BRAIN NETWORKS AND INTELLIGENCE: A GRAPH NEURAL NETWORK BASED APPROACH TO RESTING STATE fMRI DATA

*Bishal Thapaliya [1,2]   Esra Akbas [1]   Jiayu Chen [1,2]   Raam Sapkota [1,2]   Bhaskar Ray [1,2]
Pranav Suresh [1,2]   Vince Calhoun [1,2,3]   Jingyu Liu [1,2]
[1]Georgia State University   [2]TReNDs Center   [3]Georgia Institute of Technology

October 29, 2024

ABSTRACT

Resting-state functional magnetic resonance imaging (rsfMRI) is a powerful tool for investigating the relationship between brain function and cognitive processes as it allows for the functional organization of the brain to be captured without relying on a specific task or stimuli. In this paper, we present a novel modeling architecture called BrainRGIN for predicting intelligence (fluid, crystallized and total intelligence) using graph neural networks on rsfMRI derived static functional network connectivity matrices. Extending from the existing graph convolution networks, our approach incorporates a clustering-based embedding and graph isomorphism network in the graph convolutional layer to reflect the nature of the brain sub-network organization and efficient network expression, in combination with TopK pooling and attention-based readout functions. We evaluated our proposed architecture on a large dataset, specifically the Adolescent Brain Cognitive Development Dataset, and demonstrated its effectiveness in predicting individual differences in intelligence. Our model achieved lower mean squared errors, and higher correlation scores than existing relevant graph architectures and other traditional machine learning models for all of the intelligence prediction tasks. The middle frontal gyrus exhibited a significant contribution to both fluid and crystallized intelligence, suggesting their pivotal role in these cognitive processes. Total composite scores identified a diverse set of brain regions to be relevant which underscores the complex nature of total intelligence.

Keywords Graph Neural Networks · Intelligence · Resting-State fMRI Data · Static Functional Connectivity (sFNC)

1 Introduction

Intelligence is a complex construct that comprises various cognitive processes. Researchers typically rely on a range of cognitive tests that measure different aspects of cognition and form specific measures for intelligence, such as fluid intelligence (the ability to reason and solve problems in novel situations)(Kyllonen and Kell [2017]), crystallized intelligence (the ability to use knowledge and experience to solve problems), and total intelligence (a composite measure of overall cognitive ability). There is always undamped human interest to reveal neural underpinnings of such intelligence and to predict individual intelligence differences. Although traditional MRI studies in the literature focused on structural brain measures for various phenotypes (Suresh et al. [2023], Ray et al. [2023]), rapidly growing studies have emerged to investigate the prediction of intelligence based on brain functional features (Ferguson et al. [2017], He et al. [2020], Dubois et al. [2018]). The review conducted by (Vieira et al. [2022]) further concluded that functional MRI (fMRI) has become the most employed modality for predicting intelligence, and resting-state fMRI (rs-fMRI) derived static functional connectivity (FC) was the most studied predictor. rs-fMRI measures spontaneous brain activity during rest through blood oxygen level-dependent (BOLD) signals in response to neuronal activity. FC is defined as the degree of temporal correlation between regions of the brain computed using time series of BOLD signals, and rs-fMRI FC provides a comprehensive view of the brain's intrinsic organization (Lee et al. [2012]). Indeed, FC between the default mode network and frontoparietal network have been validated to contribute to individual differences in

---

*Corresponding Author: Bishal Thapaliya, bthapaliya16@gmail.com.

cognitive ability (Hearne et al. [2016]).

While the majority of intelligence prediction approaches involve linear regression methods, some studies have applied non-linear approaches including polynomial kernel SVR (Wang et al. [2015]), and kernel ridge regression (He et al. [2020]) methods and deep neural networks (He et al. [2020], Fan et al. [2020], Li et al. [2023]). In recent years, Graph Neural Networks (GNNs) have gained great interest and have evolved rapidly for end-to-end graph learning applications. GNNs are considered the state-of-the-art deep learning methods for solving graph-structured data analysis problems, as they specify a neural network to fit into the graph structure with nodes and edges, and embed node features and edge features with structural information in the graph. Various studies have investigated the effectiveness of GNNs in different applications such as social networks, protein networks, and neurological biomarkers (Kim and Ye [2020], Kazi et al. [2023], Nandakumar et al. [2021]). Given the network-structured nature of the brain, modeling brain connectome via GNNs has been implemented. Most brain GNN studies utilize the FC graph from rs-fMRI (Ktena et al. [2018], Škoch et al. [2022], Wu et al. [2021], Ma et al. [2018]) and classify a particular phenotype of the subjects, such as gender (Arslan et al. [2018], Kazi et al. [2021]) or specific disease status (Kazi et al. [2021], Ma et al. [2018]), while its predictive power for intelligence is yet to be investigated.

Most GNNs assume that the nodes learn embeddings in an identical way throughout the whole graph, which is problematic for brain connectome due to the sub-network nature of the brain (Parente and Colosimo [2020]). Recently, BrainGNN (Li et al. [2021]) proposed a new GNN architecture that tackled this limitation by proposing a clustering-based embedding method in the graph convolutional layer, which allows nodes in different clusters (representing different brain networks) to learn embeddings differently. Inspired by BrainGNN, we present a novel GNN model, Brain ROI-aware Graph Isomorphism Networks (*BrainRGIN*), for intelligence prediction. Firstly, we make use of a graph isomorphism network (GIN) (Xu et al. [2018]) to improve the expressive power of GNNs, which is designed to approximate the power of the Weisfeiler-Lehman (WL) graph isomorphism test. Similar to BrainGNN, we also tackle the limitation of node identical learning mechanism by introducing cluster representation of the region of interest (ROI). By combining these two architectures, our model can effectively capture both local and global relationships between brain regions. Furthermore, we validate various aggregation and read-out functions including attention-based readout methods. As far as we are aware, this is the first study that uses graph neural networks to uncover the patterns of the brain for intelligence prediction using resting state fMRI data. We evaluate the performance of our proposed model on a large dataset and demonstrate its effectiveness in predicting individual differences in intelligence.

$$h_i^{(l)} = MLP^{(l)} \left(\left(1 + \epsilon^{(l)}\right) h_i^{(l-1)} + \sum_{j \in \mathcal{N}^{(l)}(i)} h_j^{(l-1)} \right) \tag{1}$$

where $\epsilon^{(l)}$ is a trainable parameter that helps to avoid the over-smoothing problem, and $MLP^{(l)}$ is a 2-layered multi-layer perception to update the final feature representation for the i-th node in the l-th layer. The output of the graph is then computed as:

$$h_G = READOUT \left(h_i^{(l)}{}_{v \in G} \right) \tag{2}$$

where h_G is the final representation of the graph and $READOUT$ is a permutation invariant function that aggregates the node representations to produce a single output vector. GINs have been shown to outperform other state-of-the-art GNNs on various graph classification tasks, demonstrating their power and effectiveness in learning representations of graph-structured data (Xu et al. [2018]).

1.1 Comparision with existing deep learning based architectures for Brain Network Analysis

In recent times, there has been a growing focus on extending the applicability of Graph Neural Network (GNN)-based models to the realm of brain network analysis. Notably, BrainGNN (Li et al. [2021]) innovates by crafting Region of Interest (ROI)-aware GNNs to harness functional information within the brain and employing a specialized pooling operator to identify pivotal nodes. FBNetGen (Kan et al. [2022a]) delves into the trainable generation of brain networks while investigating their interpretability for downstream tasks. Graph Transformer Dwivedi and Bresson [2020] uses static FC as features and applies multi-head attention with residual connections and normalization blocks to perform downstream tasks. A much more recent model, Brain Network Transformer (BNT) (Kan et al. [2022b]) effectively used transformer-scaled dot product attention and introduced an orthonormal clustering readout function that leads to cluster-aware node embeddings. On the other hand, IBGNN (Cui et al. [2022]) initializes a mask based on edges and

refines it with sparse control during training to identify top connections and regions for a specific task, aiming to create a subnetwork capable of predicting information as effectively as the original network. Unlike IGBNN which pools connections (edges) based on the sparse mask, we aim to select top salient ROIs by pooling nodes. While these models seem to be very effective, we direct our focus on designing models to harness the power of isomorphism networks during graph-based convolution, and attention-based readouts for intelligence prediction. As for the conceptualization, our model relates very closely to the BNT, which uses a clustering-based embedding approach during the readout after scaled dot product attention-based node feature updates. Our model, however, implements clustering-based node feature updates during the isomorphism-based graph convolution part, followed by scaled dot product attention for the readout function. The complete idea is further elucidated in Section 2.
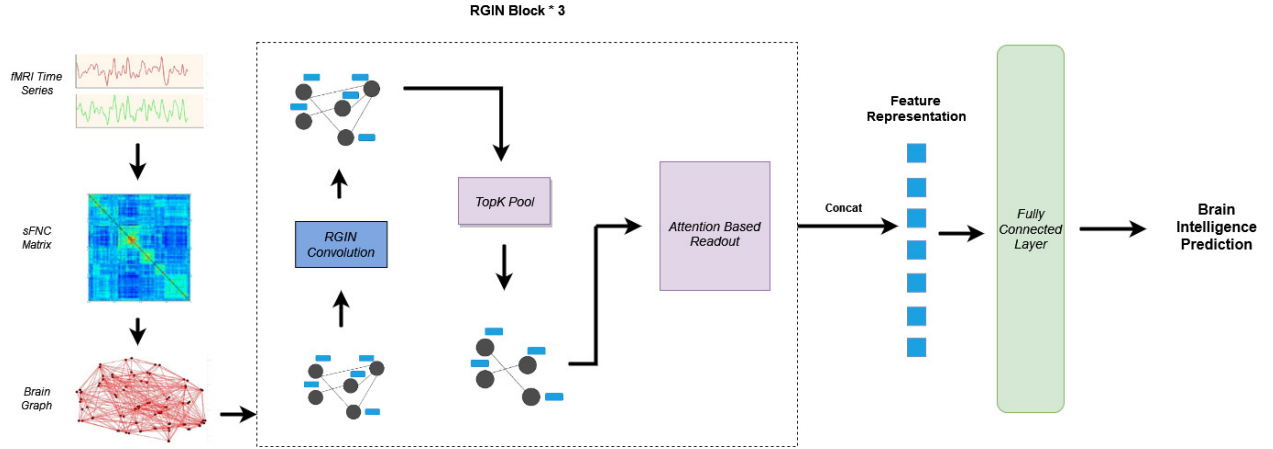
2 Proposed Architecture

Figure 1: Overall Architecture of *BrainRGIN* . The static FNC matrix is extracted from a resting state fMRI time series data. Three blocks of BrainRGIN are used with attention-based readout functions followed by a fully connected layer for prediction.

The architecture presented in Fig. 1 depicts the overall working mechanism of the *BrainRGIN* . As the initial step, we create a functional connectivity graph using functional relations between ROIs of the brain. After creating the FC graph, a three-layer RGIN (ROI-aware graph isomorphism network) block is established, where each block encompasses a sequence of RGIN convolutions, succeeded by a TopK Pooling layer. The output of RGIN is subject to an Attention-Based Readout block either with a Squeeze-Excitation Readout (SERO) function or a Graph Attention Readout (GARO) function. Finally, a single-layer fully connected neural network serves to compute the intelligence score.

2.1 Functional connectivity graph creation

The initial phase involves the processing of input from fMRI time series data to create a static FC matrix or static functional network connectivity (sFNC) matrix. The difference between FC and FNC lies in how ROIs are generated: regions selected based on an atlas lead to an FC matrix, while regions generated by independent component analyses (ICA) lead to an FNC matrix. The brain is spatially divided into N ROIs. In this study, we employed ICA ROIs resultant static FNC matrix. Using ROIs time series, Pearson's correlation between ROIs is used to create a FNC matrix. The resulting FNC matrix is used to construct an undirected graph denoted by the tuple $G = (V, E)$ with an adjacency matrix $A \in R^{N*N}$. Each ROIs represents graph nodes indexed by the set $V = 1, ..., N$. The input node features, x_i, are formed using Pearson's correlation coefficients between node i and all other nodes. An edge set E represents the functional connections between ROIs with each edge e_{ij} linking two nodes $(i, j) \in E$.

The Adjacency matrix $A_k \in R^{N \times N}$ is formed as in Equation 3, where $e_{i,j}$ is the thresholded element of FNC matrix to achieve either fully connected or sparse graph.

$$A_{i,j}^k = \begin{cases} 0, & i = j \\ e_{i,j}, & \text{otherwise} \end{cases} \qquad (3)$$

3

2.1.1 RGIN Convolution

We present RGIN convolution, an approach to effectively capture both node and edge features in the graph by merging RGCN and GIN. Similar to BrainGNN, within the context of brain function, the type of edge relation reflects the different functional clustering of brain networks. In other words, nodes in the brain are divided into functional clusters, and different brain clusters form unique types of edge relations. Without fixing cluster assignments for ROIs, the proposed model also learns cluster formation such that a group of nodes in the same cluster works together (update together) to achieve the best performance for the task. GIN is incorporated for learning node embeddings for its better graph discriminative expression.

In RGIN, we replace the aggregation function of GIN with that of RGCN, allowing the model to learn different mechanisms for different clusters. The motivation behind using this design is twofold: 1) GIN utilizes Multilayer Perceptrons (MLPs) to learn the injective functions to adapt to the complexity of brain connectivity. 2) RGCN is able to model the sub-network clustering nature of brain connectivity. The forward propagation function of RGIN is defined as follows:

$$h_i^{(l)} = \text{MLP}^{(l)} \left((1 + \epsilon^{(l)}) \cdot W_i^{(l)} \cdot h_i^{(l-1)} + \sum_{j \in N_{(i)}^{(l)}} W_j^{(l)} \cdot e_{i,j}^{(l-1)} \cdot h_j^{(l-1)} \right) \tag{4}$$

While in regular GNN models, including GIN and RGCN, W_i and W_j are learnable weight matrices, we define them as the function with parameters subject to cluster assignment of nodes to make it ROI aware convolution layer. Then, RGIN is trained in an end-to-end fashion with a 2-layered MLP. $\epsilon^{(l)}$ here is a learnable parameter that determines the importance of node i compared to its neighbors j. In addition, multiplying node features by the edge weights, similar to the graph convolution operations in (Gong and Cheng [2018]), created a greater impact for neighbors connected by stronger edges.

As given in Equation 5, $W_i^{(l)}$ is defined as a function to incorporate the position encoding r_i, with parameters θ_1 and θ_2 defined to learn the clustering specific parameter for nodes. ReLU is the rectified linear unit activation function. and $\mathbf{b}^{(l)}$ is the bias term.

$$\mathbf{W}_i^{(l)} = \theta_2^{(l)} \cdot \text{relu}(\theta_1^{(l)} \mathbf{r}_i) + \mathbf{b}^{(l)} \tag{5}$$

In the same way as BrainGNN (Li et al. [2021]), we allow RGIN to learn different update learning parameters for different clusters conditioned on the ROIs. We represent each node's location information by a vector $\mathbf{r}_i$, using one-hot encoding instead of coordinates, assuming that the ROIs are aligned in the same order for all the brain graphs.

We assume that $\theta_1^{(l)} = [\alpha_1^{(l}, \ldots, \alpha_{N^{(l)}}^{(l}]$, where $N^{(l)}$ is the number of ROIs in the lth layer, $\alpha_i^{(l)} = [\alpha_{i1}^{(l)}, \ldots, \alpha_{iK(l)}^{(l)}] \in \mathbb{R}^{K(l)}, \forall i \in \{1, \ldots, N^{(l)}\}$, where $K(l)$ is the number of clusters in $N^{(l)}$ ROIs, and $\alpha_i^{(l)}$ is the non-negative assignment scores of ROI i to clusters. Assume $\theta_2^{(l)} = [\beta_1^{(l)}, \ldots, \beta_{K(l)}^{(l)}]$ with $\beta_u^{(l)} \in \mathbb{R}^{d(l+1) \times d(l)}, \forall u \in \{1, \ldots, K(l)\}$. $\beta_u^{(l)}$ is a basis matrix. Then, we can rewrite equation (5) as:

$$\mathbf{W}_i^{(l)} = \sum_{u=1}^{K(l)} \alpha_{iu}^{(l)} \beta_u^{(l)} + \mathbf{b}^{(l)} \tag{6}$$

This reduces the number of learnable parameters while still allowing a separate embedding kernel for each ROI. The final forward propagation function of RGIN Convolution can be written as:

$$h_i^{(l)} = \text{MLP}^{(l)} \left((1 + \epsilon^{(l)}) \cdot \left(\sum_{u=1}^{K(l)} \alpha_{iu}^{(l)} \beta_u^{(l)} + \mathbf{b}^{(l)} \right) \cdot h_i^{(l-1)} \right.$$
$$\left. + \sum_{j \in N_{(i)}^{(l)}} h_j^{(l-1)} \cdot \left(\sum_{u=1}^{K(l)} \alpha_{ju}^{(l)} \beta_u^{(l)} + \mathbf{b}^{(l)} \right) \cdot e_{i,j}^{(l-1)} \right) \tag{7}$$

2.1.2 Pooling Layers

Via RGIN convolution, node-wise representations are generated. However, for the prediction task discussed in this paper, which requires graph-level prediction instead of node-level, these node-wise representations must be collected or

pooled together. We apply ROI-aware TopK Pooling as it improves interpretability for rsfMRI data, by keeping the most indicative ROIs and removing the noisy and uninformative nodes (Li et al. [2021]), along with enforcing sparsity on the network. To select the most indicative ROIs, the choice of which nodes to drop is determined based on the scores of nodes obtained by projecting the node features to 1D via a learnable projection vector $\omega^{(l)} \in \mathbb{R}^{d^{(l)}}$. The pooled graph $(V^{(l+1)}, E^{(l+1)})$ is computed as follows:

1. Calculate the scores of nodes $s^{(l)}$ with node feature matrix as follows:

$$s^{(l)} = \frac{H^{(l)} w^{(l)}}{\|w^{(l)}\|_2} \tag{8}$$

2. Normalize the score $s^{(l)}$ by subtracting its mean and dividing by its standard deviation, yielding final scores $\tilde{s}^{(l)}$:

$$\tilde{s}^{(l)} = \frac{(s^{(l)} - \mu(s^{(l)}))}{\sigma(s^{(l)})} \tag{9}$$

3. Find the top k elements in the normalized score vector $\tilde{s}^{(l)}$ with k- largest values and get their indices as i:

$$i = \text{topk}(\tilde{s}^{(l)}, k) \tag{10}$$

4. Create the pooled node features $H^{(l+1)}$ by element-wise multiplying the original node features $H^{(l)}$ with the sigmoid of the normalized scores and selecting only the elements indexed by i:

$$H^{(l+1)} = (H^{(l)} \odot \text{sigmoid}(\tilde{s}^{(l)}))_{i:} \tag{11}$$

5. Create new adjacency matrix as the new pooled graph $E^{(l+1)}$ by returning the edges between selected nodes i from previous edge matrix $E^{(l)}$

$$E^{(l+1)} = E_{i,i}^{(l)}$$

In this description, $\| \cdot \|_2$ represents the L2 norm, μ and σ are functions that calculate the mean and standard deviation of a vector, TopK identifies the indices of the largest k elements in a vector, $\odot$ denotes element-wise multiplication, and $(\cdot)_{i,:}$ selects elements in the ith row and all columns.

2.1.3   Readout

The conventional READOUT function of GNN can be thought of as a fixed decoder that decodes whole-graph features from the node features with no learnable parameters. We address this issue by incorporating attention to the READOUT function. The attention here refers to the scaling coefficient across the nodes learned by the model. Therefore, we make use of attention-based readout functions called GARO and SERO (Hu et al. [2017]) as proposed in STAGIN (Kim et al. [2021]).

GARO function is inspired by the key-query embedding-based attention mechanism of the Transformer (Vaswani et al. [2017]). The key embedding $\mathbf{K} \in \mathbb{R}^{D \times N}$ is computed from node feature matrix $\mathbf{H} \in \mathbb{R}^{D \times N}$, D is the dimension of node feature and N is a number of nodes. The query embedding $\mathbf{q} \in \mathbb{R}^D$ is computed from the vector of unattended graph representation $\mathbf{H}_{\phi mean}$, where $\phi mean$ represents the global average of node feature matrix H. $zspace$ is the final attention vector.

$$\mathbf{K} = \mathbf{W}_{key}\mathbf{H}, \tag{12}$$

$$\mathbf{q} = \mathbf{W}_{query}\mathbf{H}\phi_{mean}, \tag{13}$$

$$\mathbf{z}space = \sigma\left(\frac{\mathbf{q}^\top \mathbf{K}}{\sqrt{D}}\right), \tag{14}$$

where $\mathbf{W}_{key} \in \mathbb{R}^{D \times D}$, $\mathbf{W}_{query} \in \mathbb{R}^{D \times D}$ are learnable key-query parameter matrices, σ is the sigmoid function. The GARO function uses the attended key matrix to compute the final graph representation as follows:

$$\mathbf{h}_{\mathrm{GARO}} = \mathbf{H}z_{space}. \tag{15}$$

The GARO function allows the graph neural network to effectively capture dynamic brain connectome by attending to the important nodes.

On the other hand, the SERO function applies the squeeze and excitation operations to enhance the relevant features. Originally, the squeeze excitation module was shown to increase the performance of CNN models (Roy et al. [2019]). As we have a fixed number of nodes similar to (Kim et al. [2021]), we applied the SERO module. It is important to note that the SERO activation does not scale the channel dimension, but the node dimension. The squeeze operation computes the global information of the node features using the global average pooling, followed by the excitation operation which models the dependencies between different nodes. Specifically, we first apply a linear transformation to the squeezed feature vector and pass it through the sigmoid function to obtain the importance scores for each nodes. These importance scores are then multiplied with the original feature vector to obtain the final readout feature vector. Given a set of node features $\mathbf{H}$ with dimensionality D, the SERO function outputs the feature vector $\mathbf{h_{SERO}}$ of dimensionality $\mathbb{D}' \in \mathbb{R}^{N \times D}$ as:

$$z = \mathbf{H}\phi_{mean} \tag{16}$$
$$zspace = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 z)) \tag{17}$$
$$\mathbf{h}_{SERO} = \mathbf{H}z_{space}. \tag{18}$$

where σ is the sigmoid function, δ is the ReLU activation function, $\mathbf{W}_1 \in \mathbb{R}^{D \times D}$, $\mathbf{W}_2 \in \mathbb{R}^{N \times D}$ are learnable weight matrices. The first equation computes the global information of the node features using global average pooling, and the second equation models the channel-wise dependencies by learning the feature importance scores. The third equation combines the original feature vectors with the computed importance scores to obtain the final readout feature vector.

Furthermore, as the effectiveness of the mean and max element-wise pooling operation is proven (He et al. [2020]), we compare the attention-based readout function with the mean and max element-wise pooling readout method. For a network, G with l convolution and respective pooling layers, the output graph of the lth pooling block is summarized using a mean and max pooling operation element-wise on $H(l) = [h(l)_i : i = 1, ..., N(l)]$. The resulting vector is obtained by concatenating both the mean and max summaries. To obtain a graph-level representation, the summary vectors from each layers of RGIN block are concatenated together.

2.1.4 Loss Functions

We primarily use the SmoothL1Loss function as a criterion to train the model. SmoothL1Loss is preferred over Mean Squared Error (MSE) Loss as it provides a more robust and less sensitive measure of distance between predicted and target values, with a less steep gradient around zero that makes it less sensitive to outliers and noise, leading to better model convergence and improved performance. It is defined as:

$$L_{\mathrm{smoothL1}}(x, y) = \begin{cases} 0.5 \cdot (x-y)^2, & \text{if } |x-y| < 1 \\ |x-y| - 0.5, & \text{otherwise} \end{cases}$$

Here, x represents the predicted value and y is the target value.

Furthermore, we also add the Unit Loss and TopK pooling loss (Li et al. [2021]) for interpretation. Unit Loss is important because we use a learnable vector ω to represent nodes, but this vector can be scaled in different ways without affecting the results. To solve this issue, we add a rule that the vector must have a fixed length of 1, and we use a loss function to ensure that this rule is followed, which is Unit Loss. The loss is formulated as:

$$L_{\mathrm{unit}}(v) = ||\omega||_2 - 1$$

Here, ω represents the learnable vector representing a node, and $||\omega||_2$ is the L2 norm of the vector. The goal is to minimize this loss, which effectively encourages the learnable vectors to have a fixed length of 1. This helps in finding

meaningful and unique values for the vector while accounting for different scaling factors.

Furthermore, using TopK Pooling loss ensures that top K selected indicative ROIs should have significantly different scores than those of the unselected nodes. It helps in identifying the most relevant ROIs. In order to achieve the desired behavior in our model, where selected nodes should exhibit scores close to 1, while unselected nodes should approach a score of 0, we employ a training strategy as in (Li et al. [2021]). This strategy involves ranking the sigmoid-transformed scores, $\tilde{s}_m^{(l)}$, in descending order for each instance m, resulting in $\hat{s}_m^{(l)} = [\hat{s}_{m,1}^{(l)}, \ldots, \hat{s}_{m,N^{(l)}}^{(l)}]$. To introduce diversity in these ranked scores across all M training instances, we employ a specific constraint. In practice, we define the TPK loss using binary cross-entropy, given by:

$$L_{\text{TPK}}^{(l)} = -\frac{1}{M} \sum_{m=1}^{M} \left(\frac{1}{N^{(l)}} \sum_{i=1}^{k} \log(\hat{s}_{m,i}^{(l)}) + \sum_{i=1}^{N^{(l)}-k} \log(1 - \hat{s}_{m,i+k}^{(l)}) \right) \tag{19}$$

This loss encourages the scores of selected ROIs to be higher ensuring that they stand out from the rest of the nodes.

The total loss L_{total} is a combination of these individual loss components:

$$L_{\text{total}} = L_{\text{smoothL1}} + \sum_{l=1}^{L} L_{\text{unit}}^{(l)} + \lambda_1 \sum_{l=1}^{L} L_{\text{TPK}}^{(l)}$$

Where L represents the number of layers in the model, and λ_1 is a hyperparameter that balances the contribution of the TopK Pooling loss. This combination of loss components helps in training the model to achieve good convergence, meaningful node representations, and relevant ROIs selection.

Overall with the architecture above, our proposed model for resting state fMRI data prediction using staticFNC could bring several benefits:

- By replacing the aggregation function of GIN with that of RGCN, the resulting RGIN model takes advantage of both permutation invariant update function and aggregation functions.

- We tackle the limitation of all brain ROIs learning embeddings in an identical manner by incorporating a clustering-based embedding method in the convolutional layer of the graph.

- Instead of using conventional READOUT methods of GNNs, we make use of attention-based readout function which effectively captures the information by obtaining the importance scores for each node.

3 Experiments

3.0.1 The ABCD (Adolescent Brain Cognitive Development) Dataset

ABCD is a large ongoing study following youths from age 9-10 into late adolescence to understand factors that increase the risk of physical and mental health problems. Participants were recruited from 21 sites across the US to represent various demographic variables. Data used in this study were from 8520 children aged 9–10 at baseline, including resting state fMRI image and fluid intelligence, crystallized intelligence, and total composite scores. Data were split into training (n = 5964), validation (n = 1278), and test (n = 1278) subsets. There were 4430 male and 4089 female subjects in this study. We note that age and site effects were regressed out from the intelligence scores to make sure that only the relevant intelligence features are captured by our graph model.

We conducted preprocessing on the raw resting-state fMRI data utilizing a combination of the FMRIB Software Library (FSL) v6.0 toolbox and Statistical Parametric Mapping (SPM) 12 toolbox within the MATLAB 2019b environment. The preprocessing encompassed several key steps, namely: 1) correction for rigid body motion; 2) distortion correction; 3) removal of dummy scans; 4) standardization to the Montreal Neurological Institute (MNI) space; and 5) application

Table 1: Comparison of different *BrainRGIN* architectures for intelligence prediction

Aggregation	Readout	Fluid Intelligence		Crystallized Intelligence		Total Composite Scores	
		MSE	Corr	MSE	Corr	MSE	Corr
Sum	Attention Based Readout (SERO)	**263**	**0.23**	295.95	0.27	**261**	**0.31**
Mean		269	0.21	295	0.27	268.58	0.29
Sum	Attention Based Readout (GARO)	274.69	0.20	**263.7**	**0.30**	291	0.29
Mean		270.05	0.20	268.7	0.28	267.01	0.29
Sum	Mean + Max Readout	300	0.19	299.81	0.27	271.37	0.30
Mean		298	0.19	298.30	0.27	289.93	0.30

of a 6 mm Gaussian kernel for smoothing. Subsequently, we employed a fully automated spatially constrained independent component analysis framework to extract 53 robust intrinsic connectivity networks (ICNs) using the Neuromark_fmri_1.0 template. 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.

3.1 Experimental Setup

The neural network architecture depicted in Fig. 1 was implemented using Pytorch (Paszke et al. [2019]), and Pytorch Geometric (Fey and Lenssen [2019]) for the specific graph neural network components. The number of nodes was 53 (corresponding to NeuroMark (Du et al. [2020]) ICA Components), and the number of node features per each node was the number of pairwise correlations between each other nodes (i.e., 53). The performance of the *BrainRGIN* architecture was evaluated using the ABCD dataset for predicting fluid intelligence, crystallized intelligence, and total composite scores. We created three different variations of *BrainRGIN* for various intelligence prediction tasks. The number of edges depends on the threshold percentage used to retain only the strongest correlations, which was selected to be 100% for all the prediction tasks. The model architecture for all intelligence prediction tasks was implemented with 3 RGIN layers with node sizes 32, 128, and 256 respectively followed by pooling layers as shown in Fig. 1 and attention-based readout functions for each layer. The pooling ratio from hyperparameter tuning was set to be 0.38 from fluid intelligence, 0.46 for crystallized intelligence and 0.78 for total composite scores. The number of clustered communities is set to 7, and the motivation for this comes from the seven functional networks defined by (Yeo et al. [2011]), because these 7 networks show key brain functionality relevance. The majority of these settings were fixed with the hyperparameter search. We also tested the *BrainRGIN* architecture with a variety of aggregation (Sum, Mean) and readout (SERO, GARO and Mean+Max) architectures to identify how different choices affect the performance of the model. We first identified the best *BrainRGIN* architecture compared with a variety of graph convolution, pooling, and readout architectures for predicting intelligence. Finally, the best architecture was used for comparison with baseline models.

We compared the performance of *BrainRGIN* with popular recent models along with traditional robust ML models as baseline models. The recent baselines includes BrainGNN (Li et al. [2021]), BrainNetCNN (Kawahara et al. [2017]), Brain Network Transformer Kan et al. [2022b], FBNetGen Kan et al. [2022a], Graph Transformers Dwivedi and Bresson [2020] etc. We also made comparisons with traditional established ML methods such as Support Vector Regression (SVR), Logistic Regression (LR), and Ridge Regression (Ridge). We particularly excluded reporting the performance of other baseline models like GCN (Kipf and Welling [2016]), GAT (Veličković et al. [2017]), GraphSage (Hamilton et al. [2017]) because of proven BrainGNN superior performance over these models. For BrainGNN, we performed multiple sweeps to identify the best performance. Based on performance, setting the learning rate of both BrainRGIN and BrainGNN initially to 0.001 and reducing every 30 epochs achieved the best performance for all prediction tasks. For BrainGNN, we used two convolutional layers of size 32, as presented in the original paper (Li et al. [2021]), with a fully connected layer of size 512. We utilized the authors' provided open-source codes for BrainGNN Li et al. [2021], FBNetGen Kan et al. [2022a] and Brain Network Transformer (BNT) Kan et al. [2022b], conducting grid searches to fine-tune essential hyperparameters based on the recommended best configurations. Specifically, for BrainGNN, we also explored various learning rates (0.01, 0.005, 0.001) while for FBNetGen, we investigated how varying hidden dimensions (8, 12, 16) impacts the performance. For BNT, we used different hidden dimensions (256, 512, 1024) during scaled dot product attention, along with varying fully connected layers (2 and 3). In the case of BrainNetCNN Kawahara et al. [2017], we experimented with different dropout rates (0.3, 0.5, 0.7). Regarding GT (Graph Transformer) Dwivedi and Bresson [2020], we examined the impact of different number of transformer layers (1, 2, 4) and the number of attention heads (2, 4, 8) over 50 epochs of training. For the remaining baseline methods, we employed Polyssifier, a widely used tool for baseline model comparisons, leveraging its automatic hyperparameter search feature to identify optimal model parameters for maximizing performance. For the other traditional baseline

Table 2: Comparison of *BrainRGIN* With Baseline Models on ABCD Dataset

Models	Linear	Fluid Intelligence		Crystallized Intelligence		Total Composite Scores		
		MSE	Correlation	MSE	Correlation	MSE	Correlation	
SVR	No	275	0.20	307	0.24	307	0.24	
LR	Yes	326	0.19	354	0.22	345	0.23	
Ridge	Yes	325	0.19	353	0.22	349	0.23	
BrainGNN Li et al. [2021]	No	264	0.22	288.29	0.28	294.12	0.29	
BrainRGIN (Ours)	No	263	0.23	263.7	0.30	261	0.31	
BrainNetCNN Kawahara et al. [2017]	No	372.32	0.17	367.82	0.204	375.26	0.218	
Brain Network Transformer Kan et al. [2022b]	No	261.59	0.248	265.12	0.32	264.65	0.339	
FBNetGen Kan et al. [2022a]	No	265.21	0.21	276.27	0.315	273.74	0.30	
Graph Transformers Dwivedi and Bresson [2020]	No	267	0.22	281	0.23	278	0.32	

Table 3: Effect of edges threshold selection in model prediction

Edge Threshold	MSE		
	Fluid Intelligence	Crystallized Intelligence	Total Composite Intelligence
100%	263	263.7	261
80%	269.22	276.12	272
60%	271.35	278.38	269.24
50%	288.12	281.06	283
40%	290.05	297.12	288

methods, we used Polyssifier (https://github.com/sergeyplis/polyssifier). This tool is widely used to perform baseline model comparisons and has a feature to automatically perform hyperparameter search and identify the best parameters of the model for the highest performance. For the evaluation metrics, we considered Mean Squared Error (MSE) and Correlation. The results were reported using a separate holdout testing set (n=1278) and the average scores of experiments using four different seeds are reported for every intelligence prediction task.

4 Results

The experimental results are summarized in Table 1 and Table 2. In Table 1, it is evident that the *BrainRGIN* architecture demonstrated promising performance in predicting fluid intelligence, achieving a mean squared error (MSE) of 263 and a correlation coefficient of 0.23 when employing RGIN convolution combined with the SERO readout method. Moreover, it yielded the best results in predicting crystallized intelligence, with an MSE of 263.7 and a correlation of 0.30 using the same RGIN convolution and SERO readout method. Notably, for total composite scores, the GARO attention-based readout function in conjunction with the RGIN graph model attained the highest performance, achieving an MSE of 261 and a correlation of 0.31.

Furthermore, we observed stable and comparable performances from baseline models such as BrainNetCNN Kawahara et al. [2017], FBNetGen Kan et al. [2022a], and GT Dwivedi and Bresson [2020], indicating the reliability of these models. However, the Brain Network Transformer (BNT) Kan et al. [2022b] exhibited robust performance with higher correlation scores and lower mean squared errors. Although *BrainRGIN* occasionally reported lower mean squared errors compared to BNT, the performance appeared to be comparable. Additionally, *BrainRGIN* surpassed all traditional baseline models and exhibited superior performance compared to Support Vector Regression (SVR), Linear Regression (LR), and Ridge Regression. It consistently achieved lower MSE values and higher correlation coefficients across all metrics, namely fluid intelligence, crystallized intelligence, and total composite scores. These results underscore the effectiveness of RGIN aggregation and attention-based readout methods over the BrainGNN model and other baseline models in predicting intelligence scores. Moreover, attention-based readout methods were found to outperform other readout techniques, significantly contributing to the architecture's success in predicting intelligence scores.

These findings underscore the importance of carefully selecting appropriate graph neural network components for predicting intelligence scores from rsfMRI data and provide a valuable foundation for future research in this area.

4.1 Hyperparameter Study

We also analyzed how changing the edges threshold affected the performance of the model. Interestingly, as can be seen from Table 3, we saw that using 100% of edges gave a superior performance for prediction for all intelligence prediction tasks. Besides, inspired by BrainGNN (Li et al. [2021]) hyperparameter study for TopK Loss regularizer, we fixed the regularizer $\lambda 1$ to 0.1 to partially force the network to select the same top nodes and also make sure that the selected nodes have different scores than those of the unselected nodes.

4.2 Interpretation of brain regions and networks

To identify the top ROIs for the intelligence prediction tasks, we first extracted the pooled ROIs after passing through the first layer of TopK Pool layer of a fully trained model using a holdout test set. We then selected the ROIs with highest frequency across each of the samples for all the intelligence scores. For fluid intelligence, we selected two ROIs (Middle Temporal Gyrus, and Middle Frontal Gyrus) (Fig 2(a)) with 90% frequency threshold. Similarly, for crystallized intelligence, middle frontal gyrus and caudate were selected with a frequency threshold of 90% (Fig 2(b)). For total composite scores, as the pooling ratio from hyper-parameter tuning was higher (0.78), we were able to extract 21 brain regions for 95% frequency. We categorized the selected regions into seven connectivity networks and plotted them independently in Fig. 3.

4.3 Experiment with HCP dataset

4.3.1 HCP Dataset

For generalizability, we used HCP Essen et al. [2013] data was first minimally pre-processed following the pipeline described in Glasser et al. [2013]. The preprocessing includes gradient distortion correction, motion correction, and field map preprocessing, followed by registration to T1 weighted images. The registered EPI image was then normalized to the standard MNI152 space. To reduce noise from the data, FIX-ICA based denoising was applied Salimi-Khorshidi et al. [2014]. To minimize the effects of head motion subject scans with framewise displacement (FD) over 0.3mm at any time of the scan were discarded. The FD was computed with fsl motion outliers function of the FSL Jenkinson et al. [2012]. There were 152 discarded scans from filtering out with the FD, and 1075 scans were left. For all experiments, the scans from the first run of HCP subjects released under S1200 were used. Subsequently, similar to preprocessing with ABCD dataset, we employed a fully automated spatially constrained independent component analysis framework to extract 53 robust intrinsic connectivity networks (ICNs) using the Neuromark fmri 1.0 template which categorized into seven functional domains based on their anatomical locations and functional characteristics. Furthermore, we also regressed out age and gender effects from the intelligence scores to make sure the model captures relevant feature representations.

4.3.2 Hyperparameter settings for HCP

As the measures of intelligence on the HCP dataset are different than those of ABCD, we retrained the model with different hyperparameters as we expected configuration changes as compared to the ABCD architecture. Similar to the ABCD dataset, we created three variations of *BrainRGIN* for various intelligence prediction tasks. The edges threshold percentage was set to 100% referencing the earlier ABCD results. Similar to ABCD experiments, the model was trained with 3 RGIN layers with node sizes 32, 128, and 256 followed by a pooling layer and attention-based readout as shown in Fig 1. The pooling ratio from sweeps from hyperparameter tuning was set to be 0.42 for fluid, 0.52 for crystallized, and 0.65 for total composite scores. We used sum aggregation with SERO readout which was found to be robust with ABCD experiments. The results reported in Table 4 are the average scores of experiments using four different seeds for every intelligence prediction task on separate holdout testing sets (n=192).

4.3.3 Results from HCP dataset

The performance of various models on the HCP dataset is presented in Table 4. Notably, our model, *BrainRGIN* , demonstrates competitive performance across all intelligence scores. For fluid intelligence prediction, BrainRGIN achieves a correlation coefficient of 0.262 with a mean squared error (MSE) of 274.92. Similarly, for predicting crystallized intelligence, BrainRGIN achieves a correlation coefficient of 0.2937 with an MSE of 308.21. Additionally, for total composite scores, BrainRGIN achieves a correlation coefficient of 0.309 with an MSE of 306.45. Comparatively, Brain Network Transformer and FBNetGen also exhibit noteworthy performance, achieving very comparable results. However, similar to ABCD experiments, our model reports slightly lower mean squared errors in some cases, indicating its potential for more precise predictions based on a graph convolution method.

5 Discussion and Conclusion

In this research study, a novel technique called Brain ROI-Aware Graph Isomorphism Networks, *BrainRGIN* , was proposed to predict intelligence using static FNC matrices derived from resting-state fMRI data. *BrainRGIN* integrates

Table 4: Comparison of *BrainRGIN* With Baseline Models on HCP Dataset

Model	Fluid		Crystallized		TotalComp		
	Correlation	**MSE**	**Correlation**	**MSE**	**Correlation**	**MSE**	
BrainRGIN (Ours)	0.262	274.92	0.2937	308.21	0.309	306.45	
BrainGNN Li et al. [2021]	0.256	278.81	0.285	312.52	0.285	314.233	
Brain Network Transformer Kan et al. [2022b]	0.271	278.51	0.302	303.26	0.318	312.44	
BrainNetCNN Kawahara et al. [2017]	0.251	303.21	0.268	328.14	0.271	322.23	
FBNetGen Kan et al. [2022a]	0.257	281.21	0.291	307.38	0.301	316.34	
Graph Transformer Dwivedi and Bresson [2020]	0.256	361	0.28	318	0.267	372.32	

(a) Middle frontal gyrus and Middle temporal gyrus

(b) Middle frontal gyrus and Caudate

Figure 2: Regions Significant in Fluid and Crystallized Intelligence Prediction

the expressive power of GIN and clustering-based GCN and also incorporates attention-based readout function, in the hope of better representing brain networks and improving model prediction. Specifically, by replacing the aggregation function of GIN with that of RGCN, the model leverages the powerful representation learning capability of GIN, while still capturing the edge strength and edge type represented by a clustering-based embedding method. Another notable aspect of the proposed architecture is the use of attention-based readout functions instead of conventional readout methods, which is proven to be very effective as can be seen from Table 1. The attention mechanism assigns importance scores to each node, effectively capturing spatial relevance information for prediction. Using attention-based readout functions not only improved the overall prediction of the model but also validated the theory that different brain regions contribute in a different manner to intelligence prediction.

To ensure the model is robust to influence from demographic factors such as age, gender, site, and socioeconomic status, we have addressed demographic factors such as age, gender, and site (which is expected to have a huge effect on MRI data) being controlled for the experiment. Specifically, we effectively regressed these factors from intelligence scores during model evaluation to mitigate their influence. It's noteworthy that before regressing out site from intelligence scores, the final correlation scores were as high as 0.35 for fluid intelligence, 0.42 for crystallized intelligence, and 0.41 for total composite scores with lower mean squared errors. However, removing these covariates led to a drop in overall prediction scores based on correlation and MSE metrics as presented in Table 2. Furthermore, we did an in-depth analysis to assess the potential impact of socioeconomic status (SES), focusing on income and education, on our model evaluation for crystallized intelligence. Originally, our data revealed moderate positive correlations between SES factors and predicted (0.28) and original crystallized intelligence scores (0.33). To further evaluate the influence of SES, we retrained the model on crystallized intelligence, controlling for income and education to generate new intelligence scores, keeping the model configuration consistent with earlier training. Surprisingly, upon removing SES factors from the evaluation, the model's performance improved slightly, as evidenced by an increased correlation of 0.331 from 0.30 on crystallized intelligence and a lower mean squared error of 256.21 from 263.7. The top regions of interest were similar to the earlier model where we did not regress out SES. These findings robustly demonstrate that our model evaluation remains unaffected by socioeconomic status and other demographic factors.

The top two regions contributing to fluid intelligence are the middle frontal gyrus and the middle temporal gyrus. In fact middle frontal gyrus contributes to all three intelligence scores, and its overall significance in predicting intelligence is in line with the literature. In particular, the dorsolateral prefrontal cortex in the middle frontal gyrus is well known to play a crucial part in many domains of cognitive processes, including working memory (Barbey et al. [2013]), attention control (Knight et al. [2020]), executive function (Nejati et al. [2021]), and decision making (Patterson et al. [2007]), etc (Yu et al. [2022]). Given its multi-domain involvement in cognition, this region's engagement substantiates its relevance in all intelligence scores. The function of the middle temporal gyrus is relatively more specific to language relevant (Turken and Dronkers [2011]), such as semantic processing (Binder et al. [2009]), syntactic comprehension (Yu et al. [2022]), language comprehension, decoding intelligible speech. Its role in fluid intelligence prediction could

(a) Cognitive Control Network

(b) Default Mode Network

(c) Sensorimotor Network

(d) Subcortical Network

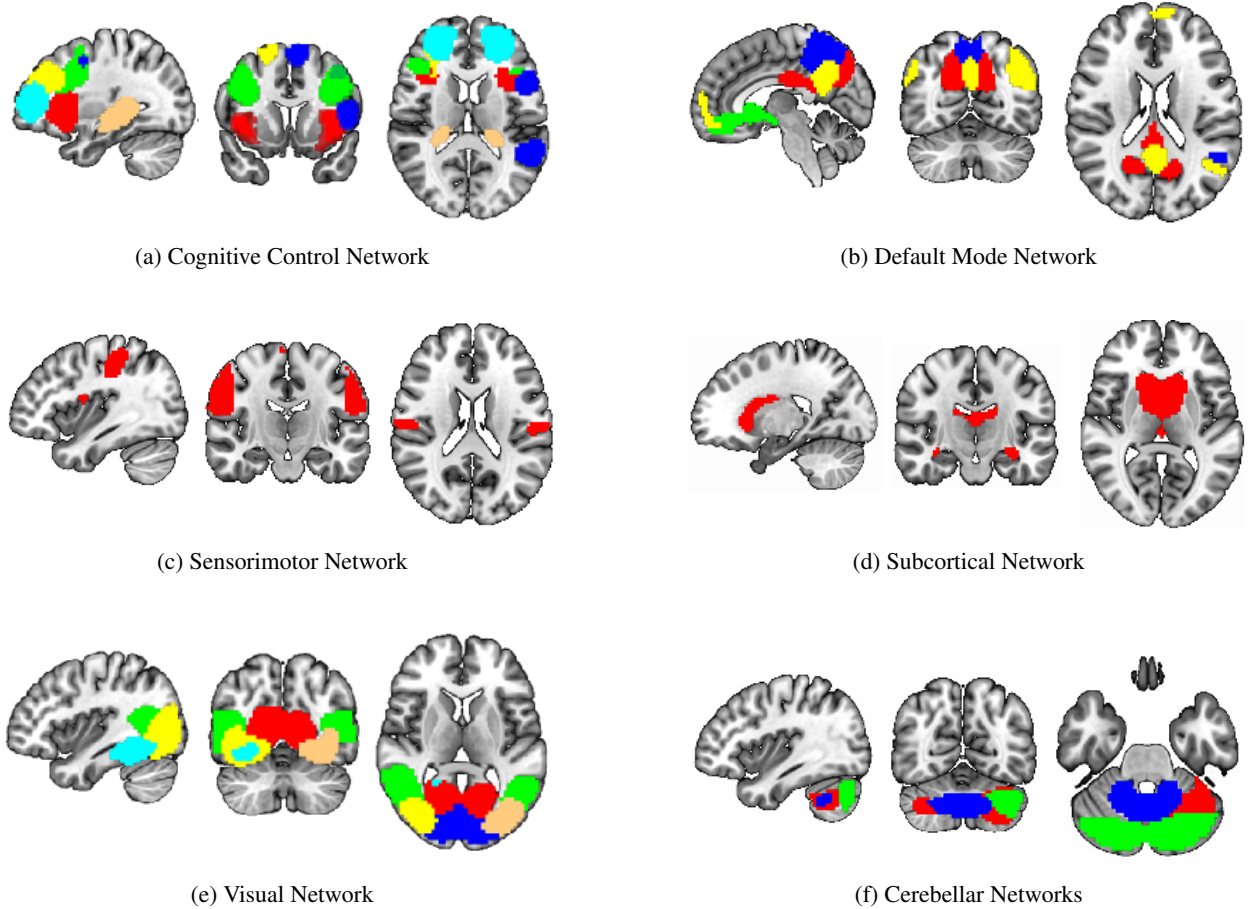(e) Visual Network

(f) Cerebellar Networks

Figure 3: Significant regions expressed as Connectivity Networks for Total Composite Scores

be attributed to its contribution to information decoding and integration. In the context of crystallized intelligence, besides the middle frontal gyrus, caudate plays the most remarked contribution. The caudate's association with learning (Choi et al. [2020]) and memory consolidation (Müller et al. [2017]) aligns with crystallized intelligence's reliance on acquired knowledge and experience (Grazioplene et al. [2014]). The combined involvement of middle frontal gyrus and caudate suggests a cooperative engagement in translating stored knowledge into practical problem-solving skills, a hallmark of crystallized intelligence.

The diverse set of brain regions identified as relevant for total composite score reflects the intricate and distributed nature of cognitive processes, as well as the general and broad attribute of 'intelligence'. These regions cover six relatively separate brain networks. Sensorimotor network (postcentral gyrus) and visual network(middle/inferior occipital gyri, calcarine cortex, cuneus, fusiform) involve basic functions of various sensory inputs (J;Tadi) and further integration of inputs (BrainMadeSimple [2022]). The cognitive control network covers particularly frontal-parietal regions (superior/middle frontal gyri, supplementary motor area, insula, hippocampal) and participates in various higher-order cognitive processes (Jacques et al. [2018]), such as executive function, attention, working memory, and planning (Tanji and Shima [1994]). The default network comprises the anterior cingulate cortex, posterior cingulate cortex, and precuneus and reflects the brain's intrinsic activity in contrast to task-oriented controls (Zanto and Gazzaley [2013]). Their functions are linked to self-referential thinking (Raichle [2015]), introspection, and episodic memory retrieval (Gusnard et al. [2001]). Subcortical network, precisely caudate as discussed before has a role in learning and memory and has been shown to be able to predict IQ (Vincent et al. [2006]). Last but not the least is the cerebellum. Its function is well recognized to be beyond motor function, but subserving cognition (Koziol et al. [2011]), and recently cerebellum has been partitioned to echo the 7 cortical networks (King et al. [2019]). In summary, the interplay of these diverse brain regions likely contributes to the complexity of cognitive processes that manifest as Total Composite Scores. The extensive network of regions suggests that intelligence is not confined to a single brain area but rather

emerges from the collaboration of numerous interconnected regions, each contributing its unique specialization to overall cognitive performance.

In conclusion, the *BrainRGIN* architecture presented in this study offers several advantages for intelligence prediction using resting-state fMRI data. It effectively captures the local and global relationships between brain regions, incorporates edge information, utilizes attention-based readout functions, and addresses the variability among brain ROIs. The superior performance of *BrainRGIN* over baseline models highlights its potential for predicting intelligence and provides valuable insights into the relationship between brain function and cognitive processes. Future research can explore the application of this architecture to other types of fMRI data, such as task-based fMRI, to further advance our understanding of brain functioning.

Acknowledgements

Author contributions statement

B.T. conducted the primary data analysis, coding implementation, experiments, writing, and revision of the manuscript. R.S, B.R., and P.S were involved in the in-depth discussion of the analyses and reviewing the manuscript. J.C. assisted in analyzing the results, and also provided assistance in interpretation. V.D.C was involved in the design of the project and the manuscript review. J.L. was involved in the design, data analysis, and revision of the manuscript. E.A was involved in writing and editing the manuscript. All authors reviewed the manuscript.

Code and Data Availability for Replication

The code for replication is available at `https://github.com/bishalth01/BrainRGIN`. Data will be made available upon request.

Financial Disclosures Section

All authors declare that they have no conflicts of interest.

References

Patrick Kyllonen and Harrison Kell. Fluid ability (gf) and complex problem solving (CPS). *Journal of Intelligence*, 5(3): 28, July 2017. doi:10.3390/jintelligence5030028. URL `https://doi.org/10.3390/jintelligence5030028`.

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)*. IEEE, April 2023. doi:10.1109/isbi53787.2023.10230722. URL `https://doi.org/10.1109/isbi53787.2023.10230722`.

Bhaskar Ray, Jiayu Chen, Zening Fu, Pranav Suresh, Bishal Thapaliya, Britny Farahdel, Vince D. Calhoun, and Jingyu Liu. Replication and refinement of brain age model for adolescent development. August 2023. doi:10.1101/2023.08.16.553472. URL `https://doi.org/10.1101/2023.08.16.553472`.

Michael A. Ferguson, Jeffrey S. Anderson, and R. Nathan Spreng. Fluid and flexible minds: Intelligence reflects synchrony in the brain's intrinsic network architecture. *Network Neuroscience*, 1(2):192–207, June 2017. doi:10.1162/netn_a_00010. URL `https://doi.org/10.1162/netn_a_00010`.

Tong He, Ru Kong, Avram J. Holmes, Minh Nguyen, Mert R. Sabuncu, Simon B. Eickhoff, Danilo Bzdok, Jiashi Feng, and B.T. Thomas Yeo. Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. *NeuroImage*, 206:116276, February 2020. doi:10.1016/j.neuroimage.2019.116276. URL `https://doi.org/10.1016/j.neuroimage.2019.116276`.

Julien Dubois, Paola Galdi, Lynn K. Paul, and Ralph Adolphs. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1756):20170284, August 2018. doi:10.1098/rstb.2017.0284. URL `https://doi.org/10.1098/rstb.2017.0284`.

Bruno Hebling Vieira, Gustavo Santo Pedro Pamplona, Karim Fachinello, Alice Kamensek Silva, Maria Paula Foss, and Carlos Ernesto Garrido Salmon. On the prediction of human intelligence from neuroimaging: A systematic review of methods and reporting. *Intelligence*, 93:101654, July 2022. doi:10.1016/j.intell.2022.101654. URL `https://doi.org/10.1016/j.intell.2022.101654`.

M.H. Lee, C.D. Smyser, and J.S. Shimony. Resting-state fMRI: A review of methods and clinical applications. *American Journal of Neuroradiology*, 34(10):1866–1872, August 2012. doi:10.3174/ajnr.a3263. URL `https://doi.org/10.3174/ajnr.a3263`.

Luke J. Hearne, Jason B. Mattingley, and Luca Cocchi. Functional brain networks related to individual differences in human intelligence at rest. *Scientific Reports*, 6(1), August 2016. doi:10.1038/srep32328. URL `https://doi.org/10.1038/srep32328`.

Liye Wang, Chong-Yaw Wee, Heung-Il Suk, Xiaoying Tang, and Dinggang Shen. MRI-based intelligence quotient (IQ) estimation with sparse learning. *PLOS ONE*, 10(3):e0117295, March 2015. doi:10.1371/journal.pone.0117295. URL `https://doi.org/10.1371/journal.pone.0117295`.

Liangwei Fan, Jianpo Su, Jian Qin, Dewen Hu, and Hui Shen. A deep network model on dynamic functional connectivity with applications to gender classification and intelligence prediction. *Frontiers in Neuroscience*, 14, August 2020. doi:10.3389/fnins.2020.00881. URL `https://doi.org/10.3389/fnins.2020.00881`.

Yang Li, Xin Ma, Raj Sunderraman, Shihao Ji, and Suprateek Kundu. Accounting for temporal variability in functional magnetic resonance imaging improves prediction of intelligence. *Human Brain Mapping*, 44(13):4772–4791, July 2023. doi:10.1002/hbm.26415. URL `https://doi.org/10.1002/hbm.26415`.

Byung-Hoon Kim and Jong Chul Ye. Understanding graph isomorphism network for rs-fMRI functional connectivity analysis. *Frontiers in Neuroscience*, 14, June 2020. doi:10.3389/fnins.2020.00630. URL `https://doi.org/10.3389/fnins.2020.00630`.

Anees Kazi, Luca Cosmo, Seyed-Ahmad Ahmadi, Nassir Navab, and Michael M. Bronstein. Differentiable graph module (DGM) for graph convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):1606–1617, February 2023. doi:10.1109/tpami.2022.3170249. URL `https://doi.org/10.1109/tpami.2022.3170249`.

Naresh Nandakumar, Komal Manzoor, Shruti Agarwal, Jay J. Pillai, Sachin K. Gujar, Haris I. Sair, and Archana Venkataraman. Automated eloquent cortex localization in brain tumor patients using multi-task graph neural networks. *Medical Image Analysis*, 74:102203, December 2021. doi:10.1016/j.media.2021.102203. URL `https://doi.org/10.1016/j.media.2021.102203`.

Sofia Ira Ktena, Sarah Parisot, Enzo Ferrante, Martin Rajchl, Matthew Lee, Ben Glocker, and Daniel Rueckert. Metric learning with spectral graph convolutions on brain connectivity networks. *NeuroImage*, 169:431–442, April 2018. doi:10.1016/j.neuroimage.2017.12.052. URL `https://doi.org/10.1016/j.neuroimage.2017.12.052`.

Antonín Škoch, Barbora Rehák Bučková, Jan Mareš, Jaroslav Tintěra, Pavel Sanda, Lucia Jajcay, Jiří Horáček, Filip Španiel, and Jaroslav Hlinka. Human brain structural connectivity matrices–ready for modelling. *Scientific Data*, 9(1), August 2022. doi:10.1038/s41597-022-01596-9. URL `https://doi.org/10.1038/s41597-022-01596-9`.

Dongya Wu, Xin Li, and Jun Feng. Connectome-based individual prediction of cognitive behaviors via graph propagation network reveals directed brain network topology. *Journal of Neural Engineering*, 18(4):0460a3, July 2021. doi:10.1088/1741-2552/ac0f4d. URL `https://doi.org/10.1088/1741-2552/ac0f4d`.

Guixiang Ma, Nesreen K. Ahmed, Ted Willke, Dipanjan Sengupta, Michael W. Cole, Nicholas B. Turk-Browne, and Philip S. Yu. Similarity learning with higher-order graph convolutions for brain network analysis, 2018. URL `https://arxiv.org/abs/1811.02662`.

Salim Arslan, Sofia Ira Ktena, Ben Glocker, and Daniel Rueckert. Graph saliency maps through spectral convolutional networks: Application to sex classification with brain connectivity, 2018. URL `https://arxiv.org/abs/1806.01764`.

Anees Kazi, Soroush Farghadani, and Nassir Navab. Ia-gcn: Interpretable attention based graph convolutional network for disease prediction, 2021. URL `https://arxiv.org/abs/2103.15587`.

Fabrizio Parente and Alfredo Colosimo. Functional connections between and within brain subnetworks under resting-state. *Scientific Reports*, 10(1), February 2020. doi:10.1038/s41598-020-60406-7. URL `https://doi.org/10.1038/s41598-020-60406-7`.

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*, 74:102233, December 2021. doi:10.1016/j.media.2021.102233. URL `https://doi.org/10.1016/j.media.2021.102233`.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks?, 2018. URL `https://arxiv.org/abs/1810.00826`.

Xuan Kan, Hejie Cui, Joshua Lukemire, Ying Guo, and Carl Yang. Fbnetgen: Task-aware gnn-based fmri analysis via functional brain network generation, 2022a. URL `https://arxiv.org/abs/2205.12465`.

Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs, 2020. URL `https://arxiv.org/abs/2012.09699`.

Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer, 2022b. URL `https://arxiv.org/abs/2210.06681`.

Hejie Cui, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, and Carl Yang. Interpretable graph neural networks for connectome-based brain disorder analysis, 2022. URL `https://arxiv.org/abs/2207.00813`.

Liyu Gong and Qiang Cheng. Exploiting edge features in graph neural networks, 2018. URL `https://arxiv.org/abs/1809.02709`.

Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks, 2017. URL `https://arxiv.org/abs/1709.01507`.

Byung-Hoon Kim, Jong Chul Ye, and Jae-Jin Kim. Learning dynamic graph representation of brain connectome with spatio-temporal attention, 2021. URL `https://arxiv.org/abs/2105.13495`.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017. URL `https://arxiv.org/abs/1706.03762`.

Abhijit Guha Roy, Nassir Navab, and Christian Wachinger. Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. *IEEE Transactions on Medical Imaging*, 38(2):540–549, February 2019. doi:10.1109/tmi.2018.2867261. URL `https://doi.org/10.1109/tmi.2018.2867261`.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, 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. URL `https://arxiv.org/abs/1912.01703`.

Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric, 2019. URL `https://arxiv.org/abs/1903.02428`.

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*, 28:102375, 2020. doi:10.1016/j.nicl.2020.102375. URL `https://doi.org/10.1016/j.nicl.2020.102375`.

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*, 106(3):1125–1165, September 2011. doi:10.1152/jn.00338.2011. URL `https://doi.org/10.1152/jn.00338.2011`.

Jeremy Kawahara, Colin J. Brown, Steven P. Miller, Brian G. Booth, Vann Chau, Ruth E. Grunau, Jill G. Zwicker, and Ghassan Hamarneh. Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, February 2017. ISSN 1053-8119. doi:10.1016/j.neuroimage.2016.09.046. URL `http://dx.doi.org/10.1016/j.neuroimage.2016.09.046`.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2016. URL `https://arxiv.org/abs/1609.02907`.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2017. URL `https://arxiv.org/abs/1710.10903`.

William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs, 2017. URL `https://arxiv.org/abs/1706.02216`.

David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, and Kamil Ugurbil. The WU-minn human connectome project: An overview. *NeuroImage*, 80:62–79, October 2013. doi:10.1016/j.neuroimage.2013.05.041. URL `https://doi.org/10.1016/j.neuroimage.2013.05.041`.

Matthew F. Glasser, Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce Fischl, Jesper L. Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R. Polimeni, David C. Van Essen, and Mark Jenkinson. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80:105–124, October 2013. doi:10.1016/j.neuroimage.2013.04.127. URL `https://doi.org/10.1016/j.neuroimage.2013.04.127`.

Gholamreza Salimi-Khorshidi, Gwenaëlle Douaud, Christian F. Beckmann, Matthew F. Glasser, Ludovica Griffanti, and Stephen M. Smith. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90:449–468, April 2014. doi:10.1016/j.neuroimage.2013.11.046. URL `https://doi.org/10.1016/j.neuroimage.2013.11.046`.

Mark Jenkinson, Christian F. Beckmann, Timothy E.J. Behrens, Mark W. Woolrich, and Stephen M. Smith. FSL. *NeuroImage*, 62(2):782–790, August 2012. doi:10.1016/j.neuroimage.2011.09.015. URL `https://doi.org/10.1016/j.neuroimage.2011.09.015`.

Aron K. Barbey, Michael Koenigs, and Jordan Grafman. Dorsolateral prefrontal contributions to human working memory. *Cortex*, 49(5):1195–1205, May 2013. doi:10.1016/j.cortex.2012.05.022. URL `https://doi.org/10.1016/j.cortex.2012.05.022`.

Helen C. Knight, Daniel T. Smith, and Amanda Ellison. The role of the left dorsolateral prefrontal cortex in attentional bias. *Neuropsychologia*, 148:107631, November 2020. doi:10.1016/j.neuropsychologia.2020.107631. URL `https://doi.org/10.1016/j.neuropsychologia.2020.107631`.

Vahid Nejati, Reyhaneh Majdi, Mohammad Ali Salehinejad, and Michael A. Nitsche. The role of dorsolateral and ventromedial prefrontal cortex in the processing of emotional dimensions. *Scientific Reports*, 11(1), January 2021. doi:10.1038/s41598-021-81454-7. URL `https://doi.org/10.1038/s41598-021-81454-7`.

Karalyn Patterson, Peter J. Nestor, and Timothy T. Rogers. Where do you know what you know? the representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12):976–987, December 2007. doi:10.1038/nrn2277. URL `https://doi.org/10.1038/nrn2277`.

Mengxia Yu, Yiying Song, and Jia Liu. The posterior middle temporal gyrus serves as a hub in syntactic comprehension: A model on the syntactic neural network. *Brain and Language*, 232:105162, September 2022. doi:10.1016/j.bandl.2022.105162. URL `https://doi.org/10.1016/j.bandl.2022.105162`.

And U. Turken and Nina F. Dronkers. The neural architecture of the language comprehension network: Converging evidence from lesion and connectivity analyses. *Frontiers in System Neuroscience*, 5, 2011. doi:10.3389/fnsys.2011.00001. URL `https://doi.org/10.3389/fnsys.2011.00001`.

Jeffrey R. Binder, Rutvik H. Desai, William W. Graves, and Lisa L. Conant. Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12):2767–2796, March 2009. doi:10.1093/cercor/bhp055. URL `https://doi.org/10.1093/cercor/bhp055`.

Yera Choi, Emily Yunha Shin, and Sungshin Kim. Spatiotemporal dissociation of fMRI activity in the caudate nucleus underlies human de novo motor skill learning. *Proceedings of the National Academy of Sciences*, 117(38):23886–23897, September 2020. doi:10.1073/pnas.2003963117. URL `https://doi.org/10.1073/pnas.2003963117`.

Nils C. J. Müller, Boris N. Konrad, Nils Kohn, Monica Muñoz-López, Michael Czisch, Guillén Fernández, and Martin Dresler. Hippocampal–caudate nucleus interactions support exceptional memory performance. *Brain Structure and Function*, November 2017. doi:10.1007/s00429-017-1556-2. URL `https://doi.org/10.1007/s00429-017-1556-2`.

Rachael G. Grazioplene, Sephira G. Ryman, Jeremy R. Gray, Aldo Rustichini, Rex E. Jung, and Colin G. DeYoung. Subcortical intelligence: Caudate volume predicts IQ in healthy adults. *Human Brain Mapping*, 36(4):1407–1416, December 2014. doi:10.1002/hbm.22710. URL `https://doi.org/10.1002/hbm.22710`.

DiGuiseppi J;Tadi. Neuroanatomy, postcentral gyrus. URL `https://pubmed.ncbi.nlm.nih.gov/31751015/`.

BrainMadeSimple, Feb 2022. URL `https://brainmadesimple.com/occipital-lobe/`.

Corentin Jacques, Jacques Jonas, Louis Maillard, Sophie Colnat-Coulbois, Laurent Koessler, and Bruno Rossion. The inferior occipital gyrus is a major cortical source of the face-evoked n170: Evidence from simultaneous scalp and intracerebral human recordings. *Human Brain Mapping*, 40(5):1403–1418, November 2018. doi:10.1002/hbm.24455. URL `https://doi.org/10.1002/hbm.24455`.

Jun Tanji and Keisetsu Shima. Role for supplementary motor area cells in planning several movements ahead. *Nature*, 371(6496):413–416, September 1994. doi:10.1038/371413a0. URL `https://doi.org/10.1038/371413a0`.

Theodore P. Zanto and Adam Gazzaley. Fronto-parietal network: flexible hub of cognitive control. *Trends in Cognitive Sciences*, 17(12):602–603, December 2013. doi:10.1016/j.tics.2013.10.001. URL `https://doi.org/10.1016/j.tics.2013.10.001`.

Marcus E. Raichle. The brain's default mode network. *Annual Review of Neuroscience*, 38(1): 433–447, July 2015. doi:10.1146/annurev-neuro-071013-014030. URL `https://doi.org/10.1146/annurev-neuro-071013-014030`.

Debra A. Gusnard, Erbil Akbudak, Gordon L. Shulman, and Marcus E. Raichle. Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(7):4259–4264, March 2001. doi:10.1073/pnas.071043098. URL `https://doi.org/10.1073/pnas.071043098`.

Justin L. Vincent, Abraham Z. Snyder, Michael D. Fox, Benjamin J. Shannon, Jessica R. Andrews, Marcus E. Raichle, and Randy L. Buckner. Coherent spontaneous activity identifies a hippocampal-parietal memory network. *Journal of Neurophysiology*, 96(6):3517–3531, December 2006. doi:10.1152/jn.00048.2006. URL `https://doi.org/10.1152/jn.00048.2006`.

Leonard F. Koziol, Deborah Ely Budding, and Dana Chidekel. From movement to thought: Executive function, embodied cognition, and the cerebellum. *The Cerebellum*, 11(2):505–525, November 2011. doi:10.1007/s12311-011-0321-y. URL `https://doi.org/10.1007/s12311-011-0321-y`.

Maedbh King, Carlos R. Hernandez-Castillo, Russell A. Poldrack, Richard B. Ivry, and Jörn Diedrichsen. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nature Neuroscience*, 22(8):1371–1378, July 2019. doi:10.1038/s41593-019-0436-x. URL `https://doi.org/10.1038/s41593-019-0436-x`.