the sliding window technique," *Brain and Behavior*, 2019.

[14] S. M. Smith, "The future of fmri connectivity," *NeuroImage*, vol. 62, pp. 1257–1266, 2012.

[15] T. O. Laumann, A. Z. Snyder, and C. Gratton, "Challenges in the measurement and interpretation of dynamic functional connectivity," *NeuroImage*, 2017.

[16] V. M. Vergara, M. Salman, A. Abrol, F. A. Espinoza, and V. D. Calhoun, "Determining the number of states in dynamic functional connectivity using cluster validity indexes," *Journal of Neuroscience Methods*, 2021.

[17] E. Damaraju, E. A. Allen, A. Belger, J. M. Ford, S. McEwen, D. H. Mathalon, B. A. Mueller, G. D. Pearlson, S. G. Potkin, A. Preda, J. A. Turner, J. G. Vaidya, T. G. van Erp, and V. D. Calhoun, "Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia," *NeuroImage: Clinical*, 2014.

[18] A. P. Spencer and M. Goodfellow, "Using deep clustering to improve fmri dynamic functional connectivity analysis," *NeuroImage*, vol. 257, p. 119288, 2022.

[19] X. Kan, W. Dai, H. Cui, Z. Zhang, Y. Guo, and C. Yang, "Brain network transformer," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

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

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

[22] C. F. J. Bond and K. Richardson, "Seeing the fisher z-transformation," *Psychometrika*, vol. 69, no. 2, pp. 291–303, 2004.

[23] 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, D. C. V. Essen, and M. Jenkinson, "The minimal preprocessing pipelines for the human connectome project," *NeuroImage*, vol. 80, pp. 105–124, 2013.

[24] S. Whitfield-Gabrieli and A. Nieto-Castanon, "Conn: A functional connectivity toolbox for correlated and anticorrelated brain networks," *Brain Connectivity*, vol. 2, no. 3, pp. 125–141, 2012.

[25] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman *et al.*, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *NeuroImage*, vol. 31, no. 3, pp. 968–980, 2006.

[26] S. Gadgil, Q. Zhao, E. H. Abdelnabi, A. Taleb, H. Lombaert, and M. Toews, "Spatio-temporal graph convolution for resting-state fmri analysis," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2020, pp. 528–538.

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

[28] B. T. 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, B. Fischl, H. Liu, and R. L. Buckner, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, vol. 106, no. 3, pp. 1125–1165, 2011.

[29] J. R. Andrews-Hanna, J. Smallwood, and R. N. Spreng, "The default network and self-generated thought: Component processes, dynamic control, and clinical relevance," *Annals of the New York Academy of Sciences*, vol. 1316, no. 1, pp. 29–52, 2014.

[30] S. Vossel, J. J. Geng, and G. R. Fink, "Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles," *The Neuroscientist*, vol. 20, no. 2, pp. 150–159, 2014.