

Graph-based deep learning models in the prediction of early-stage Alzheimers

Bishal Thapaliya^{1,2} Zundong Wu³ Ram Sapkota^{1,2} Bhaskar Ray^{1,2} Pranav Suresh^{1,2}
Santosh Ghimire⁴ Vince Calhoun^{2,3} Jingyu Liu^{1,2}

Abstract—Alzheimer’s disease is the most common age-related problem and progresses in different stages, from cognitively normal to early mild cognitive impairment, and severe dementia. This study investigates the predictive potential of resting-state functional magnetic resonance imaging (rs-fMRI) and its derived functional connectivity (FC) in understanding Alzheimer’s progression. Leveraging deep learning and graph-based models, we introduce two key contributions: 1) a comparative analysis of rs-fMRI time points and FC for Alzheimer’s prediction. 2) an innovative graph transformer variant incorporating self-clustering for enhanced prediction accuracy. Experiments on the Alzheimer’s Disease Neuroimaging Initiative dataset with 830 subjects reveal two notable conclusions. Firstly, rs-fMRI time points offer limited utility compared to functional network connectivity for transformer-based models, even when considering temporal information. Secondly, a clustering-based attention module proves effective for classifying brain networks in predicting Alzheimer’s disease progression, providing valuable insights for future research and clinical applications.

I. INTRODUCTION

Dementia continues to fly under the radar within healthcare systems, a result of insufficient educational outreach, awareness initiatives, and restricted access to diagnostic and care services for dementia [1], [2]. The diagnosis often comes later than ideal, leading to a bleak prognosis as even the most advanced FDA-approved medications show only modest effectiveness in alleviating cognitive and behavioral symptoms. Early therapeutic interventions not only enhance cognitive and behavioral functions in elderly patients but also empower them to make pivotal healthcare decisions, significantly enhancing their overall quality of life.

In recent years, mental health has become a focal point of concern, prompting significant efforts from traditional machine learning [3], genetics [4], [5], and especially deep learning [6], [7] to address underlying issues, where Alzheimer’s disease (AD) also takes the spotlight as the prevailing form of dementia in the elderly population, characterized by an irreversible decline in memory, cognition, and behavior. Mild cognitive impairment (MCI), positioned between typical age-related cognitive decline and dementia

[8], becomes a crucial target for preventive treatments for AD due to its propensity to precede various forms of dementia. The existing preventive medications, while approved, grapple with limited effectiveness over an initial period, sparking debates around their cost-to-benefit equilibrium. Early identification and regular follow-ups for individuals displaying MCI symptoms prove pivotal in discerning potential risks of progression to AD or other forms of dementia. Numerous ongoing studies are dedicated to this cause, leveraging diverse data sources such as multimodal neuroimaging, genetic data, and clinical records. For example, the Global Alzheimer’s Association Interactive Network (GAAIN), funded by the Alzheimer’s Association, hosts thirty-four live datasets (GAAIN Data 2017). Within this landscape, longitudinal studies emerge as crucial connectors [8], bridging the realms of clinical observation and neuro-pathological models.

In this paper, our focus shifts towards harnessing the potential of powerful graph-based deep learning models to discern the predictive capabilities concerning Alzheimer’s disease. Our objective is to delve into a comparative analysis, exploring the efficacy of employing functional network connectivity (FNC) matrices (where temporal information is condensed into correlations) as features, contrasting with the utilization of resting-state time points (rsfMRI) that incorporate both temporal and spatial information. Moreover, we introduce a novel variant of a self-clustering-based graph transformer tailored specifically for analyzing brain networks. This approach aims to enhance the understanding of intricate connections within brain function, providing a fresh perspective on the transformative potential of graph-based deep learning in unraveling the complexities associated with Alzheimer’s disease prediction. We leverage data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI), encompassing a cohort of 860 subjects. Through this exploration, we seek to demonstrate the efficient of deep learning models in deciphering complex brain activity patterns and identifying early-stage dementia.

II. MATERIALS AND METHODS

A. The ADNI Dataset

In this study, we extracted the relevant data from the database of ADNI, which is available publicly upon approval from the ADNI. We used the rsfMRI data of a total of 830 subjects where the labels for subjects were classified into early-stage Alzheimer’s disease (AD) (n=53), mild cognitive disorder (MCI) (n=308), and healthy controls (n=469). Data

*This work was supported by NIH grant R01AG063153.

¹ Department of Computer Science, Georgia State University, Atlanta, USA

² Tri-Institutional Center for Translational Research in Neuroimaging and Data Science

³ School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, USA

⁴ Department of Applied Sciences and Chemical Engineering, Tribhuvan University, Nepal

were split into training (n = 623), validation (n = 124), and holdout test (n = 125) subsets, and a 5-fold cross-validation was performed to ensure stable results in the experiments. We conducted preprocessing on the raw rs-fMRI data utilizing a combination of FSLv6.0 and SPM12 toolboxes, encompassing several key steps, namely: 1) rigid head motion correction; 2) distortion correction; 3) removal of dummy scans; 4) normalization to the Montreal Neurological Institute space; and 5) a 6mm Gaussian kernel smoothing. Subsequently, we employed a fully automated spatially constrained independent component analysis to extract 53 intrinsic components (ICNs) using the Neuromark_fmri_1.0 template [9] as the 53 nodes of graph models with 136 time points for each corresponding node. These ICNs were categorized into seven functional domains based on their anatomical locations and functional characteristics, which included subcortical, auditory, visual, sensorimotor, cognitive control, default mode, and cerebellar domains. Functional network connectivity (FNC) was computed as the Pearson correlation between the time courses of these intrinsic connectivity networks.

B. Proposed Architecture

The depicted architecture in Fig. 1 outlines the operational framework of our approach, which systematically evaluates and compares the predictive efficacy of two distinct inputs: rsfMRI time points and FNC. Our model was trained and optimized separately for rs-fMRI and FNC inputs. For rs-fMRI time points, we employed two methods: 1) Applying a Temporal Convolutional Networks (TCN) block to capture time series information, and concatenating it with Laplacian eigenvectors (LEs) of raw time series, extracting the Pearson correlation between output of TCN and using it as the input to the transformer model. 2) Directly using rsfMRI time points combined with Laplacian eigenvectors as the input to the model. Laplacian eigenvectors extracted from the node time series captured essential relative spatial position information. For training the architecture using only FNC, input features were FNC matrix, and there were no time series and no LEs. Subsequent steps involved a self-clustering graph attention block dynamically learning functional clusters within the brain, considering neighborhood relations through self-attention. This process was followed by a 3-layered fully connected layer for the final prediction in both cases. The architecture is specifically designed to assess and contrast the model's efficacy when processing rs-fMRI time points versus FC as inputs, offering valuable insights into their respective contributions to Alzheimer's disease prediction.

1) *Problem Definition:* The brain is spatially divided into N regions of interest (ROIs), which represent graph nodes indexed by the set $V = 1, \dots, N$. Let the characteristics of node i be denoted by $h_{i(raw)} \in \mathbb{R}^T$, representing the BOLD time series of length T , or the Pearson's correlation of BOLD time series between ROIs viewed as $h_{i(fnc)} \in \mathbb{R}^N$, constructed as the FNC matrix of dimension $N \times N$. FNC is computed as

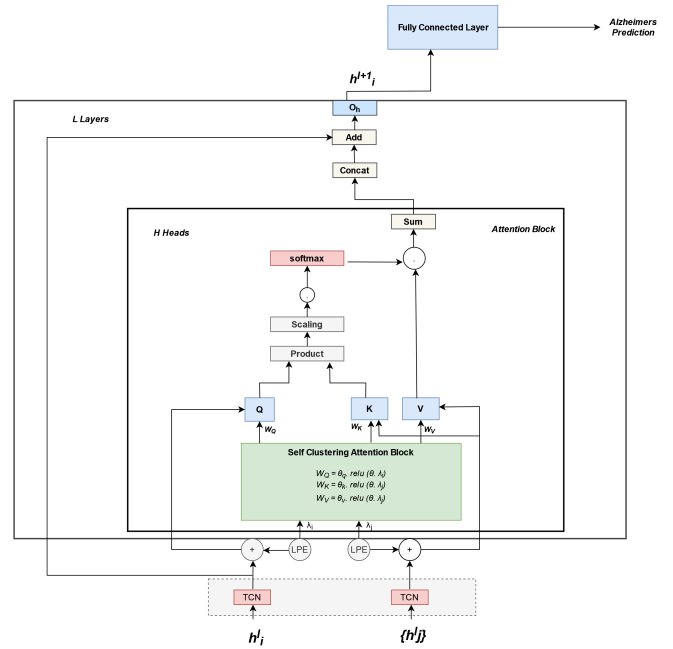


Fig. 1. Overall Architecture of clustering-based graph transformer model for early stage AD detection. TCN and LE blocks are optional.

Pearson's correlation using fMRI data series between ROIs which is also used to construct an adjacency matrix. An edge set E made up of $|E|$ unordered pairs (i, j) represents the connections between ROIs. For each edge linking two nodes $(i, j) \in E$, the connection strength, $e_{i,j}$, is the Pearson correlation coefficient between time series of node i and node j . The resultant graph is denoted by the tuple $G = (V, E)$. Let $H_{raw} \in \mathbb{R}^{N \times T}$, $H_{fnc} \in \mathbb{R}^{N \times N}$, $E \in \mathbb{R}^{|E| \times 1}$, and $A \in \mathbb{R}^{N \times N}$ symbolize the nodes features, edge features, and adjacency for the graph structure G . The Adjacency matrix $A_k \in \mathbb{R}^{N \times N}$ is formed as in Equation 1, where N_k is the number of ROIs in the k -th subject.

$$A_{i,j}^k = \begin{cases} 0, & i = j \\ e_{i,j}, & \text{otherwise} \end{cases} \quad (1)$$

2) *Temporal Convolutional Networks:* Temporal Convolutional Networks form the first block of our architecture when using rsfMRI time points as input features, and are responsible for extracting temporal features from the 1D time-series data of each brain node. It is well documented that temporal convolutional networks [10] can perform equally well or sometimes even better than Recurrent Neural Networks (RNNs) and Long-Short term memory networks (LSTMs) for many kinds of sequential data. Some advantages of this convolutional operator are, for instance, (1) low memory requirement for long input sequences, especially compared to LSTMs and Gated Recurrent Units (GRUs), which commonly consume big chunks of memory to store partial results for the multiple gates (convolutional kernels, in contrast, are shared across a layer), (2) better parallelization because a TCN layer is processed as a whole instead of sequentially as in RNNs, and (3) easier to train (e.g., it is known

that LSTM training can commonly encounter issues with vanishing gradients).

Formally, given a single ROI time series $t_i \in \mathbb{R}^T$ and a filter $f \in \mathbb{R}^K$, the dilated causal convolution operation of t with f at time t is represented as:

$$h_{i(\text{TCN})} = t_i * f(t) = \sum_{s=0}^{K-1} f(s) \times t_i(t - d \cdot s), \quad (2)$$

where $d = 2^{l_d - 1}$ is the dilation factor, and l_d denotes level of dilation in the network. When $d = 1$, a dilated convolution reduces to a regular convolution. Using larger dilation enables an output at the top level to represent a wider range of inputs, thus effectively expanding the receptive field of the network.

3) *Laplacian Eigenvector Block*: Positively acknowledged for their effectiveness in transformer-based models, positional encodings (PE) guide our attention to LEs, a pivotal element in spectral graph theory. These eigenvectors function as a representation of the inherent graph structure, mirroring the principles outlined in Graph Transformer [11]. Unlike traditional transformer PE, Laplacian PE is explicitly designed for graphs, providing an efficient means to encode distance/similarity-aware information. To address the multiplicity challenge resulting from arbitrary signs, we adopt a robust training strategy inspired by established practices in [11], [12], involving the random flipping of eigenvector signs. The Laplacian eigenvectors are computed as:

$$I - D_e^{-\frac{1}{2}} A D_e^{-\frac{1}{2}} = U^T \Lambda U$$

Here, D_e represents the degree matrix, while Λ and U denote the eigenvalues and eigenvectors, respectively. Specifically, we harness the m smallest non-trivial eigenvectors associated with each node for its positional encoding, referred to as λ_i for node i .

4) *Input Representation for Self-Clustering based Graph Model*: Incorporating various feature representations for Alzheimer's prediction, we first explore the use of rsfMRI time points in conjunction with a TCN block. The input to the self-clustering graph attention block for rsfMRI time points is defined as $\hat{h}_i = h_{i(\text{TCN})} + \lambda_i$, where λ_i signifies the PEs specific to node i . The resulting node features \hat{h}_i collectively form the matrix $\hat{\mathbf{H}}$ with dimensions $N \times D$, where N is the number of nodes, and D is the output dimension of the TCN block concatenated with m eigenvectors. Additionally, an alternative approach involves utilizing rsfMRI time points directly, incorporating Laplacian eigenvectors as $\hat{h}_i = h_{i(\text{raw})} + \lambda_i$. Furthermore, when using FNC as features, the input directly involves the FC matrix: $\hat{h}_i = h_{i(\text{FNC})}$. These diverse feature representations contribute to a comprehensive understanding of the importance of features for Alzheimer's disease progression within multiple frameworks.

5) *Self Clustering graph attention block*: Within the framework of the Transformer architecture [13], this module introduces a customized approach designed specifically for brain analysis. Diverging from conventional Transformer models, it doesn't treat all nodes uniformly during the training process. Instead, following the principles of BrainGNN [14] and BrainRGIN [15], it integrates domain-specific knowledge concerning functional clusters within the brain. This approach recognizes the intrinsic divisions in brain regions and their distinct interactions, essential for a comprehensive understanding of brain functionality. The input format for the clustering-based transformer is elucidated in Section II-B.4.

Let $W(K)_i^{(l)}$, $W(Q)_i^{(l)}$ and $W(V)_i^{(l)}$, parameters of attention block as in Fig. 1, be defined as a function of the PE (λ_i) for key, query, and value matrix. While θ_1 is the shared parameter across all key, query, and values to learn the clustering assignment for nodes, other parameters $\theta(k)_2$, $\theta(q)_2$, and $\theta(v)_2$ are defined for key, value, and query for the cluster-specific attention updates. Here, $\theta_1 \in \mathbb{R}^{N \times k_r}$, where k_r is defined as the number of clustered communities, λ_i is the LE encoded position for node i , and ReLU is the rectified linear unit activation function. The mathematical formulation for the clustering-based attention block can be defined as:

$$\mathbf{w}(\mathbf{k})_i^{(l)} = \theta(k)_2^{(l)} \cdot \text{relu}(\theta_1^{(l)} \cdot \lambda_i) + b_k^{(l)} \quad (3)$$

$$\mathbf{w}(\mathbf{q})_i^{(l)} = \theta(q)_2^{(l)} \cdot \text{relu}(\theta_1^{(l)} \cdot \lambda_i) + b_q^{(l)} \quad (4)$$

$$\mathbf{w}(\mathbf{v})_i^{(l)} = \theta(v)_2^{(l)} \cdot \text{relu}(\theta_1^{(l)} \cdot \lambda_i) + b_v^{(l)} \quad (5)$$

$$K_{ROI} = \mathbf{W}(\mathbf{K})^{(l)} \cdot \hat{\mathbf{H}} \quad (6)$$

$$Q_{ROI} = \mathbf{W}(\mathbf{Q})^{(l)} \cdot \hat{\mathbf{H}} \quad (7)$$

$$V_{ROI} = \mathbf{W}(\mathbf{V})^{(l)} \cdot \hat{\mathbf{H}} \quad (8)$$

The update equations for layer l are defined as:

$$\hat{h}_i^{l+1} = \mathcal{O}_h^l \left[\hat{h}_i + \parallel_{k=1}^K \left(\sum_{j \in N_i} w_{ij}^{k,l} \cdot V_{ROI} \cdot \hat{\mathbf{h}}_j \right) \right] \quad (9)$$

$$w_{ij}^{k,l} = \text{softmax}_j \left(\hat{w}_{ij}^{k,l} \right) \quad (10)$$

$$\hat{w}_{ij}^{k,l} = \frac{Q_{ROI} \cdot K_{ROI}^T}{\sqrt{D_k}} \quad (11)$$

Here, $k = 1$ to K denotes the number of attention heads, and \parallel denotes concatenation. $\mathcal{O}_h^l \in \mathbb{R}^{N \times D}$ represents linear transformations for node features after concatenating the outputs from multi-head attention. Upon obtaining the ultimate feature representations from the self-clustering attention block, we employ a 3-layered fully connected layer equipped with dropout and ReLU activation functions to achieve the final prediction.

In summary, we provide two major contributions:

- We conducted a *detailed comparative analysis* to assess the predictive power of rs-fMRI time points and FNC for Alzheimer's disease prediction.
- We present a *novel attention mechanism* designed for graphs with sub-networks, allowing distinct attention mechanisms for each cluster.

III. EXPERIMENTAL SETUP AND RESULTS

A. Training the model

The neural network architecture depicted in Fig. 1 was implemented using Pytorch [16], and Pytorch Geometric [17] for the specific graph neural network components. The number of nodes was 53 (corresponding to NeuroMark ICA Components), and the number of time points T for each node (i.e., 136). To ensure training robustness, we adopted a weighted random sampling strategy with a replacement inspired by the weighted random sampler concept. Each data point d_i is assigned a weight w_i based on its classification significance, forming a normalized probability distribution $P(d_i) = p_i$. This weighted random sampling is applied to create independent training and validation sets, preventing overlap and preserving model evaluation integrity. The approach mitigates biases, enhances generalization, and ensures a stable training process.

From the meticulous exploration of hyperparameters, we established fixed configurations for three models: the first model involves employing 4 TCN blocks and incorporates 8 eigenvectors (λ_i) for each node, and integrating two layers of the self-clustering graph attention block, generating output features of dimensionality D set at 128. The choice of 7 for the number of clustered communities (k_r) aligns with the seven functional networks delineated by [18]. The second model mirrors the first but excludes the TCN block, while the third model purely uses the FNC features. Our neural network training spanned 200 epochs, optimizing with the Adam optimizer [19] and utilizing the CrossEntropy loss function. The number of attention heads used for all the transformer models was fixed to be 8. To ensure efficiency, we implemented an early stopping mechanism, halting training if the validation loss failed to decrease over 30 consecutive epochs. The learning rate underwent reduction by a factor of 0.3, with a patience setting of 30. Throughout, a batch size of 32 was employed. All other hyperparameter settings remained uniform across these input configurations. The results reported are the average scores from a holdout test set with 5-fold cross-validation.

We compared our proposed model with popular models like Brain Network Transformer [20], BrainNetCNN [21], BrainRGIN [15] and Graph Transformers[11]. To evaluate the performance of our model, we conducted proof-of-concept experiments to classify the AD, MCI, and healthy controls for Alzheimer’s classification in a highly imbalanced dataset setting. Furthermore, it is important to note that all the configurations for baseline models were used as outlined in the original paper and further optimized to achieve optimal performance.

B. Results

Our model, incorporating self-clustering, demonstrates robust classification performance with a notable improvement in the area under the curve (AUC) compared to the original Graph Transformer and other baseline models. When using static FNC features, our model outperforms BrainRGIN,

TABLE I

COMPARISON OF DIFFERENT MODEL PERFORMANCES WHEN USING STATIC FNC VS RESTING-STATE fMRI AS FEATURES

Model	Features	AUC \pm Std. Dev	F1 \pm Std. Dev
Using static FNC features			
BrainRGIN [15]	FNC	0.64 \pm 0.03	0.63 \pm 0.02
Graph Transformer	FNC	0.68 \pm 0.02	0.65 \pm 0.01
Ours	FNC	0.76 \pm 0.03	0.68 \pm 0.02
BrainNetCNN [21]	FNC	0.72 \pm 0.01	0.67 \pm 0.01
Brain Network Transformer [20]	FNC	0.76 \pm 0.02	0.692 \pm 0.01
Using rsfMRI timepoints features			
Graph Transformer [11]	rsfMRI + LPE	0.62 \pm 0.02	0.65 \pm 0.01
Ours with TCN	rsfMRI + LPE	0.66 \pm 0.02	0.68 \pm 0.03
Ours	rsfMRI + LPE	0.65 \pm 0.03	0.63 \pm 0.03

Graph Transformer, and BrainNetCNN, achieving an AUC of 0.76 and an F1 score of 0.68. Brain Network Transformer achieves comparable performance with an AUC of 0.76 and an F1 score of 0.692.

However, using resting-state fMRI timepoints as features resulted in a surprising performance decline. With features extracted from TCN, our model achieved an AUC of 0.66 and an F1 score of 0.68, while without TCN, the AUC was 0.65 with an F1 score of 0.63. Notably, the original Graph Transformer model performed even worse when using time points as features compared to FNC features.

IV. DISCUSSION AND CONCLUSION

In conclusion, our study presents a preliminary version of our model, augmented with a self-clustering mechanism, demonstrating robust performance in Alzheimer’s disease classification, even amidst highly imbalanced class proportions. The integration of static functional network connectivity (FNC) features into our novel method yields significant enhancements, surpassing benchmarks set by established techniques such as BrainNetCNN, BrainRGIN, and Graph Transformer.

A pivotal aspect of our findings revolves around the detrimental impact of directly integrating timepoints as features into model. Despite attempts to enrich the model with temporal information through Temporal Convolutional Networks (TCN) in resting-state fMRI timepoints, we observed a pronounced decline in Alzheimer’s disease classification performance. This underscores the critical importance of feature selection, highlighting the superiority of connectivity-based features over timepoints, especially within transformer models.

In conclusion, our study showcases the potential of self-clustering mechanisms in enhancing neuroimaging-based classification tasks. Moving forward, further research into refining and optimizing self-clustering methodologies holds promise for advancing the field of neuroimaging-based disease classification, ultimately contributing to improved diagnostic accuracy and patient care in Alzheimer’s disease and beyond.

V. ACKNOWLEDGMENTS

This study was funded part by NIH grant R01AG063153.

REFERENCES

- [1] Andrea Bradford, Mark E. Kunik, Paul Schulz, Susan P. Williams, and Hardeep Singh, "Missed and delayed diagnosis of dementia in primary care: Prevalence and contributing factors," *Alzheimer Disease and Associated Disorders*, vol. 23, no. 4, pp. 306–314, Oct. 2009.
- [2] Amanda Connolly, Ella Gaehl, Helen Martin, Julie Morris, and Nitin Purandare, "Underdiagnosis of dementia in primary care: Variations in the observed prevalence and comparisons to the expected prevalence," *Aging and Mental Health*, vol. 15, no. 8, pp. 978–984, Nov. 2011.
- [3] Britny Farahdel, Bishal Thapaliya, Pranav Suresh, Bhaskar Ray, Vince D. Calhoun, and Jingyu Liu, "Confirmatory factor analysis on mental health status using abcd cohort," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2021, pp. 3540–3547.
- [4] Bishal Thapaliya, Vince D. Calhoun, and Jingyu Liu, "Environmental and genome-wide association study on children anxiety and depression," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2021, pp. 2330–2337.
- [5] Bishal Thapaliya, Bhaskar Ray, Britny Farahdel, Pranav Suresh, Ram Sapkota, Bharath Holla, Jayant Mahadevan, Jiayu Chen, Nilakshi Vaidya, Nora Perrone-Bizzozero, Vivek Benegal, Gunter Schumann, Vince D. Calhoun, and Jingyu Liu, "Cross-continental environmental and genome-wide association study on children and adolescent anxiety and depression," Feb. 2023.
- [6] Ram Sapkota, Bishal Thapaliya, Pranav Suresh, Bhaskar Ray, Vince D. Calhoun, and Jingyu Liu, "Multimodal imaging feature extraction with reference canonical correlation analysis underlying intelligence," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 2071–2075.
- [7] Pranav Suresh, Bhaskar Ray, Bishal Thapaliya, Britny Farahdel, Behnam Kazemivash, Jiayu Chen, Kuaikuai Duan, Vince D. Calhoun, and Jingyu Liu, "Effective training strategy for nn models of working memory classification with limited samples," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, 2023, pp. 1–4.
- [8] William R. Markesbery, "Neuropathologic alterations in mild cognitive impairment: A review," *Journal of Alzheimer's Disease*, vol. 19, no. 1, pp. 221–228, Jan. 2010.
- [9] Yuhui Du, Zening Fu, Jing Sui, Shuang Gao, Ying Xing, Dongdong Lin, Mustafa Salman, Anees Abrol, Md Abdur Rahaman, Jiayu Chen, L. Elliot Hong, Peter Kochunov, Elizabeth A. Osuch, and Vince D. Calhoun, "NeuroMark: An automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders," *NeuroImage: Clinical*, vol. 28, pp. 102375, 2020.
- [10] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *ArXiv*, vol. abs/1803.01271, 2018.
- [11] Vijay Prakash Dwivedi and Xavier Bresson, "A generalization of transformer networks to graphs," 2020.
- [12] Vijay Prakash Dwivedi, Chaitanya K. Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson, "Benchmarking graph neural networks," 2020.
- [13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," 2017.
- [14] Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H. Staib, Pamela Ventola, and James S. Duncan, "BrainGNN: Interpretable brain graph neural network for fMRI analysis," *Medical Image Analysis*, vol. 74, pp. 102233, Dec. 2021.
- [15] Bishal Thapaliya, Esra Akbas, Jiayu Chen, Raam Sapkota, Bhaskar Ray, Pranav Suresh, Vince Calhoun, and Jingyu Liu, "Brain networks and intelligence: A graph neural network based approach to resting state fmri data," 2023.
- [16] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimeshin, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala, "Pytorch: An imperative style, high-performance deep learning library," 2019.
- [17] Matthias Fey and Jan Eric Lenssen, "Fast graph representation learning with pytorch geometric," 2019.
- [18] B. T. Thomas Yeo, Fenna M. Krienen, Jorge Sepulcre, Mert R. Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L. Roffman, Jordan W. Smoller, Lilla Zöllei, Jonathan R. Polimeni, Bruce Fischl, Hesheng Liu, and Randy L. Buckner, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, vol. 106, no. 3, pp. 1125–1165, Sept. 2011.
- [19] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," 2014.
- [20] Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang, "Brain network transformer," 2022.
- [21] Jeremy Kawahara, Colin J. Brown, Steven P. Miller, Brian G. Booth, Vann Chau, Ruth E. Grunau, Jill G. Zwicker, and Ghassan Hamarneh, "Brainnetcn: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, Feb. 2017.