Mixing Temporal Graphs with MLP for Longitudinal Brain Connectome Analysis

Hyuna Cho[1], Guorong Wu[2], and Won Hwa Kim[1(✉)]

[1] Pohang University of Science and Technology (POSTECH), Pohang, South Korea
{hyunacho,wonhwa}@postech.ac.kr
[2] University of North Carolina, Chapel Hill, USA

Abstract. Analyses of longitudinal brain networks, i.e., graphs, are of significant interest to understand the dynamics of brain changes with respect to aging and neurodegenerative diseases. However, each subject has a graph of heterogeneous structure and time-points as the data are obtained over several years. Moreover, most existing datasets suffer from lack of samples as the images are expensive to acquire, which leads to overfitting with complex deep neural networks. To address these issues for characterizing progressively alternations of brain connectome and region-wise measures as early as possible, we develop Spatio-Temporal Graph Multi-Layer Perceptron (STGMLP) that mixes features over both graph and time spaces to classify sets of longitudinal human brain connectomes. The proposed model is made efficient and interpretable such that it can be easily adopted to medical imaging datasets and identify personalized features responsible for a specific diagnostic label. Extensive experiments show that our method achieves successful results in both performance and computational efficiency on Alzheimer's Disease Neuroimaging Initiative (ADNI) and Adolescence Brain Cognitive Development (ABCD) datasets independently.

1 Introduction

Consider a longitudinal brain connectome study where each participant goes through imaging protocol multiple times over the study period. Given a population of such subjects, analyzing them can be posed as a *spatio-temporal graph analysis* where each sample in a dataset is given as a set of longitudinal graphs of different cardinality. An exemplar sample corresponding to this task is shown in Fig. 1a that consists of graphs from T_m time points with multi-variate node features (denoted in different colors). The fundamental goal of such longitudinal studies is to characterize progressive change patterns of time-varying graphs due to certain factors such as aging [32] and neurodegenerative diseases [19, 22].

There are several practical bottlenecks to extract meaningful results from the longitudinal brain connectome and region-wise imaging measures. The data are temporally sparse, i.e., the participants pay a different number of visits which can be very few. Also, each brain network has a different structure of white-matter fiber connections unlike regular lattice structure in images. Last but importantly,
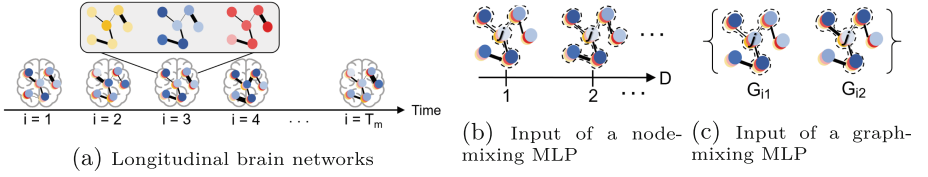
(a) Longitudinal brain networks

(b) Input of a node-mixing MLP

(c) Input of a graph-mixing MLP

Fig. 1. (a) A set of longitudinal graphs for T_m timepoints. The node colors (i.e., yellow, blue, and red) represent different node features, and the edge thickness stands for edge strength. (b) A node-mixing MLP in GSMs takes node features of a local graph centered at node j and its neighbors (in dotted lines) as an input. As the depth D of GSM increases, the range of the local graph is broadened. (c) A graph-mixing MLP in GTMs takes a graph pair p of different time-points (i.e., $p = \{i_1, i_2\}$). (Color figure online)

most neuroimaging datasets suffer from lack of samples as the data are expensive to acquire and process. These spatio-temporal heterogeneities and sample-size issue make longitudinal analyses of the brain network challenging, but it must be investigated to characterize the progressive disease-relevant variations.

Therefore, it is necessary to develop an efficient prediction model for a "set of longitudinal graphs" and corresponding regional measures (i.e., node features) over sparse time-points. Most graph neural networks are designed for a fixed template graph for predicting node values [18] or graph-level labels [33,35], where the graph topology is used as a domain and predictions heavily rely on node features. Moreover, recent spatio-temporal graph methods for a stream of images (e.g., video) often include complex architectures that require a large-scale dataset to train [2,3], which cannot be easily adopted for medical applications due to the limited sample-size. Notice that the sample-size is a much bigger issue for a longitudinal study as a single label is given for a "set" of graphs, as opposed to a cross-sectional study where the label is given for each graph.

We tackle the aforementioned issues by designing a flexible architecture with a "mix" of Multi-layer Perceptron (MLP) [30] to investigate time-varying graph structure and measurements on the nodes. We propose Spatio-Temporal Graph MLP (STGMLP) which integrates the following three mixing modules: 1) graph spatial mixer (GSM), 2) graph temporal mixer (GTM), and 3) spatio-temporal mixer (STM) that extract space, time, and spatio-temporal features, respectively. The features curated from the three components are fed into a downstream classifier to discriminate labels for the sets of longitudinal graphs. The core idea is to efficiently mix features along irregular space and time with simple MLPs: brain network structure guides spatial mixing as a graph, and temporal pooling extracts the most effective features from disjoint time-space across subjects.

Contributions of Our Work: our model 1) can be trained efficiently compared to existing spatio-temporal graph deep models with a significantly reduced number of parameters, 2) flexibly incorporates irregular space and time into prediction, 3) yields interpretable results that quantify the contribution of each node
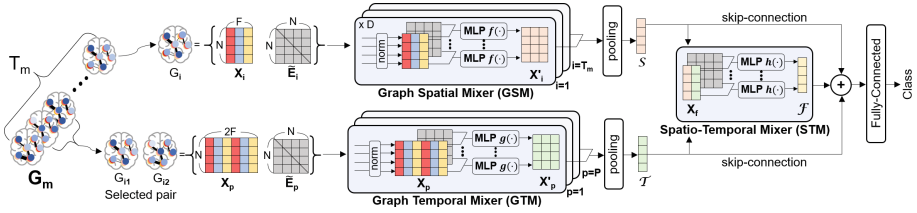
Fig. 2. Architecture of STGMLP. It consists of T_m Graph Spatial Mixers (GSM), P Graph Temporal Mixers (GTM), a Spatio-Temporal Mixer (STM), and a classifier head. GSMs and GTMs contain a node-mixing MLP and a graph-mixing MLP, respectively. STM mixes the jointly obtained features from each spatial and temporal aspects. If the GSM depth D is $D>1$, the GSM output X_i' is used as an input of the same GSM to deepen model layers and to widen an encoding range of neighbor nodes.

to classify different clinical labels. Extensive validation was performed on two independent public datasets, i.e., Alzheimer's Disease Neuroimaging Initiative (ADNI) and Adolescent Brain Cognitive Development (ABCD), for classifying pre-defined groups to demonstrate the efficiency and efficacy of our model.

2 STGMLP: Spatio-Temporal Graph MLP

Problem Definition. Consider a longitudinal graph set $\boldsymbol{G}_m = \{G_1, ..., G_{T_m}\}$ for $m = 1, 2, \cdots, M$ samples where the graphs for $i = 1, ..., T_m$ timepoints (different across samples, $T_m \geq 2$) are presented in chronological order. For each undirected graph $G_i = \{E_i, X_i\}$, $X_i \in \mathbb{R}^{N \times F}$ contains F node features for N nodes and $E_i \in \mathbb{R}^{N \times N}$ is a weighted adjacency matrix whose elements denote connection strength between two nodes. Given a population of $\boldsymbol{G}_m$ with C classes, STGMLP aims to classify the label of each $\boldsymbol{G}_m$ by leveraging both temporal and spatial variations of the graph set from different groups. Note that the label of each sample (i.e., longitudinal graph set) is consistent over time.

Overview of STGMLP. STGMLP mixes graph features across space and time with Graph Spatial Mixer (GSM) and Graph Temporal Mixer (GTM), respectively. GSM performs a per-graph operation (i.e., *node-mixing*) and GTM accounts for cross-temporal operation (i.e., *graph-mixing*) between multiple graphs in a $\boldsymbol{G}_m$. Figure 1b and 1c show inputs to node-mixing and graph-mixing MLPs. The node-mixing MLP projects node features from a local neighborhood of each node (i.e., local graph) onto a latent space. On the other hand, the graph-mixing MLP extracts hidden relationships between graphs across time by projecting spatially equivalent local graphs from multiple G_i's to the same latent space.

The overall structure of STGMLP is shown in Fig. 2, which integrates GSMs, GTMs, a Spatio-Temporal Mixer (STM), and a downstream classifier. Due to the heterogeneous number of timepoints T_m across samples, pooling operations are applied to the outputs of GSMs and GTMs to reduce them into a coherent dimension. Taking outcomes from the pooling layers, the STM fuses both spatial

and temporal features. The fused feature is combined with its inputs through a skip-connection [14] to maximize the use of multi-level (i.e., space, time, and spatio-temporal) information extracted from the input. Finally, a downstream classifier takes the mixed features to predict labels for a given longitudinal graph set $\boldsymbol{G}_m$. The details of each module and variables are given below.

Graph Spatial Mixer. GSM encodes node features and a graph structure of a graph G_i with graph convolution. The node-mixing MLP ($\mathbb{R}^F \mapsto \mathbb{R}^F$) acts on rows of X_i, and it is shared across $\boldsymbol{G}_m$ for all $N \times T_m$ nodes. Let $f(\cdot)$ be an operation of node-mixing MLP which takes $\tilde{E}_i$ and X_i as inputs, where $\tilde{E}_i$ is a normalized E_i to ensure unbiased strength of the connectivity. It includes self-connections I_N (i.e., identity matrix) and computed as $\tilde{E}_i = \tilde{D}_i^{-\frac{1}{2}}(E_i + I_N)\tilde{D}_i^{-\frac{1}{2}}$, where $(\tilde{D}_i)_{jj} = \sum_{k=1}^{N}(E_i + I_N)_{jk}$ is a diagonal degree matrix of $E_i + I_N$ for $j = 1, ..., N$ nodes. Layer normalization [1] is applied across all features to prevent biased learning from unbalanced node feature distributions. Including two fully-connected (FC) layers and a GELU nonlinearity [15] $\sigma(\cdot)$, the MLP operates independently on each j-th node with two-layer graph convolutions [18] as

$$(X_i')_j = f(\tilde{E}_i, X_i) = \tilde{E}_i\,\sigma(\tilde{E}_i\,\mathrm{LN}(X_i)_j W^0)W^1, \tag{1}$$

where W^0 and W^1 are trainable weights and $\mathrm{LN}(\cdot)$ is a layernorm function. The output $(X_i')_j \in \mathbb{R}^F$ accounts for a latent vector of local graph structure at node j. Stacking $(X_i')_j$s up to the number of nodes N, an outcome $X_i' \in \mathbb{R}^{N \times F}$ is derived for an input G_i. In this way, a set of whole outputs from T_m GSMs is derived as $\{X_i'\}_{i=1}^{T_m} \in \mathbb{R}^{T_m \times N \times F}$. Notice that the GSM can be stacked D times by iteratively taking the X_i' as an updated input to encode a wider range of local graph structures. After performing max pooling on T_m and F dimensions of $\{X_i'\}_{i=1}^{T_m}$, the condensation of spatial features across $\boldsymbol{G}_m = \{G_1, ..., G_{T_m}\}$ is obtained as a N-dimensional vector $\mathcal{S}$.

Graph Temporal Mixer. GTM performs a cross-temporal operation on multiple "pairs" of graphs. This graph-mixing encodes the relations between graphs of different time-points. Given T_m graphs from a subject, P pairs of graphs, each pair as a set $\{G_{i_1}, G_{i_2}\}$, are selected where P is a user parameter. For each pair for $\{i_1, i_2\}$, an averaged connectivity $\tilde{E}_p = (\tilde{E}_{i_1} + \tilde{E}_{i_2})/2$ and $X_p \in \mathbb{R}^{N \times 2F}$ as a concatenation of X_{i_1} and X_{i_2} are inputted into the graph-mixing MLP $g(\cdot)$. In our work, we choose to input pairs of temporally adjacent graphs together with the first-and-last graph pair to encode a temporal sequence. The $g(\cdot)$ acts on rows of X_p, mapping $\mathbb{R}^{2F} \mapsto \mathbb{R}^F$. It transforms the features of node j (i.e., $(X_p)_j$) into F-dimensional latent vector, and the projection is performed across the whole node pairs in parallel. Similar to the node-mixing MLP, graph-mixing MLP contains two FC layers with weights W^2 and W^3 and a GELU $\sigma(\cdot)$ as

$$(X_p')_j = g(\tilde{E}_p, X_p) = \sigma(\tilde{E}_p\,\mathrm{LN}(X_p)_j W^2)W^3. \tag{2}$$

As in the GSM, each $(X_p')_j$ is stacked N times to be a X_p'. For P GTMs, an output $\{X_p'\}_{p=1}^{P} \in \mathbb{R}^{P \times N \times F}$ is obtained and reduced into $\mathcal{T} \in \mathbb{R}^N$ by max pooling on P pairs and F node features. Note that, unlike GSMs, $\tilde{E}_p$ is used

only once in Eq. (2). Using $\tilde{E}_p$ multiple times causes encoding of a wider range of local graph structures, unnecessarily encompassing graph spatial relations (i.e., non-adjacent neighbors) for extracting temporal relations of j-th node pairs.

Spatio-Temporal Mixer. To capture comprehensive spatio-temporal relations across the whole graphs in $\boldsymbol{G}_m$, the spatial and temporal features, i.e., $\mathcal{S}$ and $\mathcal{T}$, are embedded into a latent vector $\mathcal{F} \in \mathbb{R}^N$ in STM. To do so, the $\mathcal{S}$ and $\mathcal{T}$ are stacked as $X_f = [\mathcal{S}, \mathcal{T}]$, where $(X_f)_j$ represents the spatial and temporal node features of the j-th node. A spatio-temporal mixing MLP $h(\cdot)$ is applied to $(X_f)_j$ for all j's in parallel with an averaged edge matrix $\tilde{E}_f$ across $\boldsymbol{G}_m$ as

$$\mathcal{F}_j = h(\tilde{E}_f, X_f) = \tilde{E}_f \ \sigma\big(\tilde{E}_f(X_f)_j W^4\big)W^5, \tag{3}$$

where W^4 and W^5 are weight matrices and $\sigma(\cdot)$ is a nonlinearity. With this STM, irregular space and time components can be flexibly integrated into a prediction.

Longitudinal Graph Classifier. To take a full advantage from the extracted features, $\mathcal{S}$ and $\mathcal{T}$ are combined together with $\mathcal{F}$ via a skip connection. These features collected from diverse branching paths contain both low and high-level information extracted from the graphs, and their integration provides strong ensemble-like results [31]. Using a FC layer and softmax, a set of predicted label probabilities $\hat{Y}$ is obtained for C classes as

$$\hat{Y}_c = \frac{\big((\mathcal{F} + \mathcal{S} + \mathcal{T})W^6\big)_c}{\sum_{c=1}^{C}\big((\mathcal{F} + \mathcal{S} + \mathcal{T})W^6\big)_c} \qquad \text{for } c = 1, \dots C, \tag{4}$$

where W^6 is a set of trainable weights of the FC layer for class prediction. Given the ground truth Y, the cross-entropy loss is defined with ℓ_2-regularization as

$$L = -\frac{1}{M}\sum_{m=1}^{M}\sum_{c=1}^{C} Y_{mc} \cdot \log(\hat{Y}_{mc}) + \frac{\lambda}{2M}||\mathbf{W}||_{\ell_2}^2, \tag{5}$$

where $\mathbf{W}$ is a set of trainable parameters and λ controls a regularization strength.

3   Experiments

In this section, we evaluate STGMLP on two independent datasets, i.e., ADNI and ABCD, whose demographics are given in Table 1. We discuss the quantitative results, model behavior, and neuroscientific interpretations below.

3.1   Materials and Setup

ADNI Dataset. The ADNI is the largest public AD dataset providing longitudinal and multimodal images such as magnetic resonance imaging (MRI) and positron emission tomography (PET). As node features, cortical thickness from MRI, standardized uptake value ratio (SUVR) from FDG-PET and Amyloid-PET at all ROIs were measured. Structural brain networks were obtained by in-house probabilistic tractography on diffusion tensor images (DTI) on Destrieux

Table 1. Demographics of the ADNI and ABCD datasets.

Category	ADNI			ABCD	
	Preclinical	MCI	AD	BP	NP
# of Subject	45	61	25	835	734
# of Record	163	253	75	1,670	1,468
Gender (M/F)	18/27	38/23	14/11	439/396	399/335
Age (Mean $\pm$ std)	73.8 ± 5.6	72.5 ± 7.3	76.0 ± 7.5	9.9 ± 0.4	9.9 ± 0.4

atlas [8] with 148 ROIs. The number of visits (i.e., T_m) by participants varied from 2 to 7. Five labels were initially given: cognitively normal (CN), significant memory concern (SMC), early mild cognitive impairment (EMCI), late MCI (LMCI) and AD. They were redefined as Preclinical (CN and SMC), MCI (EMCI and LMCI) and AD groups to secure sufficient sample size.

ABCD Dataset. The ABCD dataset (v4.0) contains two timepoints with multivariate features: baseline data for children aged 8–10 and their 2-year follow-up measurements such as fractional anisotropy, mean diffusivity, and cortical thickness obtained via DTI and MRI. Morphometric similarity network [27] was used to construct a graph per subject. As in other works [5,9,24,34] studying the relationship between socioeconomic status (SES) and brain development on the ABCD, we categorized the longitudinal samples into Below-Poverty (BP) and Non-Poverty (NP) groups based on the annual household income. The poverty criterion from U.S. Census Bureau ($27,479) is used to set the BP group, and the NP group is set whose annual household income is $200k and greater.

Setup. As baselines, we adopt various graph convolutional networks (GCNs) for spatio-temporal graph analysis such as ST-GCN [11], IT-GNN [16], infoGCN [4], ShiftGCN [3], and CTRGCN [2]. Also, typical machine learning (ML) methods such as Linear SVM, Linear Regression (LR) and MLP are used for comparisons. Along with the all node and edge features, the maximum time difference $X_{T_m} - X_1$ is used to train these ML models. We applied early stopping via test loss with 5-fold cross validation (CV) for all methods including ours.

To implement STGMLP, the learning rate, weight decay (λ), dropout rate, and depth D of GSM were set to $1e$-2, $5e$-4, 5%, and 3, respectively. For GTMs on the ADNI, total $P = T_m$ pairs are selected: $T_m - 1$ pairs for adjacent graphs in time, and one pair for the first and last (i.e., end-to-end) timepoints. For the ABCD, P is set to 1 as $T_m = 2$ for all samples. Note that, the combination of timepoints can be flexibly selected to include domain knowledge.

3.2 Evaluation and Discussions on the Results

Quantitative results (i.e., mean accuracy, precision, and recall) of all experiments and the number of trainable parameters are compared in Table 2. The results demonstrate that our model with a small computational cost outperformed baselines with vast parameters on both datasets. Also, our method showed no overfitting, as the mean training accuracies for ADNI and ABCD were 74.9% and

Table 2. Comparison of the number of trainable parameters and model performance with 5-fold cross validation from ADNI and ABCD experiments.

Method	# param	ADNI			ABCD		
		Accuracy (%)	Recall (%)	Precision (%)	Accuracy (%)	Recall (%)	Precision (%)
SVM	-	47.3 ± 4.2	42.9 ± 6.0	48.3 ± 3.2	64.4 ± 7.2	65.0 ± 5.7	70.5 ± 3.2
LR	68k	55.8 ± 1.9	45.4 ± 4.5	52.2 ± 14.0	67.8 ± 5.0	66.7 ± 5.3	68.7 ± 4.8
MLP	1,461k	55.7 ± 2.7	46.4 ± 6.6	50.8 ± 11.4	62.8 ± 6.6	61.2 ± 7.6	66.3 ± 5.1
ST-GCN [11]	62k	47.6 ± 2.2	46.0 ± 5.2	47.3 ± 1.7	54.1 ± 1.5	54.7 ± 6.1	46.2 ± 7.9
IT-GNN [16]	5k	50.4 ± 4.4	51.6 ± 5.0	50.8 ± 5.1	54.2 ± 1.5	51.4 ± 2.3	36.3 ± 6.4
ShiftGCN [3]	753k	58.9 ± 3.9	60.7 ± 6.8	57.2 ± 11.4	65.8 ± 3.3	66.6 ± 4.1	67.1 ± 3.5
CTRGCN [2]	683k	62.0 ± 2.9	57.7 ± 5.5	57.4 ± 14.8	67.4 ± 3.6	67.5 ± 3.7	67.9 ± 3.6
CTRGCN (w/ 1 layer) [2]	67k	63.5 ± 8.6	57.4 ± 11.0	55.8 ± 12.4	66.5 ± 3.3	66.6 ± 4.1	67.1 ± 3.5
infoGCN [4]	40,584k	57.3 ± 7.5	52.1 ± 10.1	43.2 ± 13.7	53.9 ± 2.9	51.1 ± 2.5	27.0 ± 0.9
infoGCN (w/ 1 layer) [4]	840k	58.9 ± 2.5	58.0 ± 5.4	51.2 ± 12.0	53.8 ± 1.3	51.5 ± 3.4	27.0 ± 0.9
STGMLP (Ours)	**1k**	**71.3 ± 4.6**	**66.4 ± 3.3**	**76.5 ± 6.1**	**72.3 ± 2.9**	**72.1 ± 2.9**	**72.4 ± 3.0**

Table 3. The nodal Grad-CAM for top 10 ROIs with the highest contribution to classify the AD class of ADNI dataset (Left) and the BP class of ABCD dataset (Right).

idx	ROI	Preclinical	MCI	AD	idx	ROI	BP	NP
1	r g.front.sup	9.0e-3	8.5e-2	**1.8e-1**	1	l g.insular.short	**2.8e-1**	2.5e-1
2	l s.temporal.sup	1.6e-1	8.6e-2	**1.7e-1**	2	r s.postcentral	**1.8e-1**	1.5e-1
3	r g.temp.sup.g.t.transv	1.4e-1	**1.9e-1**	1.7e-1	3	l s.oc.middle.and.Lunatus	**1.6e-1**	1.4e-1
4	r s.temporal.sup	7.5e-3	7.6e-2	**1.5e-1**	4	r g.occipital.sup	**1.6e-1**	1.2e-1
5	l g.temp.sup.lateral	6.0e-2	1.3e-1	**1.5e-1**	5	r s.oc.middle.and.Lunatus	**1.5e-1**	1.4e-1
6	r g.insular.short	2.0e-1	**2.1e-1**	1.5e-1	6	l g.occipital.sup	**1.4e-1**	1.3e-1
7	r g.temp.sup.plan.polar	1.9e-1	**2.0e-1**	1.5e-1	7	r g.insular.short	**1.4e-1**	1.2e-1
8	r g.temp.sup.lateral	5.9e-2	1.3e-1	**1.4e-1**	8	r. g.Ins.lg.and.s.cent.ins	**1.4e-1**	1.0e-1
9	l g.front.sup	1.1e-2	7.3e-2	**1.4e-1**	9	l g.temp.sup.Plan polar	1.4 e-1	**1.7e-1**
10	r g.orbital	3.8e-2	1.0e-1	**1.4e-1**	10	l s.postcentral	**1.4e-1**	1.3e-1

73.6%. Moreover, it is worth noting that the improvement in performance comes from the effectiveness of our method, not solely from the reduction in the number of parameters. On ADNI, our model even showed a 20.9%p accuracy margin over IT-GNN, which has a similar parameter scale (5k) as ours (1k).

Preclinical vs. MCI vs. AD on ADNI Dataset. Here, STGMLP achieved 71.3% accuracy with 7.8~24.0%p margin over baselines. Here, we provide clinically interpretable results by analyzing nodal contributions to classify each class via class-averaged Grad-CAM [28]. In Table 3, we reported the top 10 regions with the highest gradient activation for AD group classification, which are mostly distributed in *the temporal and frontal* regions. The ROI showing the highest activation is *the right superior frontal gyrus*, which is a majorly damaged area where atrophy of white matter is discovered in various AD studies [13,17,25]. Also, both sides of *the superior temporal gyri and sulci* were found, which are highly activated in auditory and verbal memory processing [20,36]. The visualization of averaged class-wise activations for all ROIs is shown in Fig 3.

Below-Poverty vs. Non-poverty on ABCD Dataset. As shown in Table 3, ROIs that played a decisive role to classify the BP class are mostly distributed
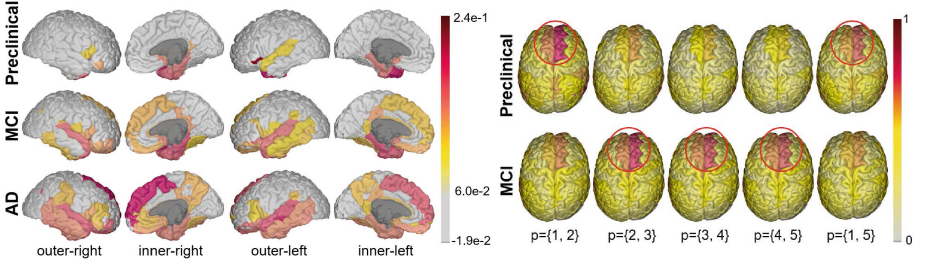
Fig. 3. (Left) Visualization of the nodal class-wise activations for Preclinical vs. MCI vs. AD group analysis. The activations are averaged on a per-class basis. Progressive variations in several ROIs are distinguishable for the labels. (Right) Averaged temporal feature maps extracted with GTM from subjects with $T_m=5$. For each timepoint pair $p=\{i_1, i_2\}$, higher values (circled in red) indicate higher activation from GELU.

along the *insular* and *occipital* regions. For example, insular subregions such as *the long insular gyrus and central sulcus of the insula* and both left and right sides of *the short insular gyri* were identified, which are implicated to social decision making [23, 26] and emotional processing [6, 12]. These ROIs responsible for somatosensory are thought to be impacted by environmental exposures such as SES and play a key role in overall cognition in children [7, 10]. Also, developmental changes in occipital subregions such as *the superior occipital gyrus* and *the middle occipital sulcus* are closely related to the parental SES [21, 29], which appear to be consistent with the result of our experiment.

3.3   Temporal Analysis on AD-Specific Activation

In Fig. 3, we also investigated pairwise temporal features from the ADNI experiment. While the Preclinical group shows strong activations in the initial and end-to-end pairs, features of later timepoints are highly activated in MCI group, showing consistency to the neurodegenerative dynamics in AD. We also observed that the use of pairwise information is sufficient to analyze the entire time series, as the signs of trained weights (i.e., W^2 in GTM) were totally opposite between adjacent timepoints, i.e., averaged weights were -7.7e-2 (std:0.56) vs. +4.4e-2 (std:0.29). In this way, our method investigates network alterations (around 1–2 year) in a pairwise manner and captures the whole temporal variation via pooling, rather than directly looking at the whole changes (over several years). These results confirm that GTM can capture the key temporal features for a given pair and label, and the following nonlinear function intensifies the difference.

3.4   Ablation Study on Hyperparamters

Ablation studies on each module (i.e., GSM, GTM, and STM) and pooling methods were performed on ADNI. The results, reported in Table 4, show that using max pooling for both GSM and GTM performed best. This suggests that there

Table 4. (Left) Ablation study on spatial ($\mathcal{S}$) and temporal ($\mathcal{T}$) features and their fusions ($\mathcal{F}$). (Right) Ablation study on pooling methods.

Fusion method	Accuracy (%)	F1-score (%)
$\mathcal{S}$	59.6 ± 6.9	50.6 ± 14.6
$\mathcal{T}$	67.4 ± 7.3	62.4 ± 7.0
$\mathcal{F}$	60.4 ± 7.4	51.2 ± 15.1
$\mathcal{S} + \mathcal{T}$	70.5 ± 6.4	$\mathbf{69.4} \pm 7.2$
$\mathcal{F} + \mathcal{S} + \mathcal{T}$	$\mathbf{71.3} \pm 4.6$	68.2 ± 2.9

Fusion Method	Accuracy (%)	F1-score (%)
GSM (avg) + GTM (avg)	65.8 ± 4.7	60.5 ± 5.9
GSM (avg) + GTM (max)	67.4 ± 7.1	65.7 ± 5.8
GSM (max) + GTM (avg)	65.0 ± 5.4	61.6 ± 7.1
GSM (max) + GTM (max)	$\mathbf{71.3} \pm 4.6$	$\mathbf{68.2} \pm 2.9$

exist particularly significant time-points (or time-points pairs) for classifying longitudinal brain networks and strongly reflecting these points in decision making is more useful than smoothing features for the entire time with average pooling.

4   Conclusion

In this work, we proposed a novel longitudinal graph mixer to investigate longitudinal variations of spatio-temporal graphs. The idea was driven by mixing features temporally and spatially along the topology of brain networks, and its structure was designed to deal with the heterogeneity of data with a significantly reduced number of parameters compared to deep graph convolutional models. Experiments validate the superiority of our framework, successfully identifying key ROIs in classifying different classes, suggesting a significant potential to be deployed for other longitudinal connectome analyses of various brain disorders.

References

1. Ba, J.L., Kiros, J.R., et al.: Layer normalization. arXiv preprint arXiv:1607.06450 (2016)
2. Chen, Y., Zhang, Z., et al.: Channel-wise topology refinement graph convolution for skeleton-based action recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 13359–13368 (2021)
3. Cheng, K., Zhang, Y., et al.: Skeleton-based action recognition with shift graph convolutional network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 183–192 (2020)
4. Chi, H.g., Ha, M.H., et al.: Infogcn: Representation learning for human skeleton-based action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20186–20196 (2022)

5. Cho, H., Park, G., Isaiah, A., Kim, W.H.: Covariate correcting networks for identifying associations between socioeconomic factors and brain outcomes in children. In: de Bruijne, M., et al. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2021: 24th International Conference, Strasbourg, France, September 27 – October 1, 2021, Proceedings, Part VII, pp. 421–431. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87234-2_40

6. Craig, A.D.: How do you feel-now? the anterior insula and human awareness. Nat. Rev. Neurosci. **10**(1), 59–70 (2009)

7. Craig, A.D., Chen, K., et al.: Thermosensory activation of insular cortex. Nat. Neurosci. **3**(2), 184–190 (2000)

8. Destrieux, C., Fischl, B., et al.: Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage **53**(1), 1–15 (2010)

9. Ellwood-Lowe, M., Irving, C., et al.: Exploring neural correlates of behavioral and academic resilience among children in poverty. Dev. Cogn. Neurosci. **54**, 101090 (2022)

10. Failla, M.D., Peters, B.R., et al.: Intrainsular connectivity and somatosensory responsiveness in young children with ASD. Molecular Autism **8**(1), 1–11 (2017)

11. Gadgil, S., Zhao, Q., Pfefferbaum, A., Sullivan, E.V., Adeli, E., Pohl, K.M.: Spatio-Temporal Graph Convolution for Resting-State fMRI Analysis. In: Martel, A.L., et al. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII, pp. 528–538. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_52

12. Gu, X., Hof, P.R., et al.: Anterior insular cortex and emotional awareness. J. Comp. Neurol. **521**(15), 3371–3388 (2013)

13. Guo, X., Wang, Z., et al.: Voxel-based assessment of gray and white matter volumes in Alzheimer's disease. Neurosci. Lett. **468**(2), 146–150 (2010)

14. He, K., Zhang, X., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

15. Hendrycks, D., Gimpel, K.: Gaussian error linear units (GELUs). arXiv preprint arXiv:1606.08415 (2016)

16. Kim, M., Kim, J., et al.: Interpretable temporal graph neural network for prognostic prediction of Alzheimer's disease using longitudinal neuroimaging data. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine, pp. 1381–1384. IEEE (2021)

17. Kim, W.H., Singh, V., Chung, M.K., Hinrichs, C., et al.: Multi-resolutional shape features via non-euclidean wavelets: applications to statistical analysis of cortical thickness. Neuroimage **93**, 107–123 (2014)

18. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)

19. Kundu, S., Lukemire, J., et al.: A novel joint brain network analysis using longitudinal Alzheimer's disease data. Sci. Rep. **9**(1), 1–18 (2019)

20. Lenzi, D., Serra, L., Perri, R., et al.: Single domain amnestic mci: A multiple cognitive domains fMRI investigation. Neurobiol. Aging **32**(9), 1542–1557 (2011)

21. Lu, Y.C., Kapse, K., et al.: Association between socioeconomic status and in utero fetal brain development. JAMA Netw. Open **4**(3), e213526–e213526 (2021)

22. Olde Dubbelink, K.T., Hillebrand, A., et al.: Disrupted brain network topology in Parkinson's disease: a longitudinal magnetoencephalography study. Brain **137**(1), 197–207 (2014)

23. Quarto, T., Blasi, G., et al.: Association between ability emotional intelligence and left insula during social judgment of facial emotions. PLoS ONE **11**(2), e0148621 (2016)

24. Rakesh, D., Zalesky, A., et al.: Similar but distinct-effects of different socioeconomic indicators on resting state functional connectivity: findings from the adolescent brain cognitive development (ABCD) study. Dev. Cogn. Neurosci. **51**, 101005 (2021)

25. Ribeiro, L.G., Busatto Filho, G.: Voxel-based morphometry in Alzheimer's disease and mild cognitive impairment: systematic review of studies addressing the frontal lobe. Dementia & Neuropsychol. **10**, 104–112 (2016)

26. Rogers-Carter, M.M., Varela, J.A., et al.: Insular cortex mediates approach and avoidance responses to social affective stimuli. Nat. Neurosci. **21**(3), 404–414 (2018)

27. Seidlitz, J., Váša, F., et al.: Morphometric similarity networks detect microscale cortical organization and predict inter-individual cognitive variation. Neuron **97**(1), 231–247 (2018)

28. Selvaraju, R.R., Cogswell, M., et al.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 618–626 (2017)

29. Spann, M.N., Bansal, R., et al.: Prenatal socioeconomic status and social support are associated with neonatal brain morphology, toddler language and psychiatric symptoms. Child Neuropsychol. **26**(2), 170–188 (2020)

30. Tolstikhin, I.O., Houlsby, N., et al.: Mlp-mixer: An all-mlp architecture for vision. In: Annual Conference on Neural Information Processing Systems, vol. 34 (2021)

31. Veit, A., Wilber, M.J., et al.: Residual networks behave like ensembles of relatively shallow networks. In: Annual Conference on Neural Information Processing Systems, vol. 29 (2016)

32. Wu, K., Taki, Y., et al.: A longitudinal study of structural brain network changes with normal aging. Front. Hum. Neurosci. **7**, 113 (2013)

33. Xu, K., Hu, W., et al.: How powerful are graph neural networks? arXiv preprint arXiv:1810.00826 (2018)

34. Yang, F., Isaiah, A., Kim, W.H.: COVLET: covariance-based wavelet-like transform for statistical analysis of brain characteristics in children. In: Martel, A.L., et al. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII, pp. 83–93. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_9

35. Yang, F., Meng, R., Cho, H., Wu, G., Kim, W.H.: Disentangled sequential graph autoencoder for preclinical Alzheimer's disease characterizations from ADNI Study. In: de Bruijne, M., et al. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II, pp. 362–372. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_34

36. Zlatar, Z.Z., Bischoff-Grethe, A., et al.: Higher brain perfusion may not support memory functions in cognitively normal carriers of the apoe $\varepsilon 4$ allele compared to non-carriers. Front. Aging Neurosci. **8**, 151 (2016)