medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

BrainGT: Multifunctional Brain Graph Transformer for Brain Disorder Diagnosis

Ahsan Shehzad, Shuo Yu, Dongyu Zhang, Shagufta Abid, Xinrui Cheng, Jingjing Zhou, Feng Xia

***Abstract*—Brain networks play a crucial role in the diagnosis of brain disorders by enabling the identification of abnormal patterns and connections in brain activities. Previous studies exploit the Pearson's correlation coefficient to construct functional brain networks from fMRI data and use graph learning to diagnose brain diseases. However, correlation-based brain networks are overly dense (often fully connected), which obscures meaningful connections and complicates subsequent analyses. This dense connectivity poses substantial performance challenges to traditional graph transformers, which are primarily designed for sparse graphs. Consequently, this results in a notable reduction in diagnostic accuracy. To address this challenging issue, we propose a multifunctional brain graph transformer model for brain disorders diagnosis, namely BrainGT, which is capable of constructing multifunctional brain networks rather than a dense brain network from fMRI data. It utilizes the fusion of self-attention and cross-attention mechanisms to learn important features within and across multiple functional brain networks. Classification (diagnosis) experiments conducted on three real fMRI datasets (i.e., ADNI, PPMI, and ABIDE) demonstrate the superiority of the proposed BrainGT over state-of-the-art methods.**

***Impact Statement*—The proposed BrainGT model represents a substantial advancement in computational neuroscience, offering a promising tool for more accurate and efficient diagnosis of brain disorders. By constructing multifunctional brain networks from fMRI data, BrainGT overcomes the limitations of traditional graph transformers and correlation-based brain networks. This innovation has profound implications across social, economic, and technological dimensions. Socially, BrainGT can enhance the quality of life for individuals with brain disorders by enabling more accurate diagnoses, leading to more effective treatments and better patient outcomes. Economically, BrainGT has the potential to reduce healthcare costs by streamlining the diagnostic process and potentially reducing the need for more expensive or invasive procedures. Technologically, BrainGT pushes the boundaries of AI and neuroscience, opening new avenues for research and development. It demonstrates the potential of AI to handle complex and dense data structures, with applications that could extend to other fields.**

***Index Terms*—Brain graph transformer, brain network analysis, brain disorder diagnosis, graph learning, graph neural networks, functional connectivity**

A. Shehzad and S. Abid are with School of Software Technology, Dalian University of Technology, Dalian 116620, China (e-mail: {ahsan;shagufta}@mail.dlut.edu.cn).

D. Zhang is with the School of Software Technology and the School of Foreign Languages, Dalian University of Technology, Dalian 116620, China (e-mail: zhangdongyu@dlut.edu.cn)

S. Yu is with School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China (e-mail: shuo.yu@ieee.org)

J. Zhou is with the School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou, China (e-mail: zhoujingjing@zjgsu.edu.cn)

F. Xia and X. Cheng are with School of Computing Technologies, RMIT University, Melbourne, VIC 3000, Australia (e-mail: f.xia@ieee.org, xcheng430@outlook.com))

I. Introduction

THE diagnosis of brain diseases using brain network analysis is a rapidly evolving field of research that holds significant promise for improving medical outcomes [1]. Functional Magnetic Resonance Imaging (fMRI) plays a crucial role in the construction of these brain networks by measuring brain activity and identifying connectivity patterns between different regions [2]. The significance of brain network analysis lies in its potential to provide a more comprehensive understanding of brain function and dysfunction enabling the early detection and diagnosis of various brain diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), and Autism pectrum disorder (ASD) [3]. By analyzing alterations in brain network connectivity, researchers and clinicians can identify biomarkers and develop targeted interventions that ultimately lead to more effective treatments and improved patient outcomes.

Numerous efforts have been made to analyze brain networks for the diagnosis of brain diseases. Conventional brain network analyzes employing graph theory-based techniques generally follow a two-step process [2], [4]. Firstly, feature engineering is applied to graphs followed by an analysis of derived features. In the feature engineering phase, graph property metrics such as clustering coefficient are used to encapsulate the functional connectivity of each node into statistical measures. Then, due to high dimensionality of fMRI data, regions of interest (ROIs) are frequently grouped into highly interconnected communities to reduce dimensionality or to enable data-driven feature selection [5]. However, in these two-step methods, inaccuracies in the first step can result in substantial errors in subsequent analysis.

Graph neural networks (GNNs) have gained popularity for end-to-end graph learning, resulting in notable advancements in brain network analysis [6], [7]. These methods autonomously extract salient features from data sets that aid in the diagnosis of brain disorders. GNNs process data from the brain network to define intricate relationships between nodes (brain regions) and edges (functional connections) [8]. However, their limited receptive field restricts comprehension of the brain network's global context, potentially omitting critical diagnostic patterns. The emergence of graph transformers represents a pivotal development in graph learning that offers substantial improvements in various applications, including the diagnosis of brain disease [9], [10]. By integrating attention mechanisms, graph transformers provide a more comprehensive and adaptable interpretation of graph structures [11]. Initial applications of graph transformer-based approaches to

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

brain networks have shown improved diagnostic accuracy for neurological conditions [12], [13].

However, because of the widespread adoption of the Pearson correlation method in constructing brain networks from fMRI data, current correlation-based brain networks are overly dense (often fully connected). This excessive density obscures meaningful connections and complicates subsequent analyses [14]. As a consequence, previous methods, that are generally optimized for sparse graphs, are facing a significant efficacy decrease. The inefficiencies and potential inaccuracies in analysis compromise the detection and understanding of neurological disorders. This underscores the need to develop more precise and efficient methods for brain disorder diagnosis.

In addressing the challenges brought by the complexities of fully connected brain networks, we propose BrainGT, a unified framework aimed at improving brain network construction and analysis. The proposed BrainGT first segments the fully connected brain network into several functional brain networks that reflect the brain's inherent structure, and then employs specialized graph transformers with a dual attention mechanism. The dual attention mechanism combines self-attention to identify intricate local relationships within each functional network and cross-attention to understand interactions between different functional networks [15]. Our proposed BrainGT offers a holistic global perspective and effectively tackles the challenges associated with fully connected networks and demonstrate superior performance in the diagnosis of brain diseases, marking a significant progression in neuroimaging analytics. Our contributions are summarized as follows.

- We propose an effective solution for brain disorders diagnosis, namely BrainGT, which is an unified framework designed for fully connected brain networks and thus improving the diagnosis of neurological conditions.
- The proposed BrainGT generates multiple functional brain networks based on the organization of different brain functions, as opposed to a fully connected network from fMRI scan. It utilizes fusion of self-attention and cross-attention mechanisms to learn important features within and across multiple functional brain networks, respectively.
- We validate our model using three different datasets: ADNI, PPMI, and ABIDE, each representing a different brain disease: AD, PD, and ASD, respectively. The experimental results demonstrate that BrainGT outperforms state-of-the-art methods.

The remainder of this paper is organized as follows. Section II examines related work, emphasizing significant contributions and pinpointing gaps that our study seeks to address. Section III offers a comprehensive background that is essential for grasping the context of this paper. Section IV explores the design of BrainGT, highlighting its architecture and implementation. Section V presents the experimental results and discussions, evaluating the performance and implications of our design. Lastly, Section VI wraps up the paper and proposes directions for future research.

II. Related Work

A. Brain Network Analysis for Brain Disorder Diagnosis

The human brain is a complex network of interconnected regions involved in sophisticated communication patterns [16]. Brain network analysis is crucial in neuroscience to understand functional connectivity between brain regions and helps to identify biomarkers for the early diagnosis of neurological disorders [17]. Brain networks are defined using fMRI data and the Pearson correlation coefficient to measure linear relationships between BOLD signals in different areas of the brain [18], resulting in a network model in which nodes represent regions of the brain and edges represent functional connectivity. The initial methodologies employing graph theory have been pivotal in defining the structural and functional organization of the brain [5]. These models have been instrumental in identifying critical hub regions and pathways essential for neural communication and cognitive processes. However, despite their significant contributions, graph theory-based approaches exhibit inherent limitations. They often require manual feature selection, which introduces subjectivity and potentially overlooks the intricate complexity of brain connectivity.

Graph learning, particularly GNNs, represents a significant advancement in brain network analysis [19]. These models autonomously learn and discern patterns in brain connectivity data, improving the understanding of brain functions and pathologies. Cui et al. [6] provided a benchmark called BrainGB for brain network analysis using GNNs, highlighting challenges like the lack of useful initial node features and real-valued connection weights. Anwar et al. [20] offered an evaluation framework for graph machine learning models in brain connectomics, introducing benchmark datasets and standardized metrics for consistent comparisons. Wang et al. [9] proposed an unsupervised contrastive graph learning framework for fMRI data analysis, enhancing discriminative feature learning for the detection of brain disorders. Li et al. [10] developed an interpretable GNN framework for fMRI data, focusing on the interpretability of predictions by identifying key brain regions and interactions. Ma et al. [12] presented a multiscale dynamic graph learning framework that takes advantage of spatio-temporal dynamics in fMRI data, improving the robustness and accuracy of brain disorder detection. However, previous graph learning approaches face challenges such as limited receptive fields, which hinder their ability to capture comprehensive network properties, affecting models' effectiveness in tasks requiring a holistic understanding of brain connectivity.

B. Graph Transformers

Graph transformers are a novel type of graph neural network that extends the transformer architecture to various graphs [21]. By using self-attention mechanisms to derive node and edge representations from graph-structured data, graph transformers capture complex dependencies and interactions among nodes and edges. Unlike GNNs, which rely on localized message passing and can suffer from over-smoothing, graph transformers effectively capture long-range

medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

dependencies within graphs [22]. These models have been applied in fields such as natural language processing [23], computer vision [24], social network analysis [25], and drug discovery [26]. Graph transformers have also shown promise in the diagnosis of brain diseases by using brain networks and self-attention mechanisms to capture extensive dependencies between regions and brain modalities.

Kan et al. [14] proposed the Brain Network Transformer, which uses connection profiles as node features and learns pairwise connection strengths with efficient attention weights. This model incorporates an Orthonormal Clustering Readout operation, resulting in cluster-aware node embeddings and informative graph embeddings. Fang et al. [27] introduced the Path-based Heterogeneous Brain Transformer Network (PH-BTN), which constructs brain graphs from rs-fMRI data and learns compact edge features through heterogeneous graph paths, enhancing brain network analysis. Other approaches address specific challenges in brain network analysis. Zuo et al. [28] developed a Distribution-Regularized Adversarial Graph Autoencoder (DAGAE) with a Transformer generator for dementia diagnosis, preprocessing fMRI data to construct graph data, and using adversarial training to generate robust functional networks of the brain. Cai et al. [29] proposed a model to estimate brain age as a biomarker for AD diagnosis. Bannadabhavi et al. [30] introduced Com-BrainTF, a transformer that predicts Autism by integrating community information in fMRI analysis. Dai et al. [11] proposed THC, a model for identifying brain modules and classifying networks capable of handling dynamic brain networks and predicting lesion locations. A notable challenge in applying Graph Transformers to brain networks is their typical design for sparse graph data, which is inefficient for fully connected brain networks. This leads to computational inefficiencies and hinders significant pattern detection. Some approaches mitigate this by sparsifying the network or modifying the attention mechanism to focus on critical connections, but these adaptations do not fully leverage the potential of Graph Transformers.

III. Preliminaries

A. Brain Networks

A brain network can be mathematically defined as a graph $G = (V, E)$, where V is the set of nodes representing distinct brain regions, and E is the set of edges representing the functional or structural connections between these regions [1]. Each node $v_i \in V$ is associated with a specific ROI, and each edge $e_{ij} \in E$ between nodes v_i and v_j quantifies the strength or presence of a connection, which can be derived from neuroimaging data. In the context of functional brain networks, let A be the adjacency matrix representing the network, where A_{ij} is the weight of the edge between nodes i and j. For functional connectivity, A_{ij} can be computed using the Pearson correlation coefficient between the time series of neural activity from regions i and j, as previously described:

$$A_{ij} = r_{ij} = \frac{\sum_{t=1}^{T}(X_i(t)-\overline{X}_i)(X_j(t)-\overline{X}_j)}{\sqrt{\sum_{t=1}^{T}(X_i(t)-\overline{X}_i)^2}\sqrt{\sum_{t=1}^{T}(X_j(t)-\overline{X}_j)^2}}, \quad (1)$$

Here, the matrix A captures the connectivity pattern of the brain network, where the value of A_{ij} indicates the strength of the functional connection between regions i and j. For structural brain networks, A_{ij} might represent the number of white matter tracts connecting regions i and j, derived from diffusion MRI data. By representing the brain as a network, graph property metrics can be applied to analyze its properties and understand the differences between healthy and diseased states, ultimately contributing to the diagnosis and treatment of brain disorders [31]. This process of constructing a brain network from fMRI data using the Pearson correlation method is illustrated in Fig. 1.

B. Graph Learning and Graph Transformers

Graph learning is a subfield of machine learning that focuses on the analysis and interpretation of data structured as graphs [32]. In a graph $G = (V, E)$, V represents the set of nodes (or vertices), and E represents the set of edges that connect pairs of nodes. The graph learning problem can be mathematically defined as learning a function $f : G \to Y$ that maps a graph G or its components to target values Y, where Y could represent node labels, edge labels, or global graph properties. Formally, this involves optimizing a loss function $\mathcal{L}(f(G), Y)$ over a training dataset $\mathcal{D} = \{(G_i, Y_i)\}_{i=1}^{N}$. Graph learning encompasses a broad range of tasks including node classification, link prediction, and graph classification [33]. In the context of brain network analysis, graph learning techniques are employed to uncover patterns and connections within brain networks that can aid in diagnosing brain diseases [34]. The application of these techniques leverages the structural and functional connectivity data obtained from neuroimaging modalities like fMRI and diffusion MRI [35].

Graph transformers represent a significant evolution in the field of graph learning combining the power of transformers with graph-structured data to capture complex relationships and dependencies [36]. Traditionally, transformers have been highly successful in natural language processing and computer vision due to their ability to model long-range dependencies and interactions through self-attention mechanisms [37]. The extension of transformers to graphs has opened new avenues for analyzing and interpreting complex graph data including brain networks. Traditional graph learning methods are mainly based on localized operations to aggregate information from neighboring nodes. Although these methods are effective for various applications, they often face challenges due to their limited receptive fields. This limitation hinders their ability to capture long-range dependencies and the global structure of the graph [38]. Graph transformers address this limitation by using a global attention mechanism that allows each node to attend to all other nodes in the graph, thus capturing more comprehensive and nuanced relationships [39]. Graph transformers incorporate several key features and mechanisms that distinguish them from traditional graph learning methods. At their core is self-attention mechanism which can be mathematically described as follows:

Given a set of node features $X \in \mathbb{R}^{N \times d}$, where N is the number of nodes and d is the feature dimension, the self-attention mechanism computes attention scores using query,

medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

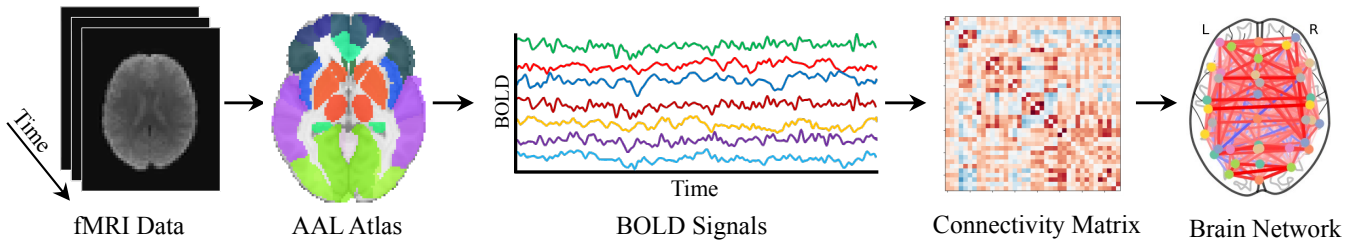


Fig. 1. Traditional brain network construction using Pearson correlation method on fMRI data.

key and value matrices Q, K, and V, derived from the input features:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V, \tag{2}$$

The attention scores are then computed as:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right), \tag{3}$$

where d_k is the dimension of the key vectors. The output of the self-attention layer is given by:

$$\text{Attention}(Q, K, V) = AV. \tag{4}$$

This mechanism allows each node to aggregate information from all other nodes weighted by attention scores enabling model to capture global dependencies. Recent advancements in graph transformer models have introduced various enhancements to improve their efficiency and effectiveness [40], [41]. For example, Graphormer model [42] integrates additional structural encodings such as shortest path distances and centrality measures into the attention mechanism, improving its ability to capture structural information from graphs. Another model SAN (Structure-Aware Transformer) [43] incorporates structural priors directly into the self-attention mechanism to enhance its performance on graph tasks.

IV. The Design of BrainGT

A. The Overview of BrainGT

BrainGT represents a novel framework designed for the diagnosis of brain diseases by addressing the complexity of dense functional connectivity in brain networks. This framework constructs multifunctional brain networks using the AAL[1] and Yeo2011[2] atlases to bypass the challenges associated with dense network analysis. The multifunctional graph transformer module takes these networks as input embeddings, incorporates absolute positional encoding for spatial context, and processes each brain network independently with network-wise encoders. Internetwork relationships are captured through cross-network encoders, and the system culminates in adaptive fusion classification to predict the presence of brain disease. The framework and workflow of BrainGT are depicted in Fig. 2.

[1]https://www.gin.cnrs.fr/en/tools/aal

[2]https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation_Yeo2011

B. Multifunctional Brain Networks Construction

1) Preprocessing of fMRI Data: This pivotal procedure converts raw fMRI data into a processed dataset with slice timing correction to synchronize the acquisition times of fMRI image slices. Let T_i represent the acquisition time of the ith slice and T_{ref} the reference time. The corrected time T_{corr} for each slice is computed as follows:

$$T_{corr}(i) = T_i - T_{ref}. \tag{5}$$

Subsequently, motion correction is implemented to compensate for any head movements during the scan. This process entails the estimation of motion parameters M_t at each time point t, and the alignment of each volume to a reference volume by minimizing the cost function $C(M)$, defined as the sum of squared differences between the image at time t and the reference image:

$$C(M_t) = \sum_{x,y,z} [I_t(x, y, z) - I_{ref}(M_t(x, y, z))]^2. \tag{6}$$

Following motion correction, spatial normalization is conducted to ensure that each individual's brain images conform to a standardized template. The transformation matrix N is calculated to map the brain images to the template space T, aiming to minimize the discrepancy between the individual image I and the template T:

$$N(x, y, z) = argmin_N \|I(x, y, z) - T(N(x, y, z))\|. \tag{7}$$

To enhance the signal-to-noise ratio, the images are smoothed by convolving them with a Gaussian kernel $G(\sigma)$, where σ represents the standard deviation of the kernel, dictating the degree of smoothing:

$$S(x, y, z) = I(x, y, z) * G(\sigma). \tag{8}$$

The preprocessing concludes with detrending, which involves the removal of linear trends from the time series data. For a time series $X(t)$, the detrended series $D(t)$ is derived by subtracting the linear trend, characterized by coefficients α and β:

$$D(t) = X(t) - (\alpha t + \beta). \tag{9}$$

2) Atlas Registration: The Atlas registration module is crucial for constructing multifunctional brain networks. It uses Yeo2011 and AAL atlases to define functional networks and anatomical ROIs. Both atlases are registered to a standardized space using the MNI152 template. This alignment ensures a

medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

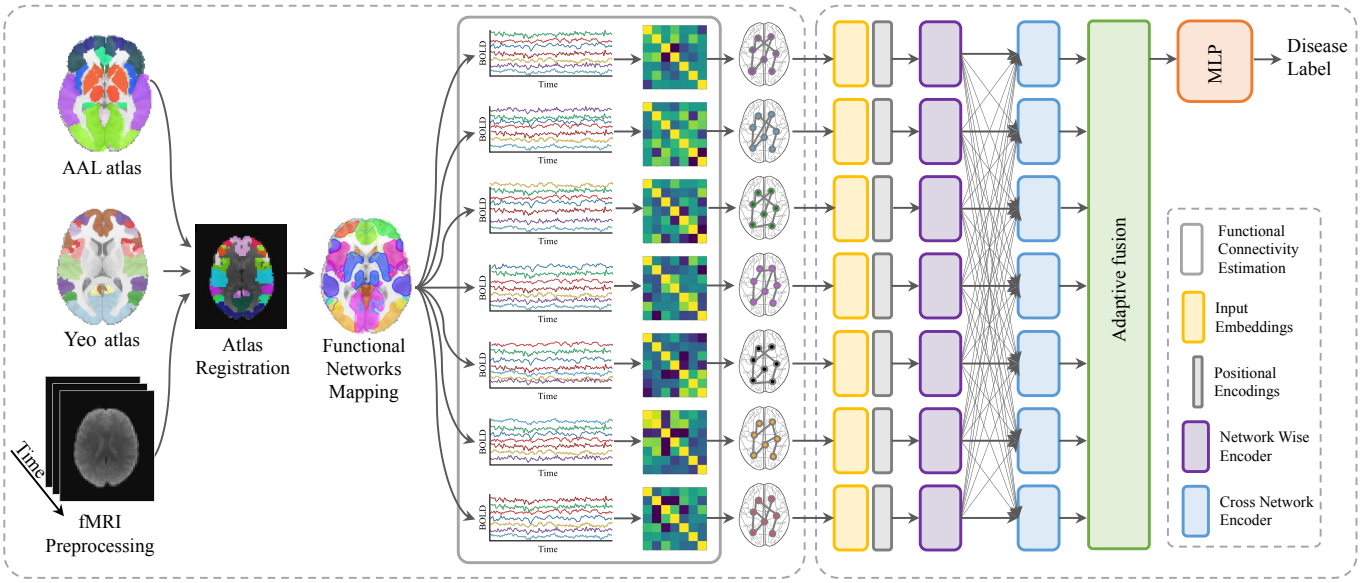


Fig. 2. The framework of BrainGT.

precise overlay and comparison. The registration involves a transformation matrix R, which adjusts the atlas coordinates A to the template space T:

$$R(A) = T. \tag{10}$$

The AAL atlas is resampled to match the voxel dimensions of the Yeo2011 atlas, ensuring alignment of anatomical labels with Yeo2011's functional networks. The resampling function f adjusts V_{AAL} to match V_{Yeo}:

$$f(V_{AAL}) = V_{Yeo}. \tag{11}$$

After resampling, anatomical labels from the AAL atlas are superimposed onto Yeo2011 functional networks. Each voxel v receives a functional network label F_{Yeo} and an anatomical label A_{AAL}, providing dual descriptors:

$$L(v) = (F_{Yeo}(v), A_{AAL}(v)). \tag{12}$$

The preprocessed fMRI data is then transformed into a common space using the registration matrix R, ensuring accurate alignment with the atlases. The transformed fMRI data D' is represented as:

$$D' = R(D). \tag{13}$$

3) Functional Networks Mapping: This module precisely maps anatomical ROIs to their respective functional networks by aggregating anatomical ROIs within each functional network, as defined by the Yeo2011 atlas. For a specific functional network f_i, the aggregation C_{f_i} represents the union of all anatomical ROIs R_{a_j} contained within f_i:

$$C_{f_i} = \bigcup_j R_{a_j} \text{ where } R_{a_j} \text{ is within } f_i. \tag{14}$$

Subsequent to aggregation, majority voting is utilized to determine the most representative anatomical Region of Interest (ROI) within each functional network. This determination is made by counting the occurrences n_{a_j} of each ROI R_{a_j}, and selecting the ROI with the maximal frequency for the network f_i:

$$R^*_{f_i} = \arg\max_{R_{a_j}} n_{a_j} \text{ within } f_i. \tag{15}$$

The ROIs that frequently recur within each functional network are subsequently defined as nodes, thereby converting the aggregated data into a structured network. The set of nodes N_{f_i} for each functional network f_i is defined as follows:

$$N_{f_i} = \{R^*_{f_i} | R^*_{f_i} \text{ is the majority ROI within } f_i\}. \tag{16}$$

The output of this module comprises a series of functional networks, each characterized by nodes representing the predominant anatomical ROIs.

4) Functional Connectivity Estimation: The functional connectivity is estimated using temporal correlations between ROIs within each functional network. Specifically, for a node corresponding to ROI $R^*_{f_i}$ in a functional network f_i, the BOLD time series $T_{R^*_{f_i}}$ is derived by averaging the signal intensities across all time points T:

$$T_{R^*_{f_i}} = \frac{1}{T} \sum_{t=1}^{T} S_{R^*_{f_i}}(t), \tag{17}$$

where $S_{R^*_{f_i}}(t)$ denotes the signal intensity at time t. Subsequently, a correlation analysis was performed with a seed region of interest (ROI), R_{seed}, serving as a reference. The Pearson correlation coefficient, r, was calculated between the time series of the seed ROI, $T_{R_{seed}}$, and the time series of each respective ROI, $T_{R^*_{f_i}}$, to generate a correlation map:

medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

$$r_{R_{seed},R^*_{f_i}} = \frac{\sum(T_{R_{seed}} - \bar{T}_{R_{seed}})(T_{R^*_{f_i}} - \bar{T}_{R^*_{f_i}})}{\sqrt{\sum(T_{R_{seed}} - \bar{T}_{R_{seed}})^2 \sum(T_{R^*_{f_i}} - \bar{T}_{R^*_{f_i}})^2}}. \tag{18}$$

This analysis across all ROIs produced whole-brain connectivity maps showing functional interconnections. These maps reveal synchronous activity patterns, with positive correlations indicating similarity and negative correlations indicating anticorrelations. Significant connections are determined by applying a threshold θ to the correlation coefficients C. The binary connectivity matrix B is defined such that $B_{ij} = 1$ if the absolute value of the correlation coefficient between ROIs i and j is at least θ, and 0 otherwise:

$$B_{ij} = \begin{cases} 1 & \text{if } |C_{ij}| \geq \theta \\ 0 & \text{otherwise} \end{cases}. \tag{19}$$

The binary matrices are transformed into a set of graphs $\{G_f\}_{f=1}^N$ where $G_f = (N, E)$, where nodes N represent ROIs and edges E indicate significant correlations.

C. Multifunctional Brain Graph Transformers

1) Input Embeddings: This module structures raw brain network data for graph transformers. Brain networks are graphs $G = (V, E)$, where V are brain regions and E are functional connections. Each node $v_i \in V$ has feature vectors x_i, transformed into embeddings h_i by a learnable function f, such that $h_i = f(x_i)$. Edge weights w_{ij} are projected into higher-dimensional space h'_{ij} by function g, where $h'_{ij} = g(w_{ij})$. An aggregation function AGG combines node and edge embeddings for each node i, yielding h''_i. This process is mathematically expressed as:

$$h''_i = \text{AGG}(h_i, \{h'_{ij} | j \in \mathcal{N}(i)\}), \tag{20}$$

where $\mathcal{N}(i)$ denotes the set of neighboring nodes connected to node i.

2) Positional Encoding: The Positional Encoding module facilitates the model's comprehension of the absolute positions of nodes, which is essential for discerning spatial relationships in the diagnosis of brain diseases. It assigns a distinct positional encoding to each node based on its connectivity profile within the functional network [14]. This profile, represented as vector c_i, encapsulates the node connectivity pattern and is transformed into an absolute position encoding p_i by a mapping function ϕ, such that $p_i = \phi(c_i)$. This mechanism augments initial embeddings by integrating positional encoding into each node, thereby producing spatially aware embeddings h'''_i:

$$h'''_i = h''_i + p_i. \tag{21}$$

This addition ensures each node's embedding reflects its features, connectivity, and specific brain network location.

3) Network-wise Encoders: The Network-Wise Encoders module within the BrainGT framework is responsible for encoding structural information into the model's representations by processing each brain network independently. This module takes as input spatially aware node embeddings $\{h'''_i\}$, which contain detailed information regarding the nodes' features and spatial positions but lack structural context. To mitigate this deficiency, the module utilizes a self-attention mechanism that uses a connectivity map to integrate structural information directly into the attention process [15]. The connectivity map delineates the structural connections between nodes, thereby forming the foundation of the model's self-attention mechanism.

$$\text{Attention}(Q, K, V, C) = \text{softmax}\left(\frac{QK^T + \alpha C}{\sqrt{d_k}}\right)V, \tag{22}$$

where, Q, K, and V represent the query, key, and value matrices, respectively, which are derived from the node embeddings. The parameter α serves as a scaling factor that adjusts the impact of the connectivity map on the attention scores. Additionally, d_k denotes the dimensionality of the key vectors.

4) Cross-network Encoders: This module elucidates interactions among brain networks. The Cross-network encoders receive encoded features from Network-Wise encoders and use cross-attention to learn inter-network dynamics. It generates a query Q from one network's features and compares it with keys K from other networks. Attention scores from these interactions capture the influence of one network's features on another [44]. The mathematical formulation of the cross-attention process is as follows.

$$\text{Cross-Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{23}$$

where V denotes the value matrices and d_k represents the dimensionality of the keys. This formulation enhances the attention mechanism's sensitivity to the interplay among brain networks, aiding the model in discerning their collective influence on brain functionality and potential pathological conditions. The output is an integrated feature set encapsulating aggregated data across all brain networks.

5) Adaptive Fusion and Classification: This module processes features from Cross-Network Encoders, aggregating information across brain networks. The adaptive fusion mechanism assigns weights to features based on their classification significance. The classification phase uses a Multi-Layer Perceptron (MLP) with multiple neuron layers and activation functions to predict brain disease.

$$y = \sigma(W_n \ldots \sigma(W_2 \sigma(W_1 x + b_1) + b_2) \cdots + b_n), \tag{24}$$

where x is the fused input features, W and b are the weights and biases of the MLP layers, σ is the activation function, and y is the output prediction vector. To enhance classification robustness and reduce overfitting, the module uses a cross-entropy loss function with L2 regularization, defined as:

medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

$$L = -\sum_{i=1}^{N} t_i \log(y_i) + \lambda \sum_{w} w^2, \tag{25}$$

where t_i are the true labels, y_i are the predicted labels, λ is the regularization parameter, and w represents the MLP weights. An optimization algorithm like SGD adjusts weights and biases to minimize the loss function and improve classifier performance. The Adaptive Fusion and Classification module outputs a diagnostic prediction for brain disease likelihood. The pseudocode of BrainGT is shown in Algorithm 1.

Algorithm 1 BrainGT Algorithm

Input: Raw fMRI data
Output: Disease prediction: y
procedure BRAINGT(fMRI_{Raw})
  $\text{fMRI}_{\text{P}} \leftarrow \text{Preprocess}(\text{fMRI}_{\text{Raw}})$
  $\text{fMRI}_{\text{R}} \leftarrow \text{AtlasRegistration}(\text{fMRI}_{\text{P}})$
  $\{N_f\}_{f=1}^{N} \leftarrow \text{NetworksMapping}(\text{fMRI}_{\text{R}})$
  $\{C_f\}_{f=1}^{N} \leftarrow \text{ConnectivityEstimation}(\{N_f\}_{f=1}^{N})$
  $\{G_f\}_{f=1}^{N} \leftarrow \text{NetworksConstruction}(\{C_f\}_{f=1}^{N})$
  $h_v \leftarrow \text{InputEmbeddings}(\{G_f\}_{f=1}^{N})$
  Initialize weights θ for Transformer encoders
  $p_v \leftarrow \phi(c_v)$ for each node v
  $E \leftarrow$ number of epochs
  for $e \leftarrow 1$ to E **do**
    for $f \leftarrow 1$ to N **do**
      $h'_f \leftarrow \text{NetworkEncoder}(\{h_v\}_{v \in G_f})$
      $h''_f \leftarrow \text{CrossNetworkEncoder}(\{h'_v\}_{v \in G_f})$
    end for
  end for
  $h_{\text{final}} \leftarrow \text{AdaptiveFuse}(\{h''_f\}_{f=1}^{N})$
  $y \leftarrow \sigma(W_n \ldots \sigma(W_1 h_{\text{final}} + b_1) \ldots + b_n)$
  Compute loss $\mathcal{L}$
  Backpropagate error and update θ
  return y
end procedure

V. Experimental Results

A. Experimental Settings

1) Datasets:

a) ADNI Dataset: , namely The Alzheimer's Disease Neuroimaging Initiative (ADNI)[3] dataset, derived from the ADNI database, comprises an extensive collection of fMRI data designed to explore the progression of Alzheimer's Disease (AD). Researchers can access the dataset following approval of their application, which supports research into neuroimaging and biomarker analysis. The primary objective with this dataset involves the identification of biomarkers for the detection and monitoring of AD. Subjects are classified into three categories: cognitively normal, mild cognitive impairment, and AD. A specific subset of the ADNI dataset, chosen for its inclusion of resting-state fMRI data, consists of 426 participants, 199 of whom (46.7%) are female and 146 (34.2%) have been diagnosed with AD, ranging in age from 50 to 100 years. The selection process was rigorous, ensuring that only subjects with confirmed diagnoses were included.

TABLE I
THE STATISTICES OF DATASETS

Dataset	Subjects	Age Range	Male/Female	Labels
ADNI	426	50-100	227/199	CN (280), AD (146)
PPMI	324	40-85	194/130	Control (94), PD (230)
ABIDE	1118	8-40	957/161	Control (581), ASD (537)

b) PPMI Dataset: , namely The Parkinson's Progression Markers Initiative (PPMI)[4] dataset, available on the PPMI website, contains fMRI data to study Parkinson's Disease (PD) progression. Researchers meeting specific criteria can access this dataset to explore PD biomarkers. The dataset aims to identify diagnostic and progression markers for PD, distinguishing between healthy controls and PD patients. The study cohort includes fMRI data from 823 individuals, with 230 (70.9%) diagnosed with PD, 130 (40.1%) female, and ages 40 to 85, representing a cross-section of the PD-affected population.

c) ABIDE Dataset: , namely The Autism Brain Imaging Data Exchange (ABIDE)[5] dataset provides fMRI data for Autism Spectrum Disorder (ASD) research. This dataset, sourced from multiple international locations, aims to elucidate ASD neural underpinnings and identify diagnostic biomarkers. Participants are categorized into ASD or control groups. It includes fMRI data from 1118 subjects aged 8 to 40 years, engaged in social cognition tasks. Of these, 537 (48%) are diagnosed with ASD and 161 (14.5%) are female. A detailed summary of the datasets is presented in Table I.

2) Baselines: In this paper, we performed a detailed assessment of BrainGT by benchmarking it against a comprehensive suite of cutting-edge techniques, categorized into four types of groups: conventional Machine Learning (ML) algorithms including Support Vector Machines (SVM) [45], Random Forest, and Multi-Layer Perceptron (MLP) [46]; Graph Learning (GL) approaches such as Graph Neural Network (GNN) [47], Graph Isomorphism Network (GIN) [48], and Graph Attention Networks (GAT) [49]; Graph Transformer techniques like SAN [22], Graphormer [42], BRAINNETTF [14]; and specialized Brain Network Models, specifically BrainNetCNN [50], BrainGNN[10], and BrainGB [6]. These baselines were strategically selected to reflect their proven effectiveness in graph analysis and their applicability to brain network research. ML algorithms were utilized to process graph theory features derived from functional brain networks, whereas GL and Graph Transformer approaches were used directly on the networks, exploiting their inherent connectivity data. Brain Network Models were executed according to their original protocols, tailored for dense network data. To ensure consistency and fairness in our comparative evaluation, all methods were applied to dense functional brain networks generated using the BrainGB pipeline. ML algorithms analyzed extracted graph theory features, GL and Graph Transformer approaches

[3] http://adni.loni.usc.edu/data-samples/access-data

[4] https://www.ppmi-info.org/access-data-specimens/download-data

[5] http://fcon_1000.projects.nitrc.org/indi/abide

medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

handled the complete networks without prior feature extraction, and Brain Network Models adhered to their specified implementations, each fine-tuned for the specific analysis. This rigorous configuration methodology guarantees that our findings are reproducible and transparent in the application of each technique to the analyzed datasets.

3) Evaluation Metrics: The efficacy of the BrainGT framework was assessed using three principal metrics: accuracy, which reflects the overall performance; the F1 score, vital for medical diagnostics as it harmonizes precision and recall, particularly advantageous for datasets with class imbalances; and the Area Under the Curve (AUC), which encapsulates the model's performance across various thresholds, especially pertinent for imbalanced datasets. These metrics are crucial for the diagnosis of brain diseases, offering a holistic evaluation of the effectiveness of the model. To verify that performance enhancements were statistically significant and not attributable to random variations, a paired t-test was performed, comparing the performance of our model against other models on identical datasets, with a significance threshold of 0.05.

4) Implementation Details: The BrainGT framework, developed in PyTorch, comprises two principal components: the Multifunctional Brain Networks Construction module and the Multifunctional Brain Graph Transformers module. The former module processes fMRI data through atlas registration, ROI extraction and assignment to functional networks, functional connectivity estimation, and construction of multifunctional brain network datasets. The latter module embeds these networks, implements positional encoding, utilizes network-specific and cross-network encoders, and adaptively fuses the output for classification purposes. Data were partitioned into training, validation, and testing subsets with proportions of 70%, 15%, and 15%, respectively. Training was performed using the Adam optimizer, configured with a learning rate of 0.001, a batch size of 32, and stopping criteria set at 100 epochs. Hyperparameter tuning was performed using a grid search approach. All computational experiments were carried out on a workstation equipped with an NVIDIA 4090 RTX Ti 16 GB GPU.

5) Computational Complexity: computational complexity of BrainGT is a crucial multifaceted aspect for its application in neuroimaging analysis. preprocessing and network construction stages involve linear operations with respect to the number of voxels V and the regions R, resulting in a complexity of $O(V \times R)$. functional connectivity estimation that includes the pairwise correlation analysis across N nodes introduces a quadratic complexity of $O(N^2)$. transformer model which incorporates self-attention and cross-attention mechanisms in functional networks F with feature dimensionality D further contributes to the scaling of complexity as $O(F \times N^2 \times D)$. Finally, adaptive fusion and classification steps, depending on the architecture of the MLP classifier, add a complexity of $O(F \times D)$. In general, the total computational complexity of BrainGT can be approximated by $O(V \times R + F \times N^2 \times D + F \times D)$.

B. Performance of BrainGT

BrainGT framework has undergone a comprehensive evaluation using three distinct datasets: ADNI PPMI and ABIDE to determine its effectiveness in the diagnosis of various neurological disorders. Empirical results indicate significant promise. Specifically for the ADNI dataset that includes subjects with Alzheimer's disease (AD) and cognitively normal (CN) BrainGT achieved an accuracy of 72.76% an F1 score of 66. 72% and an AUC of 76. 37%, which affirms its precision and reliability. In the PPMI data set that focuses on the framework of Parkinson's Disease (PD) achieved an accuracy of 78. 89%, an F1 score of 65.82% and an AUC of 6868. 14% demonstrating its ability to discern complex patterns within PD. Furthermore, in the ABIDE dataset that refers to Autism Spectrum Disorder (ASD) in various age groups, BrainGT exhibited an accuracy of 76. 4%, an F1 score of 70.0% and an AUC of 78. 7%, which supports its robust performance in detecting subtle manifestations of ASD.

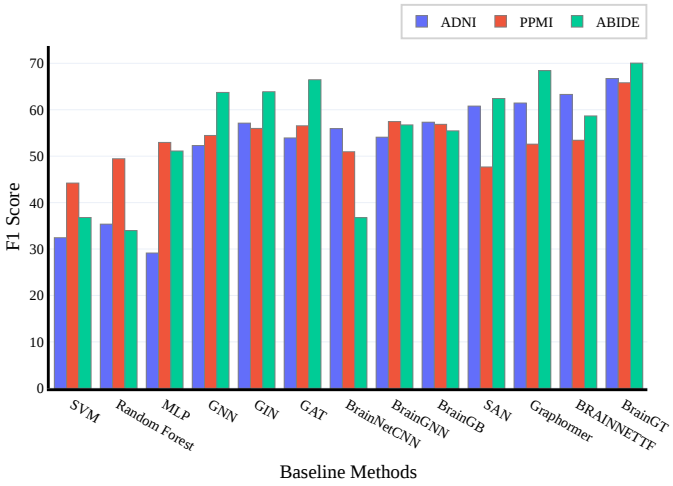


Fig. 3. Comparison of our method with baseline methods using the F1 score.

C. Comparison with Baseline Methods

The performance of BrainGT is assessed through a comparative analysis with established baseline methods in different categories of models. The detailed results are presented in Table II. BrainGT outperforms traditional machine learning techniques such as SVM, Random Forest, and MLP, commonly used in medical diagnostics. For example, in the ADNI dataset, BrainGT achieves an accuracy of 72. 76%, exceeding the accuracy of 61. 90% of MLP, the traditional method that performs the best. This discrepancy in performance is primarily due to the traditional methods' limitations in capturing complex patterns and managing the high dimensionality and interconnectedness of brain network data. The comparison with baseline methods on the F1 score is shown in Fig. 3.

Graph Learning Models such as GNN, GIN, and GAT are tailored for processing graph-structured data. These models generally surpass traditional machine learning techniques in performance metrics. However, BrainGT demonstrates superior robustness, achieving higher accuracy and F1 scores. For example, on the PPMI dataset, BrainGT achieves an F1 score

medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

TABLE II
COMPARATIVE PERFORMANCE ANALYSIS (%). BRAINGT DEMONSTRATES STATISTICALLY SIGNIFICANT IMPROVEMENTS OVER BASELINE MODELS, CONFIRMED BY T-TESTS (P-VALUE < 0.05).

Type	Method	ADNI			PPMI			ABIDE		
		Accuracy	F1	AUC	Accuracy	F1	AUC	Accuracy	F1	AUC
Machine Learning	SVM	60.34±0.17	32.44±0.17	62.58±4.80	74.09±0.24	44.21±0.08	49.21±1.48	59.45±2.31	36.79±24.03	62.41±3.52
	Random Forest	60.51±2.85	35.36±2.86	63.38±3.50	77.02±1.97	49.45±3.51	58.51±0.11	51.25±5.44	33.97±34.22	55.37±1.94
	MLP	61.90±1.22	29.14±0.57	65.74±2.37	77.99±2.78	52.96±4.52	59.65±6.80	60.72±12.69	51.12±40.97	66.72±3.87
Graph Learning	GNN	65.85±6.42	52.29±5.15	69.36±5.09	78.41±1.62	54.46±3.08	63.24±3.77	63.63±1.98	63.73±8.34	69.75±3.35
	GIN	64.63±2.35	57.12±1.81	65.32±1.42	79.34±2.21	55.98±3.08	64.15±2.67	67.87±2.75	63.89±9.78	74.93±2.43
	GAT	66.62±4.85	53.93±4.01	72.17±3.23	78.42±3.53	56.55±6.48	63.07±7.75	65.22±1.23	66.47±11.47	76.44±1.25
Brain Network Models	BrainNetCNN	64.93±4.90	55.94±5.42	67.94±5.42	77.30±1.52	50.96±4.20	63.93±4.89	59.43±2.35	36.75±24.05	62.4±3.54
	BrainGNN	67.09±2.77	54.09±2.77	73.36±4.26	79.52±6.23	57.45±6.28	64.09±3.25	69.34±6.37	56.74±34.04	72.42±3.71
	BrainGB	65.97±0.13	57.32±0.06	70.00±0.20	78.91±3.61	56.86±2.68	63.14±5.37	68.97±8.21	55.46±15.08	70.59±6.34
Graph Transformers	SAN	67.46±0.82	60.79±0.83	70.18±1.16	79.39±0.52	47.68±3.12	61.29±2.11	67.43±1.73	62.42±7.85	73.58±5.42
	Graphormer	68.92±1.29	61.45±1.28	72.16±1.41	78.29±2.58	52.59±2.1	65.35±4.71	69.41±5.31	68.43±7.84	76.52±5.48
	BRAINNETTF	70.42±011	63.30±2.08	73.33±0.78	**79.63±1.92**	53.43±4.6	64.23±2.79	70.24±4.94	58.67±16.07	73.25±4.57
Ours	BrainGT	**72.76± 1.2**	**66.72±1.17**	**76.37±1.46**	78.89±0.26	**65.82±6.43**	**68.14±5.01**	**76.47±1.07**	**70.05±1.24**	**78.73±22.34**

of 65.82%, which is significantly higher than GIN's 55.98%. This enhanced performance can be attributed to BrainGT's integration of both structural and functional brain atlases in network construction, a feature absent in the aforementioned models. Moreover, BrainGT's sophisticated attention mechanisms facilitate a deeper understanding of node relevance and feature extraction.

Graph Transformers such as SAN, Graphormer, and BRAINNETTF represent the latest advancements in graph neural networks by integrating attention mechanisms to enhance performance on graph structured data. BrainGT demonstrates a significant improvement in performance metrics across datasets compared to other Graph Transformers. Specifically, on the ABIDE dataset, BrainGT achieves an AUC of 78.7%, markedly higher than BRAINNETTF's 73.2%. The architecture of BrainGT, incorporating both self-attention and cross-attention mechanisms, facilitates more effective feature extraction from complex brain networks. This distinct design element sets BrainGT apart from other Graph Transformers, particularly in handling densely connected brain network data.

Brain network models, including BrainNetCNN, BrainGNN, and BrainGB, are specialized tools for brain network analysis. In particular, BrainGT has shown superior performance, as evidenced by its improved accuracy in the ADNI dataset, surpassing BrainGNN by 5.67%. This improvement can be attributed to BrainGT's adaptive fusion technique, which is not present in the other models. This technique synergistically integrates features across various functional brain networks, thereby enhancing the capabilities for disease diagnosis.

D. Ablation Study

In ablation study of the BrainGT framework, we methodically deconstructed the model by sequentially removing its key components to create a series of progressively simplified models. Subsequently, these ablated models were subjected to rigorous testing using ADNI PPMI and ABIDE datasets to assess the impact of each component on the diagnostic accuracy of the framework for brain disorders. study started with the complete BrainGT model which includes all components and systematically evaluated the effects of excluding multifunctional brain networks network-wise encoders cross-network encoders and adaptive fusion component. most simplified model devoid of all specialized components was utilized as a baseline for comparative analysis. Performance metrics for each variant were computed and contrasted with both this baseline and fully integrated BrainGT framework, thereby elucidating the contribution of each individual component to overall efficacy of the model. The results of the ablation study are presented in Table III.

The findings indicate a significant decline in performance with sequential removal of components that highlight their collective importance. The full model (none) exhibits the highest performance in all metrics, emphasizing the synergistic effect of the combined components. In particular, removal of multifunctional brain networks results in a marked decrease in performance, underscoring its critical role in managing dense connectivity of brain networks. Similarly, network-wise and cross-network encoders are crucial for capturing intra- and inter-network interactions. adaptive fusion component is also essential for the effective integration of the extracted features. most substantial decline in performance is observed when all specialized components are ablated (All Components), reaffirming the necessity of each element in BrainGT framework for accurate diagnosis of brain diseases.

E. Parameter Analysis

In an effort to optimize the BrainGT model for diagnosing brain disorders, a thorough parameter analysis was carried out, focusing on two main variables: the number of functional networks, denoted by n, and the number of layers within the encoders, represented by l. The analysis began with an exploration of the parameter space, in which n was varied from 1 to 17 to encompass the range of predefined functional networks within the Yeo2011 atlases. Simultaneously, l was adjusted from 1 to 20 to determine the optimal depth of the network-wise and cross-network encoders. The experiments were carried out using three datasets: ADNI, PPMI, and ABIDE. The initial parameter settings were derived from preliminary trials, serving as a baseline for the analysis. As n and

medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

TABLE III
PERFORMANCE IMPACT OF COMPONENT ABLATION ON BRAINGT FRAMEWORK: A COMPARATIVE ANALYSIS ACROSS ADNI, PPMI, AND ABIDE DATASETS

Component Ablated	ADNI			PPMI			ABIDE		
	Accuracy	F1-Score	AUC	Accuracy	F1-Score	AUC	Accuracy	F1-Score	AUC
None (Full BrainGT)	72.76	66.72	76.37	78.89	65.82	68.14	76.47	70.05	78.73
Multifunctional brain networks	68.62	63.43	73.17	74.35	62.66	64.63	73.21	65.89	75.31
Network-wise Encoders	65.48	59.99	68.71	70.97	59.23	61.31	68.81	63.14	70.83
Cross-network Encoders	70.16	65.05	74.08	77.13	63.19	66.21	73.97	68.25	75.71
Adaptive Fusion	71.30	66.24	73.16	78.10	62.85	66.78	74.18	69.35	77.15
All Components	58.21	54.04	60.33	63.11	53.31	53.83	61.18	56.74	62.20

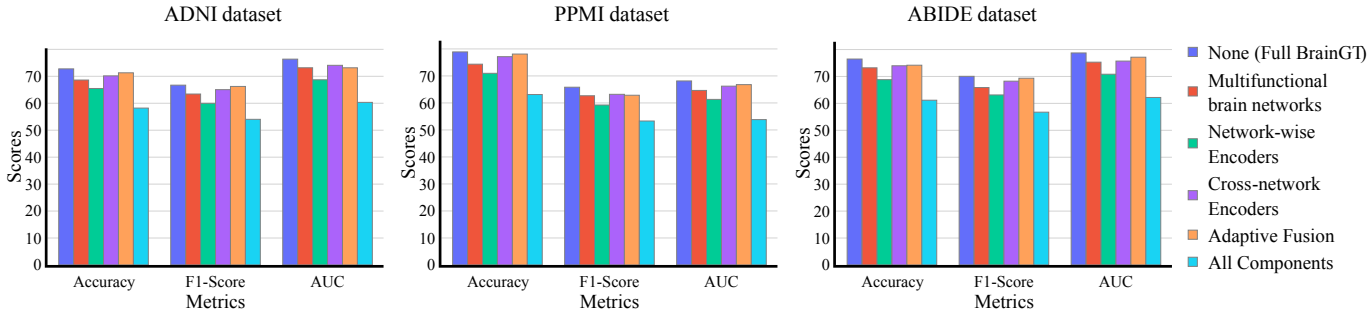


Fig. 4. Ablation studies of BrainGT.

l were systematically varied, the performance of the model was meticulously recorded, with particular attention to accuracy and F1-score metrics. The results were visualized through performance curves, providing an intuitive understanding of the impact of parameter variations on the model's efficacy.

Our findings indicate that the relationship between the number of functional networks and the model's performance is non-linear, with an optimal range of functional networks beyond which performance improvements become negligible. Similarly, increasing the number of layers exhibits diminishing returns, suggesting an optimal architecture that balances complexity with computational efficiency. For the ADNI dataset, the optimal parameters were identified as $n = 11$ and $l = 2$. The PPMI and ABIDE datasets yielded optimal values of $n = 9, l = 3$ and $n = 14, l = 2$, respectively, with the selection criteria based on achieving the highest combined accuracy and F1-score. The parameters n and l are critical to the performance of the BrainGT model, with their optimal values varying across different datasets and brain diseases. Although these results may represent a locally optimal solution, they underscore the importance of parameter tuning in computational models for neuroscience. Recognizing the limitations of our study, we recommend further exploration of the parameter space to enhance the model's diagnostic capabilities more comprehensively.

F. Discussion

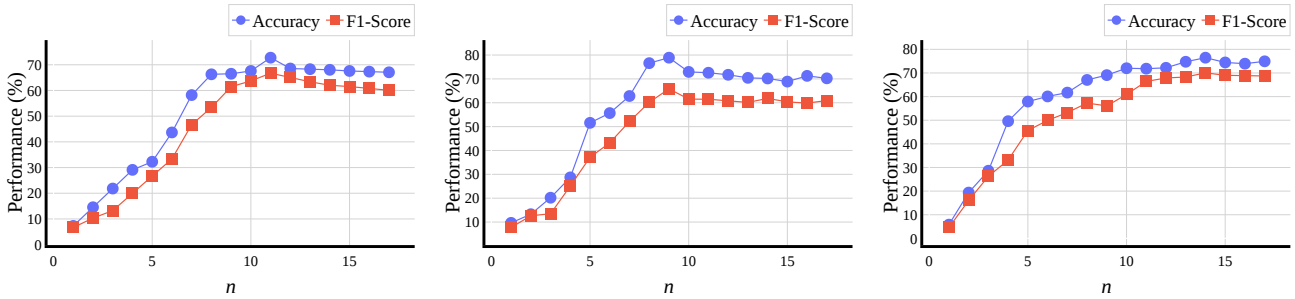
Our proposed BrainGT shows its effectiveness in diagnosing brain diseases. Using a multifunctional brain network construction method and attention mechanisms, BrainGT outperforms traditional techniques in precision F1, AUC score, and paired t-test results. To validate these findings and provide strong evidence to support our hypothesis, we performed a comparative baseline analysis, showing that BrainGT consistently outperforms existing neuroimaging methods using a two-sample t-test. Ablation studies confirmed the importance of each component (functional network segmentation, self-attention, and cross-attention) by observing performance drops when removed. Parameter analysis demonstrated the model's robustness to hyperparameter changes. Overall, BrainGT captures intricate brain connectivity patterns essential for identifying neurological disorders. Its superior performance suggests that it could become a valuable tool for clinicians and researchers, potentially leading to earlier and more precise diagnoses of conditions such as Alzheimer's, Parkinson's and autism spectrum disorder, allowing timely interventions and improving patient outcomes.

However, it is imperative to acknowledge the constraints inherent in our study. Despite the promising capabilities demonstrated by BrainGT, the current research has been conducted on a limited array of datasets. The generalizability of our findings to other datasets and populations remains to be substantiated. Additionally, the computational demands of BrainGT may pose significant challenges for its integration into clinical practice. Future investigations should aim to address these limitations to further substantiate the efficacy of BrainGT.

VI. CONCLUSION

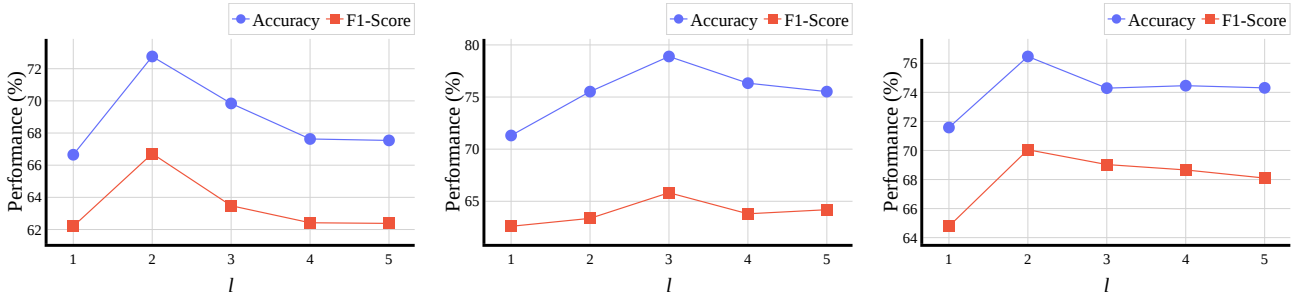
This study introduces BrainGT an innovative multifunctional brain graph transformer model that represents a substantial advancement in the domain of neuroimaging and diagnosis of brain disorders. In contrast to conventional methodologies that depend on densely connected brain networks, BrainGT constructs multifunctional networks that emphasize salient features and connections, thereby offering a more precise depiction of brain activity. Incorporation of self-attention and cross-attention mechanisms facilitates a detailed comprehension of both intra- and inter-network dynamics which is essential for

medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



(a) ADNI Dataset. (b) PPMI Dataset. (c) ABIDE Dataset.

Fig. 5. The accuracy and F1-score of BrainGT w.r.t. different n values.



(a) ADNI Dataset. (b) PPMI Dataset. (c) ABIDE Dataset.

Fig. 6. The accuracy and F1-score of BrainGT w.r.t. different l values.

accurate disease diagnosis. Our empirical evaluations on ADNI PPMI and ABIDE datasets have demonstrated BrainGT's superior performance in classification of brain diseases underscoring its potential as a valuable instrument for medical professionals.

Promising results obtained by BrainGT open several avenues for future research. One immediate direction is exploration of BrainGT's applicability to a wider range of brain disorders beyond those studied here. Additionally, further refinement of model's attention mechanisms could yield even more precise network representations potentially leading to breakthroughs in early diagnosis and intervention strategies. Another exciting prospect is the integration of BrainGT with real-time fMRI data processing, which could transform the model into a dynamic diagnostic system. Lastly, the development of a more interpretable version of BrainGT would not only enhance its clinical utility but also provide deeper insights into neural underpinnings of brain diseases.

References

[1] J. Ji, A. Zou, J. Liu, C. Yang, X. Zhang, and Y. Song, "A survey on brain effective connectivity network learning," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 34, no. 4, pp. 1879–1899, 2023. DOI: 10.1109/TNNLS.2021.3106299.

[2] Q. Yu, Y. Du, J. Chen, *et al.*, "Application of graph theory to assess static and dynamic brain connectivity: Approaches for building brain graphs," *Proc. IEEE*, vol. 106, no. 5, pp. 886–906, 2018. DOI: 10.1109/JPROC.2018.2825200.

[3] M. R. Ahmed, Y. Zhang, Z. Feng, B. Lo, O. T. Inan, and H. Liao, "Neuroimaging and machine learning for dementia diagnosis: Recent advancements and future prospects," *IEEE Rev. Biomed. Eng.*, vol. 12, pp. 19–33, 2019. DOI: 10.1109/RBME.2018.2886237.

[4] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nat. Rev. Neurosci.*, vol. 10, no. 3, pp. 186–198, Feb. 2009. DOI: 10.1038/nrn2575.

[5] J. C. Reijneveld, S. C. Ponten, H. W. Berendse, and C. J. Stam, "The application of graph theoretical analysis to complex networks in the brain," *Clin. Neurophysiol.*, vol. 118, no. 11, pp. 2317–2331, Nov. 2007. DOI: 10.1016/j.clinph.2007.08.010.

[6] H. Cui, W. Dai, Y. Zhu, *et al.*, "Braingb: A benchmark for brain network analysis with graph neural networks," *IEEE Trans. Med. Imaging*, vol. 42, no. 2, pp. 493–506, 2023. DOI: 10.1109/TMI.2022.3218745.

[7] C. Peng, M. Liu, C. Meng, S. Yu, and F. Xia, "Adaptive brain network augmentation based on group-aware graph learning," in *International Conference on Learning Representations (ICLR)*, Vienna, Austria, May 2024.

[8] W. Yin, L. Li, and F. Wu, "Deep learning for brain disorder diagnosis based on fmri images," *Neurocomputing*, vol. 469, pp. 332–345, 2022. DOI: 10.1016/J.NEUCOM.2020.05.113.

[9] X. Wang, Y. Chu, Q. Wang, *et al.*, "Unsupervised contrastive graph learning for ¡scp¿resting-state¡/scp¿ functional ¡scp¿mri¡/scp¿ analysis and brain disorder detection," *Hum. Brain Mapp.*, vol. 44, no. 17, pp. 5672–5692, Sep. 2023. DOI: 10.1002/hbm.26469.

[10] X. Li, Y. Zhou, N. Dvornek, *et al.*, "Braingnn: Interpretable brain graph neural network for fmri analysis," *Med. Image Anal.*, vol. 74, p. 102 233, Dec. 2021. DOI: 10.1016/j.media.2021.102233.

[11] W. Dai, H. Cui, X. Kan, Y. Guo, S. van Rooij, and C. Yang, "Transformer-based hierarchical clustering for brain network analysis," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, Cartagena, Colombia, Apr. 2023, pp. 1–5. DOI: 10.1109/ISBI53787.2023.10230606.

[12] Y. Ma, Q. Wang, L. Cao, *et al.*, "Multi-scale dynamic graph learning for brain disorder detection with functional mri," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3501–3512, 2023. DOI: 10.1109/TNSRE.2023.3309847.

[13] A. G. Alharthi and S. M. Alzahrani, "Do it the transformer way: A comprehensive review of brain and vision transformers for autism spectrum disorder diagnosis and classification," *Comput. Biol. Medicine*, vol. 167, p. 107 667, 2023. DOI: 10.1016/J.COMPBIOMED.2023.107667.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[14] 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)*, New Orleans, LA, USA, Dec. 2022. DOI: 10.48550/arxiv.2210.06681.

[15] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, I. Guyon, U. von Luxburg, S. Bengio, *et al.*, Eds., Long Beach, CA, USA, Dec. 2017, pp. 5998–6008.

[16] D. S. Bassett and O. Sporns, "Network neuroscience," *Nat. Neurosci.*, vol. 20, no. 3, pp. 353–364, Feb. 2017. DOI: 10.1038/nn.4502.

[17] X. Kong and P. S. Yu, "Brain network analysis: A data mining perspective," *SIGKDD Explor. Newsl.*, vol. 15, no. 2, pp. 30–38, Jun. 2014. DOI: 10.1145/2641190.2641196.

[18] W. Tong, Y.-X. Li, X.-Y. Zhao, *et al.*, "Fmri-based brain disease diagnosis: A graph network approach," *IEEE. Trans. Med. Robot. Bionics*, vol. 5, no. 2, pp. 312–322, 2023. DOI: 10.1109/TMRB.2023.3270481.

[19] P. Veličković, "Everything is connected: Graph neural networks," *Curr. Opin. Struc. Biol.*, vol. 79, p. 102 538, Apr. 2023. DOI: 10.1016/j.sbi.2023.102538.

[20] A. Said, R. G. Bayrak, T. Derr, *et al.*, "Neurograph: Benchmarks for graph machine learning in brain connectomics," in *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., New Orleans, LA, USA, Dec. 2023.

[21] L. Rampásek, M. Galkin, V. P. Dwivedi, A. T. Luu, G. Wolf, and D. Beaini, "Recipe for a general, powerful, scalable graph transformer," in *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., New Orleans, LA, USA, Nov. 2022.

[22] D. Kreuzer, D. Beaini, W. L. Hamilton, V. Létourneau, and P. Tossou, "Rethinking graph transformers with spectral attention," in *35th Conference on Neural Information Processing Systems (NeurIPS 2021)*, Virtual, Dec. 2021. DOI: 10.48550/arxiv.2106.03893.

[23] B. Min, H. Ross, E. Sulem, *et al.*, "Recent advances in natural language processing via large pre-trained language models: A survey," *ACM Computing Surveys*, vol. 56, no. 2, pp. 1–40, Sep. 2023. DOI: 10.1145/3605943.

[24] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM Computing Surveys*, vol. 54, no. 10s, pp. 1–41, Jan. 2022. DOI: 10.1145/3505244.

[25] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," *Expert Syst. Appl.*, vol. 223, p. 119 862, Aug. 2023. DOI: 10.1016/j.eswa.2023.119862.

[26] J. Yoo, T. Y. Kim, I. Joung, and S. O. Song, "Industrializing ai/ml during the end-to-end drug discovery process," *Curr. Opin. Struc. Biol.*, vol. 79, p. 102 528, Apr. 2023. DOI: 10.1016/j.sbi.2023.102528.

[27] R. Fang, Y. Li, X. Zhang, *et al.*, "Path-based heterogeneous brain transformer network for resting-state functional connectivity analysis," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, Vancouver, BC, Canada: Springer Nature Switzerland, Oct. 2023, pp. 328–337. DOI: 10.1007/978-3-031-43993-3_32.

[28] Q. Zuo, J. Hu, Y. Zhang, *et al.*, "Brain functional network generation using distribution-regularized adversarial graph autoencoder with transformer for dementia diagnosis," *Computer Modeling in Engineering & Sciences*, vol. 137, no. 3, pp. 2129–2147, 2023. DOI: 10.32604/cmes.2023.028732.

[29] H. Cai, Y. Gao, and M. Liu, "Graph transformer geometric learning of brain networks using multimodal mr images for brain age estimation," *IEEE Trans. Med. Imaging*, vol. 42, no. 2, pp. 456–466, 2023. DOI: 10.1109/TMI.2022.3222093.

[30] A. Bannadabhavi, S. Lee, W. Deng, R. Ying, and X. Li, "Community-aware transformer for autism prediction in fmri connectome," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, H. Greenspan, A. Madabhushi, P. Mousavi, *et al.*, Eds., Vancouver, BC, Canada: Springer Nature Switzerland, Oct. 2023, pp. 287–297.

[31] Z. Wang, J. Xin, Z. Wang, Y. Yao, Y. Zhao, and W. Qian, "Brain functional network modeling and analysis based on fmri: A systematic review," *Cogn. Neurodynamics*, vol. 15, no. 3, pp. 389–403, Aug. 2020. DOI: 10.1007/s11571-020-09630-5.

[32] F. Xia, K. Sun, S. Yu, *et al.*, "Graph learning: A survey," *IEEE Trans. Artif. Intell.*, vol. 2, no. 2, pp. 109–127, 2021. DOI: 10.1109/TAI.2021.3076021.

[33] Y. Rong, T. Xu, J. Huang, *et al.*, "Deep graph learning: Foundations, advances and applications," in *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, R. Gupta, Y. Liu, J. Tang, and B. A. Prakash, Eds., Virtual Event, CA, USA: ACM, Aug. 2020, pp. 3555–3556. DOI: 10.1145/3394486.3406474.

[34] R. Li, X. Yuan, M. Radfar, *et al.*, "Graph signal processing, graph neural network and graph learning on biological data: A systematic review," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 109–135, 2023. DOI: 10.1109/RBME.2021.3122522.

[35] C. Peng, M. Liu, C. Meng, S. Xue, K. Keogh, and F. Xia, "Stage-aware brain graph learning for alzheimer's disease," in *The 2024 IEEE Conference on Artificial Intelligence (CAI)*, Singapore, Jun. 2024, pp. 25–27.

[36] C. Liu, Y. Zhan, X. Ma, *et al.*, "Exploring sparsity in graph transformers," *Neural Networks*, vol. 174, p. 106 265, 2024. DOI: 10.1016/J.NEUNET.2024.106265.

[37] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI Open*, vol. 3, pp. 111–132, 2022. DOI: 10.1016/j.aiopen.2022.10.001.

[38] P. Quan, Y. Shi, M. Lei, J. Leng, T. Zhang, and L. Niu, "A brief review of receptive fields in graph convolutional networks," in *2019 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2019, Thessaloniki, Greece, October 14-17, 2019 - Companion Volume*, P. M. Barnaghi, G. Gottlob, D. Katsaros, *et al.*, Eds., Thessaloniki, Greece: ACM, Oct. 2019, pp. 106–110. DOI: 10.1145/3358695.3360934.

[39] W. Zhu, T. Wen, G. Song, L. Wang, and B. Zheng, "On structural expressive power of graph transformers," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ser. KDD '23, New York, NY, USA: Association for Computing Machinery, Aug. 2023, pp. 3628–3637. DOI: 10.1145/3580305.3599451.

[40] Y. Tay, M. Dehghani, D. Bahri, and D. Metzler, "Efficient transformers: A survey," *ACM Comput. Surv.*, vol. 55, no. 6, Dec. 2022. DOI: 10.1145/3530811.

[41] X. Ma, Q. Chen, Y. Wu, G. Song, L. Wang, and B. Zheng, "Rethinking structural encodings: Adaptive graph transformer for node classification task," in *Proceedings of the ACM Web Conference 2023*, ser. WWW '23, Austin, TX, USA: Association for Computing Machinery, Apr. 2023, pp. 533–544. DOI: 10.1145/3543507.3583464.

[42] C. Ying, T. Cai, S. Luo, *et al.*, "Do transformers really perform badly for graph representation?" In *35th Conference on Neural Information Processing Systems (NeurIPS 2021)*, Virtual, Dec. 2021. DOI: 10.48550/arxiv.2106.05234.

[43] D. Chen, L. O'Bray, and K. Borgwardt, "Structure-aware transformer for graph representation learning," in *Proceedings of the 39th International Conference on Machine Learning*, K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, Eds., ser. Proceedings of Machine Learning Research, vol. 162, Baltimore, MD, USA: PMLR, Jul. 2022, pp. 3469–3489.

[44] Y. Ma, W. Cui, J. Liu, Y. Guo, H. Chen, and Y. Li, "A multi-graph cross-attention-based region-aware feature fusion network using multi-template for brain disorder diagnosis," *IEEE Trans. Med. Imaging*, vol. 43, no. 3, pp. 1045–1059, 2024. DOI: 10.1109/TMI.2023.3327283.

[45] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995. DOI: 10.1007/bf00994018.

[46] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986. DOI: 10.1038/323533a0.

[47] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Networks*, vol. 20, no. 1, pp. 61–80, Jan. 2009. DOI: 10.1109/tnn.2008.2005605.

[48] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" In *2019 International Conference on Learning Representations (ICLR)*, New Orleans, LA, USA, May 2019.

[49] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *6th International Conference on Learning Representations (ICLR 2018)*, Vancouver, BC, Canada: OpenReview.net, Apr. 2018. DOI: 10.48550/arxiv.1710.10903.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.30.24312819; this version posted August 31, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[50] J. Kawahara, C. J. Brown, S. P. Miller, *et al.*, "Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, Feb. 2017. DOI: 10.1016/j.neuroimage.2016.09.046.