BIG-GRAPH: BRAIN IMAGING GENETICS BY GRAPH NEURAL NETWORK

Anonymous authors
Paper under double-blind review

ABSTRACT

Imaging genetics is one of the foremost emerging fields in neuroscience research that aims to combine neuroimaging and genetic information with phenotypes to shed light on inherent underlying mechanisms. While significant progress has been made in integrating brain imaging, like functional magnetic resonance imaging (fMRI), with genetic data, such as single nucleotide polymorphisms (SNPs), little progress has been made in studying them jointly using graph structures. To raise a new perspective and overcome challenges in analyzing data with high dimensionality and inherently complex relationships, we developed a graphical neural network model (BIG-Graph) that jointly learns to effectively represent both neuroimaging and genetic data in a nonlinear manner without any prior knowledge. Here, we demonstrate that joint learning of imaging-genetics using BIG-Graph largely outperforms existing state-of-the-art Imaging genetics models and networks trained separately on neuroimaging or genetic data in predicting a variety of phenotypes.

1 INTRODUCTION

Imaging genetics has attracted considerable interest in recent years. Linking genetics and brain phenotypes as they relate to biomarkers or clinical phenotypes is vital in advancing our understanding of biological associations and fundamental mechanisms (Shen & Thompson, 2019). Recent data-driven insights have investigated how the brain manages cognition (Kong et al., 2021), how neural connectivity affects typical and disordered brain function and behavior (Fornito et al., 2013), and the contribution of genomics to brain features and behavior (Bassett & Sporns, 2017; Bassett et al., 2008; Lydon-Staley & Bassett, 2018; Hao et al., 2018).

Although studying imaging genetics has the potential to make significant contributions to biomedical discoveries, there are computational and statistical challenges that must be overcome to achieve the full benefits of these valuable data. The challenge lies in the unprecedented scale, dimensionality, and complexity of brain imaging genetics data, including evaluating and testing over a million SNPs in the genome for associations and analyzing hundreds or thousands of MRI images, which require effective models to unlock shared genetic and molecular underpinnings of neural systems. An additional difficulty is the relatively small effect size of genetic variance (SNPs) on neurobiological systems, with most SNPs accounting for less than 1% of variance in brain function/phenotypes. Even common genetic variants with large effects on brain functioning have been difficult to detect (Ansarifar & Wang, 2019). Hence, studies have been expanded to tens of thousands of subjects to build an effective framework with adequate detection power to decipher imaging genetics associations and to develop effective diagnostic, therapeutic, and preventive processes for complex brain disorders.

Existing imaging genetic datasets have widely included behavioral or cognitive phenotypes, a number of genotypes, such as SNPs, and MRI images modalities (structural and functional MR) or brain phenotypes (Consortium, 2009). In recent years, imaging genetics research has become increasingly focused on studying the brain via resting-state fMRI (rs-fMRI) because this acquisition can capture the brain's spontaneous functional brain architecture that occurs in a task-negative or resting state when an explicit task is not being performed. Then, using rs-fMRI, functional connectivity (FC) has been computed using correlations or covariance between spatially distant regions across rs-fMRI data (Van Den Heuvel & Pol, 2010; Rogers et al., 2007). Accordingly, the functional activation patterns of the brain can be used to investigate associations of particular behavioral or cognitive phenotypes with neural networks encompassing multiple related brain regions (Birnbaum & Weinberger, 2022)

and explore interactions between the resultant neuroimaging components and genetic features. To this end, several models have been applied mainly for imaging genetic studies approaches, such as correlation analysis (Sheng et al., 2014), partial least squares (Le Floch et al., 2012), reduced-rank regression (Vounou et al., 2010), and machine learning models (Sebenius et al., 2021; Hu et al., 2021). Among these techniques, as deep learning models, especially Graph Neural Network (GNN), have notably achieved promising and acceptable results in many applications, this paper focuses on developing a framework through GNNs for joint learning/modeling of genetic and imaging.

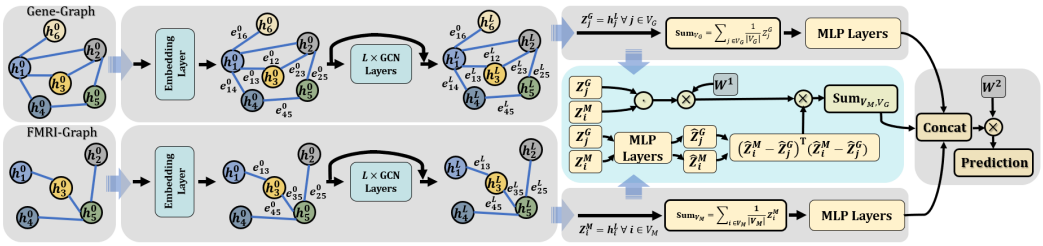

Figure 1: The overview of the proposed BIG-Graph model. FMRI and gene graphs are inputs to the model. Graphs are sent parallelly into the two GNNs with batch normalization and residual connections. Convolutional feature maps of the last layers from two networks are fed into average pooling layers, followed by a deep neural network transformation. The latent representation vectors of two graphs' nodes are combined via the element-wise product, followed by a linear transformation. Also, they are fed into MLP to apply multilayer. Then, the output of MLP is used to calculate the generalized distance function between all two graphs' nodes. The general form of imaging and genetics interactions is computed by multiplying the distance function's results and their transformed element-wise product. Ultimately, the final prediction was achieved using a linear classification or regression of average pooling of general form interactions and transformed latent representations of genes and fMRI.

This paper proposes a new graph convolution network called Brain Imaging Genetics by Graph Neural Network (BIG-Graph) to simultaneously model associations between imaging and genetics data to predict cognitive and behavioral phenotypes. Instead of starting from one brain graph where nodes represent the anatomical brain regions, BIG-Graph jointly models gene and brain graphs. Figure 1 summarizes the overview of the proposed BIG-Graph model. The contributions of this paper can be summarized as follows:

- We describe a new algorithm that jointly learns from structural and functional brain networks and genetic data by quantifying the contributions of neuroimaging and the genetic data and their interactions in the prediction of phenotypes.
- We use an interaction detection framework equipped with generalized metric learning techniques to properly formulate the fine-grained feature interactions of high-dimensional neuroimaging and genetic data.
- We construct graph-based genetic data utilizing GWAS analysis and random forest and build a brain graph using analysis of functional connectivity of rs-fMRI and structural MRI images.
- We benchmark different GNNs structures such as Vanilla Graph ConvNets (GCN) (Kipf & Welling, 2016), GraphSage (Hamilton et al., 2017), and Graph Attention Network (GAT) (Veličković et al., 2017) on the proposed model and benchmark coupling neuroimaging and the genetic data versus neuroimaging and the genetic data. Also, we benchmark various complex levels of fMRI and gene graphs to investigate dense graphs' prediction power versus more sparse graphs.
- We demonstrate the proposed framework's performance in the prediction of sex, heights, and age from the Philadelphia Neurodevelopmental Cohort (PNC) (Satterthwaite et al., 2016).

2 RELATED WORKS AND BACKGROUND

Targeted reviews of the literature in neuroimaging genetics include statistical and machine learning approaches (Shen & Thompson, 2019; Nathoo et al., 2019), multivariate methods (Liu & Calhoun, 2014), and multimodal analysis strategies (Liu et al., 2018).

A number of statistical and machine learning models have been developed to investigate associations between genetic variations and brain imaging quantitative traits (QTs) in connection with other clinical and cognitive biomarkers and behavioral phenotypes. Early imaging genetics studies largely focused on estimating genetic contributions to phenotypic variation. One class of approaches uses a GWAS analysis, for instance, a fast voxel-wise GWAS framework (Huang et al., 2015), a kernel machine method (Ge et al., 2015), gene-environment mixed effect model (Wang et al., 2017), brain imaging GWAS (Hua et al., 2015), multivariate regression methods (Hao et al., 2018; Wang et al., 2012; Vounou et al., 2012), regularized sparse canonical correlation analysis (Du et al., 2014; Yan et al., 2014), and Bayesian methods (Smeland et al., 2018; Wang et al., 2012). Pairwise analyses represent a major computational challenge and requires numerous univariate SNP–QT association tests. To overcome these challenges, neural network models have been employed in (Wang et al., 2018; Schmidt et al., 1992).

Another active research field in imaging genetics is to predict cognitive and behavioral phenotypes using imaging and genomics to better understand the relationship between these data and behavioral, cognitive, and clinical outcomes. Besides conventional prediction methods such as naive Bayes classifiers (Dukart et al., 2016) and support vector machines (Fan et al., 2006), more recent machine learning models have also been employed, such as multiple kernel learning (Rakotomamonjy et al., 2008; Peng et al., 2016), sparse multi-model learning (Wang et al., 2013), cascaded multi-view canonical correlation (Morris et al., 2017), a multi-task collaborative regression Zille et al. (2017), a neural network (Ning et al., 2018; Zhou et al., 2019b; Venugopalan et al., 2021; Yu et al., 2021; Zhou et al., 2019b), latent representation learning method for multi-modality Zhou et al. (2019a), and robust reduced rank graph regression (Zhu et al., 2018).

Even with recent advances in deep learning models (especially GNNs) and their methodological and practical impact on prediction problems, few studies in imaging genetics have taken advantage of jointly assessing brain imaging and genetics using graphs. Existing graph-based works in neuroimaging mainly focus on discovering a brain network from MRI images alone, including: single modality multi-view brain network GCN classifier (Zhang et al., 2018b;a), joint GCN model (Liu et al., 2019; Kawahara et al., 2017), multi-view GCN (Wen et al., 2022), graph attention network (Huang et al., 2022; Hu et al., 2021; Yang et al., 2019; Filip et al., 2020), pooling regularized GCN for fMRI biomarker analysis (Li et al., 2020; 2021), hierarchical GCN framework (Sebenius et al., 2021), multiplex GCN (Kong et al., 2021), and dynamic GCN (Zhao et al., 2022).

Deviating from the large body of previous works on imaging genetics, the most related study to this research is from Ko et al. (2022) in which authors built a novel deep generative and discriminative learning framework that jointly analyzes phenotypic and genotypic data for Alzheimer's disease diagnosis and cognitive score prediction. While our methodology as a novel GNN framework addresses relevant data integration challenges and introduces a generalized metric learning technique to not only identify associations between neuroimaging and genetic data but also to predict phenotypes accurately. By achieving valuable insights into outcome-relevant neurobiological mechanisms at the genetic level, this study represents the first attempt to our knowledge to predict phenotypes from the PNC cohort (Satterthwaite et al., 2016) using a GNN framework.

3 METHODS

3.1 PRELIMINARIES

Graph Neural Networks. In this paper, we used message passing-based graph neural networks, which iteratively update node representations locally from one layer to another using neighborhood nodes. The updating formula is independent of graph size and is defined as $h_i^{\ell+1} = f(h_i^{\ell}, \{h_j^{\ell}\}_{j \in \mathcal{N}_i})$, where $\mathcal{N}_i$ denotes the set of nodes connected to node i on the graph, h_i^{ℓ} is the d-dimensional embedding representation of node i at layer ℓ, and f is a mapping function defined in various forms

such as GCN (Kipf & Welling, 2016), GraphSage (Hamilton et al., 2017), and GAT (Veličković et al., 2017).

Modeling Interaction with Metric Learning. Factorization Machines (FMs), as the most promising interaction-based models to estimate target, can be formulated as

$$\hat{y}(\boldsymbol{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle x_i x_j, \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is feature vector, w_0 denotes the global bias, w_i represents the strength of the i-th feature x_i, $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ are embedded features corresponding to i-th and j-th features, and $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle$ computes interactions between the i-th and j-th features. In the original FMs paper (Rendle, 2010), $\boldsymbol{v}_i \in \mathbb{R}^k$ denotes the factorized feature vector for feature x_i, and $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle$ is inner product of $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$. To consider feature correlations in interaction term of FMs, Guo et al. (2020) proposed to use generalized metric learning with deep neural network (DNN) based distance function in FMs as

$$\hat{y}(\boldsymbol{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij} D(\boldsymbol{v}_i, \boldsymbol{v}_j) x_i x_j, \tag{2}$$

where w_{ij} denotes a transformation weight, and D is known as dissimilar pairs in metric learning. We use element-wise product of embedded features $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ and trainable vector $\boldsymbol{W}^1 \in \mathbb{R}^k$ to compute transformation weight. This presentation of w_{ij} increases the representation ability of prediction by enabling FMs to overcome the distance function limitations (non-negativity of distances). Also, instead of using a linear correlations function, we model this function by applying DNN to capture the nonlinear or more complex correlations and interactions between features. Therefore, the distance function becomes

$$w_{ij} = \boldsymbol{w}^{1T}(\boldsymbol{v}_i \odot \boldsymbol{v}_j), \qquad D(\boldsymbol{v}_i, \boldsymbol{v}_j) = (\hat{\boldsymbol{v}}_i - \hat{\boldsymbol{v}}_j)^T(\hat{\boldsymbol{v}}_i - \hat{\boldsymbol{v}}_j), \tag{3}$$

where both $\hat{\boldsymbol{v}}_i$ and $\hat{\boldsymbol{v}}_j$ are non-linear transformations of $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ via a deep neural network. Consequently, FMs with generalized metric distance with DNN can be presented as

$$\hat{y}(\boldsymbol{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \boldsymbol{w}^{1T}(\boldsymbol{v}_i \odot \boldsymbol{v}_j)(\hat{\boldsymbol{v}}_i - \hat{\boldsymbol{v}}_j)^T(\hat{\boldsymbol{v}}_i - \hat{\boldsymbol{v}}_j) x_i x_j \tag{4}$$

3.2 Proposed Model Framework

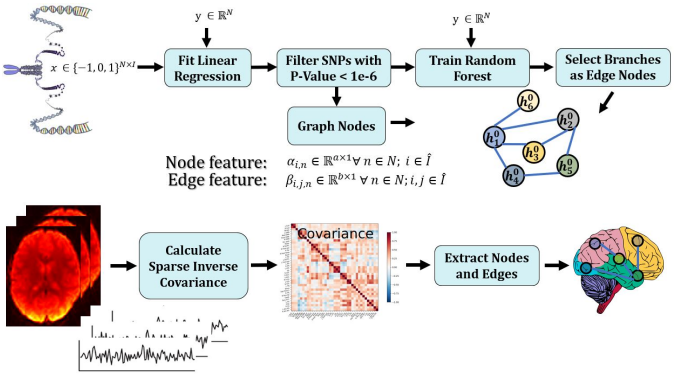

Figure 2: Overview of the proposed approach for constructing gene (top) and brain (bottom) graphs. For the gene graph, we use GWAS analysis of single-SNP–single-phenotype and random forest model. The brain graph is constructed using functional connectivity of fMRI and structural MRI.

Graph construction. As we focus on the development of a new GNN, we start a model presentation by definition of graph G with node features $\alpha_i \in \mathbb{R}^{a \times 1}$ for each node $i \in \mathcal{V}$ and (optionally) edge features $\beta_{ij} \in \mathbb{R}^{b \times 1}$ for each adjacent nodes i and j. Here, $\mathcal{V}$ is a set of nodes within the graph, and a and b denote feature dimensions of nodes and edges. Fig. 2 illustrates the construction process of

graphs. The gene graph is constructed using GWAS analysis and random forest (Ho, 1995) model sequentially to identify nodes and determine edges, respectively. First, a linear regression analysis is conducted at each SNP-by-phenotype pair to examine its genetic effect on each phenotype to measure the contribution of each SNP on phenotype and then filter SNPs based on the p-value. We choose SNPs with the minimum p-value (p-value $\leq 1 - e6$) (Meng et al., 2020). Each gene graph's node represents one selected SNP with a one-hot representation of SNPs as feature nodes. Then, a random forest is employed to predict phenotypes by one-hot coding of selected SNPs. Then, we interpret branches of trained trees as gene graph's edges. The power of this type of graph creation is that variation in random forest's hyperparameters leads to different graph complexity. In the experiment, we demonstrate these hyperparameters' impact on a model's performance in predicting phenotype.

For constructing a brain graph, the brain is parcellated into brain regions by the atlas. All parcellated ROIs were considered as graph nodes in this graph. Then, functional connectivity as the pairwise sparse inverse covariance of the mean blood-oxygen-level-dependent (BOLD) time series activity of each ROI is computed. In this research, we use sparse inverse covariance instead of Pearson's correlation as Liégeois et al. (2020) showed precision-based functional connectivity yields a better match to brain structural connectivity than correlation-based functional connectivity. Therefore, a covariance matrix was constructed for each subject, containing covariance coefficients of pairwise ROIs. A sparsity threshold can be applied on top of this matrix to define edges between ROIs. FMRI and structural MRI (sMRI) of each subject are stacked together to specify the feature nodes in a brain graph. Feature nodes are Freesurfer (Fischl, 2012) derivatives of sMRI, such as cortical thickness, volume, surface area, and subcortical volume.

Model. This paper jointly passes neuroimaging and genetic data after transformation to graph representation to BIG-Graph architecture (shown in Fig. 1). Before feeding gene and brain graphs to the graph neural network, their input features α_i and β_{ij} for both graphs are embedded to d-dimensional hidden features $h_i^{\ell=0}$ and $e_{ij}^{\ell=0}$ by a simple linear transformation $h_i^0 = U^0\alpha_i + u^0 \;\; ; \;\; e_{ij}^0 = V^0\beta_{ij} + v^0$, where $U^0 \in \mathbb{R}^{d\times a}$, $V^0 \in \mathbb{R}^{d\times b}$ and $u^0, v^0 \in \mathbb{R}^d$.

Then, d-dimensional representations of the nodes/edges are passed parallelly to the two GNNs with batch normalization and residual connections. Message-passing GNN layer updates feature representations through recursive neighborhood diffusion, where each graph node inherits features from its adjacent nodes. Stacking L GNN layers enables models to gather node representations from each node's L-hop neighborhood. We augment each GNN layer with batch normalization (BN) (Ioffe & Szegedy, 2015) and residual connections (He et al., 2016). GNN layer can be replaced with any GNN structures, such as GCN (Kipf & Welling, 2016), GraphSage (Hamilton et al., 2017), and GAT (Veličković et al., 2017). Then, convolutional feature maps of the last layer from two networks are fed into the aggregation layer by applying average pooling over graph nodes to compute latent representations of all nodes, followed by passing them to the multilayer perceptron (MLP). Additionally, the last convolutional feature maps of both gene and brain networks are passed into FMs with generalized metric distance and a deep neural network to quantify and formulate interactions between latent representations of SNPs and brain ROIs. As a final step, the resultant of FMs network and MLP of two graphs' latent representations are concatenated and connected to a linear classifier or regressor.

4 EXPERIMENT

4.1 DATASET

This study was conducted on the PNC cohort (Satterthwaite et al., 2016) by assessing 2304815 SNPs, 1071 participants with rs-fMRI and sMRI acquisitions, and 104 psychiatric and cognitive traits. Genotypes and phenotypes were acquired through the database of dbGaP (Satterthwaite et al., 2016; 2014), and bed, bim, and fam files were used to extract individualized genotyping data. This research considers age, sex, and height in inches as targets for prediction purposes. Demographics of the subjects in each target are presented in Table 1. The performance of the proposed model was tested over 1071 subjects (including 649 females) with ages ranging from 8 to 21 years (mean age 14.6 years).

Table 1: Demographics of the subjects included from PNC cohort.

Dataset	Number	Mean age (SD)	Age range	Mean height (SD) in inches	height range in inches
PNC	1071	14.61 (3.45)	(8-21)	62.63 (6.33)	(40-76)
Female	649	14.82 (3.49)	(8-21)	61.97 (5.37)	(43-74)
Male	547	14.36 (3.40)	(8-21)	63.43 (7.25)	(40-76)

4.2 BASELINE METHODS AND EXPERIMENTAL SETTINGS

For generating a brain graph, the brain was parcellated into brain regions by the Destrieux atlas (Fischl et al., 2004; Destrieux et al., 2009) with spherical 148 ROIs for cortical regions and Harvard-Oxford parcellations from FSL (Makris et al., 2006; Frazier et al., 2005; Desikan et al., 2006; Goldstein et al., 2007) with 21 spherical ROIs for subcortical regions. The time series of rs-fMRI acquisition for each ROI was extracted from the preprocessing data using a robust preprocessing pipeline (fMRIPrep) Esteban et al. (2019) to construct the functional networks. Therefore, a 169*169 covariance matrix for each subject was computed using the GraphLassoCV function of nilearn package (Abraham et al., 2014) on pairwise blood-oxygen-level-dependent (BOLD) time series activity of ROIs. To define edges between ROIs, four sparsity thresholds, including ± 0.2, ± 0.3, ± 0.4, and ± 0.5 were applied on top of these matrices to generate four brain graphs with different complexity levels. Freesurfer (Fischl, 2012) derivatives of sMRI, such as cortical thickness, volume, surface area, and subcortical volume, were considered as feature nodes of graphs. Thickness, volume, surface area were normalized globally over all participants.

To construct a gene graph, we ran 2,304,815 linear models to measure the contribution of each SNP from the PNC cohort on three phenotypes (sex, age, and height). Genomic variants with minor allele frequency (MAF) $\leq 0.01\%$ and p-value less than $1e-6$ were filtered out. Then, the random forest model identified the gene graph's edges using these selected SNPs. Each node in the gene graph represented one selected SNP with a one-hot representation of SNPs as feature nodes. As the random forest model's hyperparameters influence the generating graph, we considered four levels for the maximum depth of tree in the random forest model, including 5, 10, 15, and 20, to build four different graphs for genes with different complexity levels. Linear regression and random forest were implemented in Python using the Sklearn package (Pedregosa et al., 2011).

Table 2: Summary statistics of gene graphs (brain graphs) in different complexity levels included in the experiment.

Phenotype (Task)	**#Graphs**	**#Nodes**	**Total #Nodes**	**#Edges**	**Total #Edges**
Age (Regression)	1K (1K)	8-490 (169)	311K (181K)	53-580 (1-1K)	407K (5.8K)
				56-4.4K (2-3.4K)	2.7M (53K)
				56-8.8K (2-5.9K)	5.4M (303K)
				56-12K(2-8.6K)	7.6M (1.3M)
Height (Regression)	725 (725)	6-1.2K (169)	272K (122K)	27-969 (2-1K)	379K (4.3K)
				30-5K (2-3.4K)	1.9M (38K)
				30-9.9K (2-5.9K)	3.9M (216K)
				30-14K (2-8.6K)	5.8M (968K)
Sex (Classification)	1K (1K)	76-2.5k (169)	483K (181K)	107-682 (2-1K)	466K (5.8K)
				451-3K (2-3.4K)	2.1M (53K)
				569-4.7K (2-5.9K)	3.4M (303K)
				616-4.9K (2-8.6K)	3.6M (1.3M)

To assess the models' performance and benchmark them, we used a 10-fold cross-validation scheme. The training set is divided into 90% for updating weights and 10% for stopping criteria and reducing the learning rate when a metric has stopped improving on a validation set. To evaluate the effect of the complexity of graphs on prediction efficiency, we stacked four gene graphs with four brain graphs in terms of the level of complexity. The summary statistics of graphs included in the experiment are reported in Table 2. We benchmarked the prediction performance of the proposed GNN method with state-of-the-art message passing-based networks such as GCN, GraphSage, and GAT trained separately on neuroimaging or genetic data to illustrate the impact of the new framework in enhancing the model's performance. We carried out the experiments with Pytorch (Paszke et al., 2019) and deep graph library (DGL) (Wang et al., 2019) in Python. Adam optimizer was used to train a model with adaptive learning rates (Kingma & Ba, 2014). Stepwise learning rate decay was also applied if the validation loss stopped decreasing, with the smallest learning rate of $1e-6$. For all GNNs models in

the benchmark, we applied 32 convolutional layers followed by 2 fully connected layers, with 128 features graph convolutional and embedding layers.

5 RESULTS

To evaluate the effectiveness of the proposed framework, we conducted experiments on the prediction of age (regression problem), height (regression problem), and sex (classification problem) using the PNC cohort to answer the following three questions: Q1: The performance of BIG-Graph compared with state-of-the-art methods. Q2: The effectiveness of graph complexity in prediction outcomes. Q3: The explainable insights of combining gene and brain graphs in the BIG-graph framework.

Table 3: Benchmarking results for the BIG-graph and GNN models separately using gene and brain graphs. Mean/std of metrics are reported over test sets of 10-fold cross-validation.

Model	**Complexity**	**Age**		**Height**		**Sex**	
		MAE	RRMSE%	MAE	RRMSE%	ACC	AUC
BIG-graph GAT	Sparse	2.5 / 0.14	20.41 / 0.7	4.53 / 0.46	11.71 / 0.9	87.31 / 0.8	0.874 / 0.008
	Semi-sparse	2.36 / 0.08	19.38 / 0.7	4.3 / 0.24	10.99 / 0.6	88.13 / 0.8	0.887 / 0.044
	Semi-dense	2.33 / 0.09	19.33 / 0.4	4.18 / 0.32	10.77 / 0.2	88.69 / 0.4	0.891 / 0.048
	Dense	2.35 / 0.08	19.38 / 0.7	4.21 / 0.39	10.79 / 0.3	88.9 / 0.6	0.889 / 0.028
BIG-graph GraphSage	Sparse	2.55 / 0.11	20.57 / 0.6	4.56 / 0.13	11.81 / 0.2	87.14 / 0.2	0.87 / 0.018
	Semi-sparse	2.4 / 0.12	19.48 / 0.6	4.2 / 0.14	10.84 / 0.2	88.43 / 0.1	0.871 / 0.017
	Semi-dense	2.29 / 0.18	19.08 / 0.4	4.14 / 0.33	10.71 / 0.7	88.58 / 0.7	0.889 / 0.035
	Dense	2.39 / 0.08	19.42 / 0.9	4.23 / 0.34	10.85 / 0.5	88.79 / 0.5	0.894 / 0.038
BIG-graph GCN	Sparse	2.47 / 0.14	19.98 / 0.5	4.42 / 0.48	11.35 / 0.4	88.04 / 0.6	0.884 / 0.04
	Semi-sparse	2.24 / 0.2	18.96 / 0.7	4.27 / 0.26	10.93 / 0.8	90.5 / 0.6	0.899 / 0.012
	Semi-dense	**2.22 / 0.14**	**18.62 / 0.3**	**4.1 / 0.2**	**10.21 / 0.5**	90.82 / 0.4	0.899 / 0.047
	Dense	2.31 / 0.11	19.27 / 0.7	4.13 / 0.23	10.6 / 0.4	**91.05 / 0.1**	**0.91 / 0.027**
GAT Gene	Sparse	3.78 / 0.19	30.76 / 0.9	6 / 0.12	15.37 / 0.4	74.06 / 0.7	0.745 / 0.022
	Semi-sparse	3.54 / 0.13	29.06 / 0.4	5.71 / 0.16	14.39 / 0.9	75.09 / 0.6	0.758 / 0.01
	Semi-dense	3.55 / 0.16	29.02 / 0.4	5.57 / 0.16	14.17 / 0.2	75.45 / 0.7	0.743 / 0.034
	Dense	3.67 / 0.13	29.57 / 0.2	5.53 / 0.34	14.14 / 0.8	75.21 / 0.8	0.755 / 0.032
GAT MRI	Sparse	3.45 / 0.13	28.34 / 0.3	5.04 / 0.39	13.07 / 0.6	78.32 / 0.3	0.785 / 0.035
	Semi-sparse	3.06 / 0.08	24.24 / 0.4	4.9 / 0.13	12.46 / 0.6	79.4 / 0.3	0.79 / 0.03
	Semi-dense	3.07 / 0.15	24.14 / 0.3	4.75 / 0.34	12.52 / 0.8	79.65 / 0.9	0.791 / 0.023
	Dense	3.2 / 0.15	25.36 / 0.3	4.69 / 0.12	12.37 / 0.8	79.81 / 0.1	0.782 / 0.022
GraphSage Gene	Sparse	3.84 / 0.14	30.86 / 0.6	6.06 / 0.52	15.49 / 0.3	74.17 / 0.4	0.743 / 0.008
	Semi-sparse	3.62 / 0.13	29.31 / 0.7	5.62 / 0.33	14.76 / 0.8	74.71 / 0.8	0.745 / 0.044
	Semi-dense	3.45 / 0.14	28.13 / 0.6	5.57 / 0.5	14.19 / 0.2	75.37 / 0.2	0.755 / 0.005
	Dense	3.64 / 0.16	29.2 / 0.4	5.65 / 0.39	14.52 / 0.5	75.42 / 0.8	0.768 / 0.02
GraphSage MRI	Sparse	3.38 / 0.2	27.05 / 0.5	5.16 / 0.37	13.33 / 0.4	78.53 / 0.3	0.792 / 0.027
	Semi-sparse	3.29 / 0.16	26.79 / 0.5	4.68 / 0.39	12.08 / 0.2	79.11 / 0.7	0.782 / 0.035
	Semi-dense	3.06 / 0.14	24.37 / 0.5	4.74 / 0.26	11.83 / 0.4	79.83 / 0.3	0.8 / 0.049
	Dense	3.16 / 0.14	24.97 / 0.4	4.74 / 0.21	11.96 / 0.3	79.78 / 0.5	0.79 / 0.022
GCN Gene	Sparse	3.76 / 0.19	30.09 / 0.2	5.82 / 0.96	14.95 / 0.5	74.74 / 0.6	0.75 / 0.013
	Semi-sparse	3.4 / 0.11	28.08 / 0.8	5.68 / 0.61	14.79 / 0.2	76.85 / 0.4	0.76 / 0.011
	Semi-dense	3.47 / 0.18	27.95 / 0.3	5.35 / 0.37	13.92 / 0.5	77.19 / 0.2	0.771 / 0.03
	Dense	3.59 / 0.19	28.94 / 0.5	5.37 / 0.18	13.83 / 0.3	77.23 / 0.3	0.782 / 0.041
GCN MRI	Sparse	3.25 / 0.19	26.31 / 0.7	4.86 / 0.38	12.52 / 0.7	79.13 / 0.8	0.79 / 0.004
	Semi-sparse	2.99 / 0.11	24 / 0.7	4.76 / 0.13	12.1 / 0.4	81.25 / 0.3	0.815 / 0.009
	Semi-dense	2.94 / 0.16	23.84 / 0.6	4.59 / 0.51	11.85 / 0.9	81.72 / 0.1	0.81 / 0.038
	Dense	3.1 / 0.18	24.85 / 0.6	4.63 / 0.42	12.01 / 0.9	81.89 / 0.2	0.821 / 0.011

Table 4: Benchmarking results for the proposed model and State-of-the-art. Mean/std of metrics are reported over test sets of 10-fold cross-validation.

Model	**Age**		**Height**		**Sex**	
	MAE	RRMSE%	MAE	RRMSE%	ACC	AUC
BIG-graph GCN (Semi-dense)	**2.22/0.14**	**18.62 / 0.3**	**4.1 / 0.2**	**10.21 / 0.5**	**90.82 / 0.4**	**0.899 / 0.047**
Ko et al. (2022)	3.82 / 0.4	30.05 / 0.5	6.16 / 0.37	16.33 / 0.4	71.38 / 0.4	0.703 / 0.09
Venugopalan et al. (2021)	4.19 / 0.27	32.75 / 0.3	6.2 / 0.85	17.54 / 0.6	70.14 / 0.9	0.706 / 0.012
Zhou et al. (2019b)	4.23 /0.21	33.20 / 0.4	6.84 / 0.7	17.89 / 0.3	71.02 / 0.3	0.711 / 0.02

Prediction errors for three prediction tasks over test sets of 10-fold cross-validation using three algorithms with three different GNN structures are summarized in Table 3. Comparison in terms of the relative root mean square error (RRMSE), the mean absolute error (MAE) for regression and accuracy, and the area under the receiver operator characteristic curve (AUC-ROC) for classification reveals that the proposed model outperformed these two GNNs trained separately on large-scale neuroimaging or genetic data over gene graph and brain graph specifically for all graph configurations. As such, the different performances of our model and others can be attributed to how our model could better capture the complicated graph features from brain and genetics data and formulate feature interactions. The GCN layer yields better performance compared to GAT and GraphSage layers in terms of the graph convolution layer. Models' performances over various graph complexities suggest

that models performed better using semi-dense graphs in regression and dense in classification problems. Therefore, it highlights the role of a model's structure in prediction rather than the complexity of graphs. Based on our comparison of the performance of GNNs using gene and brain graphs separately, we can see that the brain graph was more successful at predicting age, height, and sex.

The proposed model provided more accurate predictions than most alternative methods because of its architecture. Although the proposed model inherently uses more information (genetics and neuroimaging) compared to other models trained separately on neuroimaging or genetic data, it could help better understand the complex and interactive relationship between genetics and neuroimaging. One advantage of the proposed model is to quantify the contribution of genetic information, brain features, and their interaction on targets. We define contributions as the multiplication of the feature vector and its weight in the last linear classifier or regressor in the proposed model. We visualized percentage contribution from the best-trained model (BIG-graph with GCN layer) as violin plots in Fig 3 for three feature groups. The size of the violin plot is denoted as the contribution of features to our three targets, and each point refers to one participant. Although their contributions are changed from participant to participant, high-impact features are brain features. Although the performances of GNNs using gene and brain graphs are close to each other, the joined GNN structure's performance indicates that genetic information is insufficient to contribute considerably to prediction due to the limited available genetic markers. Table 3 also hints that MRI-based GNN performed way better than gene-based GNN for sex.

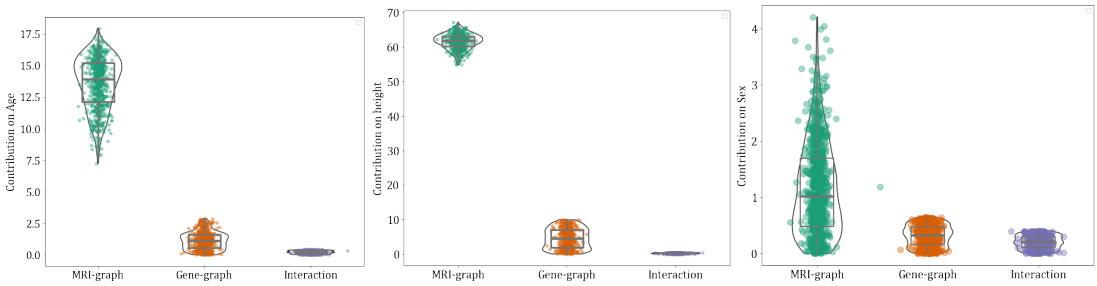

Figure 3: Violin plots of estimated percentage contributions of brain graph, gene graph, and their interaction on age (left), height (middle), and sex (right). Each dot on a violin plot represents one participant.

To quantify the importance of features, we used a game theoretical approach called SHAP (SHapley Additive exPlanation) (Lundberg & Lee, 2017). This method estimates the contribution of each feature towards a specific prediction by generating perturbations of a given instance in the dataset and estimating the impact of these perturbations on the predicted output as SHAP values, which are averaged over all possible conditions. Figures 4-6 indicate the measure of feature importance for the top 10 features in predicting age, height, and sex. The higher the SHAP value of a feature, the higher contribution to the target phenotype. Every instance in the dataset is run through the BIG-graph model and illustrated by a dot. The colors of dots are associated with the feature's value. The density of violin plots for each feature shows how different contributions for this feature were observed in the dataset.

Figure 4: Summary plot of contributions of SNPs, the volume of sub-cortical ROIs, and the thickness/volume/area of cortical ROIs on age prediction from left to right, respectively. Every participant is run through the model, creating a dot for each feature attribution value. The colors of dots for each feature are associated with the magnitude of feature values.

Figure 4 shows the contributions of features on estimated age. SHAP values indicate increased volume of sub-cortical ROIs is associated with older age. In contrast, decreased thickness, volume, and surface area in specific cortical ROIs are positively impact age estimation. Genetic markers (SNPs) values were 0, 1, and 2 (where 0 is homozygous for the first allele, 1 indicates heterozygosity, and 2 represents being homozygous for the second allele). Here, the results show heterozygous SNPs at these locations contributed most to the prediction of higher age.

Figure 5 reveals feature contributions to estimated height. Subcortical and cortical contributions to height mirror similar findings as in the age models. The observed trend of long tails reaching to the left of cortical ROIs reveals that greater thickness, volume, and area of cortical ROIs greatly impact the estimation of individuals' height compared with decreased thickness, volume, and area of cortical ROIs, and height is more sensitive to the greater thickness, volume, and area of cortical. Moreover, the results show homozygous SNP with having the second allele can push the instances to a greater height.

According to Figure 6, the extremely high and low volume of sub-cortical ROIs explicitly contributes to sex. Bigger sub-cortical ROIs tend to be male. Higher volume of sub-cortical leads to more chance of males. Lower volume leads to a higher chance of females. Hence, sub-cortical ROIs are stronger predictors of sex because of the disordering of cortical ROIs' values. SNP contribution to sex classification indicates that some genetic markers play an important role in some data points because the SHAP value is not very robust against the perturbations. Also, the results show homozygous SNP with having the second allele leads to a higher chance of being female.

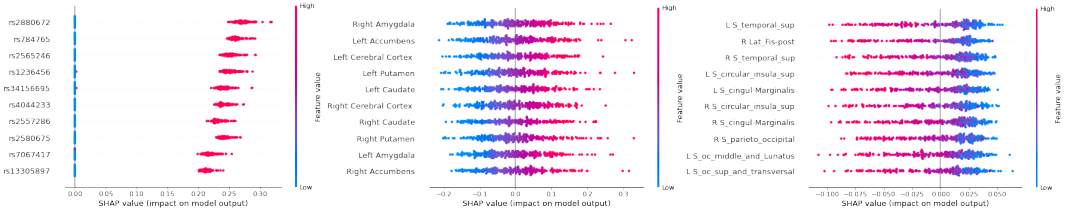

Figure 5: Summary plot of contributions of SNPs, the volume of sub-cortical ROIs, and the thickness/volume/area of cortical ROIs on height prediction from left to right, respectively. Every participant is run through the model, creating a dot for each feature attribution value. The colors of dots for each feature are associated with the magnitude of feature values.

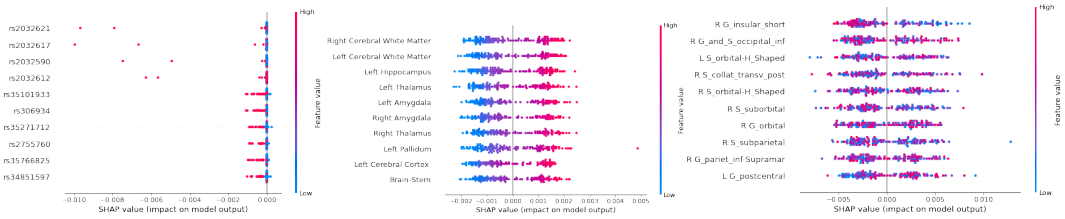

Figure 6: Summary plot of contributions of SNPs, the volume of sub-cortical ROIs, and the thickness/volume/area of cortical ROIs on sex classification from left to right, respectively. Every participant is run through the model, creating a dot for each feature attribution value. The colors of dots for each feature are associated with the magnitude of feature values.

Several challenges remain in network analysis of brain imaging genetics for future work. First, the proposed network is limited by the coverage and quality of graph definition and interaction between features. Because of variability in functional connectivity matrices and gene networks, future work must adopt a dynamic framework for updating across convolutional layers. Second, using multimodal MRI acquisition jointly in the network design increases the complexity of models and running time at least two times more than traditional models. With these limitations, our proposed model can be used in various prediction applications such as disorder and cancer studies and drug discovery. For example, cancer studies deal with multi-omics data (e.g., epigenomics, metabolomics, proteomics, and transcriptomics) from the tumor tissues. The proposed model can reveal biological insights from multi-omics data by quantifying their contribution and identifying their interaction. Incorporating prior knowledge and structure into model development is another future direction.

REFERENCES

Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, 8:14, 2014.

Javad Ansarifar and Lizhi Wang. New algorithms for detecting multi-effect and multi-way epistatic interactions. *Bioinformatics*, 35(24):5078–5085, 2019.

Danielle S Bassett and Olaf Sporns. Network neuroscience. *Nature neuroscience*, 20(3):353–364, 2017.

Danielle S Bassett, Edward Bullmore, Beth A Verchinski, Venkata S Mattay, Daniel R Weinberger, and Andreas Meyer-Lindenberg. Hierarchical organization of human cortical networks in health and schizophrenia. *Journal of Neuroscience*, 28(37):9239–9248, 2008.

Rebecca Birnbaum and Daniel R Weinberger. Functional neuroimaging and schizophrenia: a view towards effective connectivity modeling and polygenic risk. *Dialogues in clinical neuroscience*, 2022.

International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia that overlaps with bipolar disorder. *Nature*, 460(7256):748, 2009.

Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

C Destrieux, B Fischl, AM Dale, and E Halgren. A sulcal depth-based anatomical parcellation of the cerebral cortex. *NeuroImage*, (47):S151, 2009.

Lei Du, Jingwen Yan, Sungeun Kim, Shannon L Risacher, Heng Huang, Mark Inlow, Jason H Moore, Andrew J Saykin, and Li Shen. A novel structure-aware sparse learning algorithm for brain imaging genetics. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 329–336. Springer, 2014.

Juergen Dukart, Fabio Sambataro, Alessandro Bertolino, Alzheimer's Disease Neuroimaging Initiative, et al. Accurate prediction of conversion to alzheimer's disease using imaging, genetic, and neuropsychological biomarkers. *Journal of Alzheimer's Disease*, 49(4):1143–1159, 2016.

Oscar Esteban, Christopher J Markiewicz, Ross W Blair, Craig A Moodie, A Ilkay Isik, Asier Erramuzpe, James D Kent, Mathias Goncalves, Elizabeth DuPre, Madeleine Snyder, et al. fmriprep: a robust preprocessing pipeline for functional mri. *Nature methods*, 16(1):111–116, 2019.

Yong Fan, Dinggang Shen, Ruben C Gur, Raquel E Gur, and Christos Davatzikos. Compare: classification of morphological patterns using adaptive regional elements. *IEEE transactions on medical imaging*, 26(1):93–105, 2006.

Alexandru-Catalin Filip, Tiago Azevedo, Luca Passamonti, Nicola Toschi, and Pietro Lio. A novel graph attention network architecture for modeling multimodal brain connectivity. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1071–1074. IEEE, 2020.

Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

Bruce Fischl, André Van Der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H Salat, Evelina Busa, Larry J Seidman, Jill Goldstein, David Kennedy, et al. Automatically parcellating the human cerebral cortex. *Cerebral cortex*, 14(1):11–22, 2004.

Alex Fornito, Andrew Zalesky, and Michael Breakspear. Graph analysis of the human connectome: promise, progress, and pitfalls. *Neuroimage*, 80:426–444, 2013.

Jean A Frazier, Sufen Chiu, Janis L Breeze, Nikos Makris, Nicholas Lange, David N Kennedy, Martha R Herbert, Eileen K Bent, Vamsi K Koneru, Megan E Dieterich, et al. Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. *American Journal of Psychiatry*, 162(7):1256–1265, 2005.

Tian Ge, Thomas E Nichols, Debashis Ghosh, Elizabeth C Mormino, Jordan W Smoller, Mert R Sabuncu, Alzheimer's Disease Neuroimaging Initiative, et al. A kernel machine method for detecting effects of interaction between multidimensional variable sets: An imaging genetics application. *Neuroimage*, 109:505–514, 2015.

Jill M Goldstein, Larry J Seidman, Nikos Makris, Todd Ahern, Liam M O'Brien, Verne S Caviness Jr, David N Kennedy, Stephen V Faraone, and Ming T Tsuang. Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. *Biological psychiatry*, 61(8):935–945, 2007.

Yangyang Guo, Zhiyong Cheng, Jiazheng Jing, Yanpeng Lin, Liqiang Nie, and Meng Wang. Enhancing factorization machines with generalized metric learning. *IEEE Transactions on Knowledge and Data Engineering*, 2020.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Xiaoke Hao, Xiaohui Yao, Shannon L Risacher, Andrew J Saykin, Jintai Yu, Huifu Wang, Lan Tan, Li Shen, and Daoqiang Zhang. Identifying candidate genetic associations with mri-derived ad-related roi via tree-guided sparse learning. *IEEE/ACM transactions on computational biology and bioinformatics*, 16(6):1986–1996, 2018.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pp. 278–282. IEEE, 1995.

Jinlong Hu, Lijie Cao, Tenghui Li, Shoubin Dong, and Ping Li. Gat-li: a graph attention network based learning and interpreting method for functional brain network classification. *BMC bioinformatics*, 22(1):1–20, 2021.

Wen-Yu Hua, Thomas E Nichols, Debashis Ghosh, and Alzheimer's Disease Neuroimaging Initiative. Multiple comparison procedures for neuroimaging genomewide association studies. *Biostatistics*, 16(1):17–30, 2015.

Meiyan Huang, Thomas Nichols, Chao Huang, Yang Yu, Zhaohua Lu, Rebecca C Knickmeyer, Qianjin Feng, Hongtu Zhu, Alzheimer's Disease Neuroimaging Initiative, et al. Fvgwas: Fast voxelwise genome wide association analysis of large-scale imaging genetic data. *Neuroimage*, 118: 613–627, 2015.

Shih-Gu Huang, Jing Xia, Liyuan Xu, and Anqi Qiu. Spatio-temporal directed acyclic graph learning with attention mechanisms on brain functional time series and connectivity. *Medical Image Analysis*, 77:102370, 2022.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pp. 448–456. PMLR, 2015.

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, 2017.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Wonjun Ko, Wonsik Jung, Eunjin Jeon, and Heung-Il Suk. A deep generative–discriminative learning for multi-modal representation in imaging genetics. *IEEE Transactions on Medical Imaging*, 2022.

Zhaoming Kong, Lichao Sun, Hao Peng, Liang Zhan, Yong Chen, and Lifang He. Multiplex graph networks for multimodal brain network analysis. *arXiv preprint arXiv:2108.00158*, 2021.

Édith Le Floch, Vincent Guillemot, Vincent Frouin, Philippe Pinel, Christophe Lalanne, Laura Trinchera, Arthur Tenenhaus, Antonio Moreno, Monica Zilbovicius, Thomas Bourgeron, et al. Significant correlation between a set of genetic polymorphisms and a functional brain network revealed by feature selection and sparse partial least squares. *Neuroimage*, 63(1):11–24, 2012.

Xiaoxiao Li, Yuan Zhou, Nicha C Dvornek, Muhan Zhang, Juntang Zhuang, Pamela Ventola, and James S Duncan. Pooling regularized graph neural network for fmri biomarker analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 625–635. Springer, 2020.

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, 2021.

Raphael Liégeois, Augusto Santos, Vincenzo Matta, Dimitri Van De Ville, and Ali H Sayed. Revisiting correlation-based functional connectivity and its relationship with structural connectivity. *Network Neuroscience*, 4(4):1235–1251, 2020.

Jiahao Liu, Guixiang Ma, Fei Jiang, Chun-Ta Lu, S Yu Philip, and Ann B Ragin. Community-preserving graph convolutions for structural and functional joint embedding of brain networks. In *2019 IEEE International Conference on Big Data (Big Data)*, pp. 1163–1168. IEEE, 2019.

Jingyu Liu and Vince D Calhoun. A review of multivariate analyses in imaging genetics. *Frontiers in neuroinformatics*, pp. 29, 2014.

Jingyu Liu, Jiayu Chen, Nora Perrone-Bizzozero, and Vince D Calhoun. A perspective of the cross-tissue interplay of genetics, epigenetics, and transcriptomics, and their relation to brain based phenotypes in schizophrenia. *Frontiers in genetics*, pp. 343, 2018.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

David M Lydon-Staley and Danielle S Bassett. Network neuroscience: a framework for developing biomarkers in psychiatry. *Biomarkers in Psychiatry*, pp. 79–109, 2018.

Nikos Makris, Jill M Goldstein, David Kennedy, Steven M Hodge, Verne S Caviness, Stephen V Faraone, Ming T Tsuang, and Larry J Seidman. Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia research*, 83(2-3):155–171, 2006.

Xianglian Meng, Jin Li, Qiushi Zhang, Feng Chen, Chenyuan Bian, Xiaohui Yao, Jingwen Yan, Zhe Xu, Shannon L Risacher, Andrew J Saykin, et al. Multivariate genome wide association and network analysis of subcortical imaging phenotypes in alzheimer's disease. *BMC genomics*, 21 (11):1–12, 2020.

John Morris, Leslie Shaw, Beau Ances, Maria Carroll, Erin Franklin, Mark Mintun, Stacy Schneider, Angela Oliver, et al. Cascaded multi-view canonical correlation (camcco) for early diagnosis of alzheimer's disease via fusion of clinical, imaging and omic features. 2017.

Farouk S Nathoo, Linglong Kong, Hongtu Zhu, and Alzheimer's Disease Neuroimaging Initiative. A review of statistical methods in imaging genetics. *Canadian Journal of Statistics*, 47(1):108–131, 2019.

Kaida Ning, Bo Chen, Fengzhu Sun, Zachary Hobel, Lu Zhao, Will Matloff, Arthur W Toga, Alzheimer's Disease Neuroimaging Initiative, et al. Classifying alzheimer's disease with brain imaging and genetic data using a neural network framework. *Neurobiology of aging*, 68:151–158, 2018.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

Jailin Peng, Le An, Xiaofeng Zhu, Yan Jin, and Dinggang Shen. Structured sparse kernel learning for imaging genetics based alzheimer's disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 70–78. Springer, 2016.

Alain Rakotomamonjy, Francis Bach, Stéphane Canu, and Yves Grandvalet. Simplemkl. *Journal of Machine Learning Research*, 9:2491–2521, 2008.

Steffen Rendle. Factorization machines. In *2010 IEEE International conference on data mining*, pp. 995–1000. IEEE, 2010.

Baxter P Rogers, Victoria L Morgan, Allen T Newton, and John C Gore. Assessing functional connectivity in the human brain by fmri. *Magnetic resonance imaging*, 25(10):1347–1357, 2007.

Theodore D Satterthwaite, Mark A Elliott, Kosha Ruparel, James Loughead, Karthik Prabhakaran, Monica E Calkins, Ryan Hopson, Chad Jackson, Jack Keefe, Marisa Riley, et al. Neuroimaging of the philadelphia neurodevelopmental cohort. *Neuroimage*, 86:544–553, 2014.

Theodore D Satterthwaite, John J Connolly, Kosha Ruparel, Monica E Calkins, Chad Jackson, Mark A Elliott, David R Roalf, Ryan Hopson, Karthik Prabhakaran, Meckenzie Behr, et al. The philadelphia neurodevelopmental cohort: A publicly available resource for the study of normal and abnormal brain development in youth. *Neuroimage*, 124:1115–1119, 2016.

Wouter F Schmidt, Martin A Kraaijveld, Robert PW Duin, et al. Feed forward neural networks with random weights. In *International conference on pattern recognition*, pp. 1–1. IEEE Computer Society Press, 1992.

Isaac Sebenius, Alexander Campbell, Sarah E Morgan, Edward T Bullmore, and Pietro Liò. Multimodal graph coarsening for interpretable, mri-based brain graph neural network. In *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6. IEEE, 2021.

Li Shen and Paul M Thompson. Brain imaging genomics: integrated analysis and machine learning. *Proceedings of the IEEE*, 108(1):125–162, 2019.

Ren Sheng, Hyunjoon Kim, Hyeyoon Lee, Yao Xin, Yong Chen, Wen Tian, Yang Cui, Jong-Cheol Choi, Junsang Doh, Jin-Kwan Han, et al. Cholesterol selectively activates canonical wnt signalling over non-canonical wnt signalling. *Nature communications*, 5(1):1–13, 2014.

Olav B Smeland, Yunpeng Wang, Oleksandr Frei, Wen Li, Derrek P Hibar, Barbara Franke, Francesco Bettella, Aree Witoelar, Srdjan Djurovic, Chi-Hua Chen, et al. Genetic overlap between schizophrenia and volumes of hippocampus, putamen, and intracranial volume indicates shared molecular genetic mechanisms. *Schizophrenia bulletin*, 44(4):854–864, 2018.

Martijn P Van Den Heuvel and Hilleke E Hulshoff Pol. Exploring the brain network: a review on resting-state fmri functional connectivity. *European neuropsychopharmacology*, 20(8):519–534, 2010.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Janani Venugopalan, Li Tong, Hamid Reza Hassanzadeh, and May D Wang. Multimodal deep learning models for early detection of alzheimer's disease stage. *Scientific reports*, 11(1):1–13, 2021.

Maria Vounou, Thomas E Nichols, Giovanni Montana, Alzheimer's Disease Neuroimaging Initiative, et al. Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *Neuroimage*, 53(3):1147–1159, 2010.

Maria Vounou, Eva Janousova, Robin Wolz, Jason L Stein, Paul M Thompson, Daniel Rueckert, Giovanni Montana, Alzheimer's Disease Neuroimaging Initiative, et al. Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in alzheimer's disease. *Neuroimage*, 60(1):700–716, 2012.

Changqing Wang, Jianping Sun, Bryan Guillaume, Tian Ge, Derrek P Hibar, Celia MT Greenwood, Anqi Qiu, Alzheimer's Disease Neuroimaging Initiative, et al. A set-based mixed effect model for gene-environment interaction and its application to neuroimaging phenotypes. *Frontiers in neuroscience*, 11:191, 2017.

Hua Wang, Feiping Nie, Heng Huang, Shannon L Risacher, Andrew J Saykin, Li Shen, and Alzheimer's Disease Neuroimaging Initiative. Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning. *Bioinformatics*, 28(12):i127–i136, 2012.

Hua Wang, Feiping Nie, Heng Huang, and Chris Ding. Heterogeneous visual features fusion via sparse multimodal machine. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3097–3102, 2013.

Minjie Wang, Lingfan Yu, Da Zheng, Quan Gan, Yu Gai, Zihao Ye, Mufei Li, Jinjing Zhou, Qi Huang, Chao Ma, et al. Deep graph library: Towards efficient and scalable deep learning on graphs. 2019.

Xiaoqian Wang, Hong Chen, Jingwen Yan, Kwangsik Nho, Shannon L Risacher, Andrew J Saykin, Li Shen, Heng Huang, and ADNI. Quantitative trait loci identification for brain endophenotypes via new additive model with random networks. *Bioinformatics*, 34(17):i866–i874, 2018.

Guangqi Wen, Peng Cao, Huiwen Bao, Wenju Yang, Tong Zheng, and Osmar Zaiane. Mvs-gcn: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. *Computers in Biology and Medicine*, pp. 105239, 2022.

Jingwen Yan, Lei Du, Sungeun Kim, Shannon L Risacher, Heng Huang, Jason H Moore, Andrew J Saykin, Li Shen, and Alzheimer's Disease Neuroimaging Initiative. Transcriptome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm. *Bioinformatics*, 30(17): i564–i571, 2014.

Huzheng Yang, Xiaoxiao Li, Yifan Wu, Siyi Li, Su Lu, James S Duncan, James C Gee, and Shi Gu. Interpretable multimodality embedding of cerebral cortex using attention graph network for identifying bipolar disorder. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 799–807. Springer, 2019.

Chenglin Yu, Dingnan Cui, Muheng Shang, Shu Zhang, Lei Guo, Junwei Han, Lei Du, and Alzheimer's Disease Neuroimaging Initiative. A multi-task deep feature selection method for brain imaging genetics. *arXiv preprint arXiv:2107.00388*, 2021.

Xi Zhang, Jingyuan Chou, and Fei Wang. Integrative analysis of patient health records and neuroimages via memory-based graph convolutional network. In *2018 IEEE International Conference on Data Mining (ICDM)*, pp. 767–776. IEEE, 2018a.

Xi Zhang, Lifang He, Kun Chen, Yuan Luo, Jiayu Zhou, and Fei Wang. Multi-view graph convolutional network and its applications on neuroimage analysis for parkinson's disease. In *AMIA Annual Symposium Proceedings*, volume 2018, pp. 1147. American Medical Informatics Association, 2018b.

Kanhao Zhao, Boris Duka, Hua Xie, Desmond J Oathes, Vince Calhoun, and Yu Zhang. A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in adhd. *NeuroImage*, 246:118774, 2022.

Tao Zhou, Mingxia Liu, Kim-Han Thung, and Dinggang Shen. Latent representation learning for alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. *IEEE transactions on medical imaging*, 38(10):2411–2422, 2019a.

Tao Zhou, Kim-Han Thung, Xiaofeng Zhu, and Dinggang Shen. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Human brain mapping*, 40(3):1001–1016, 2019b.

Xiaofeng Zhu, Weihong Zhang, and Yong Fan. A robust reduced rank graph regression method for neuroimaging genetic analysis. *Neuroinformatics*, 16(3):351–361, 2018.

Pascal Zille, Vince D Calhoun, and Yu-Ping Wang. Enforcing co-expression within a brain-imaging genomics regression framework. *IEEE transactions on medical imaging*, 37(12):2561–2571, 2017.