

RESEARCH

Open Access



TGNet: tensor-based graph convolutional networks for multimodal brain network analysis

Zhaoming Kong^{1†}, Rong Zhou^{2†}, Xinwei Luo², Songlin Zhao², Ann B. Ragin³, Alex D. Leow⁴ and Lifang He^{2*}

[†]Zhaoming Kong and Rong Zhou contributed equally to this work.

*Correspondence: lih319@lehigh.edu

¹ School of Software Engineering, South China University of Technology, 382 Waihuan Dong Road, Guangzhou 510006, China

² Department of Computer Science and Engineering, Lehigh University, 113 Research Drive, Bethlehem 18015, PA, USA

³ Department of Radiology, Northwestern University, 737 N. Michigan Avenue, Chicago 60611, IL, USA

⁴ Department of Psychiatry, University of Illinois Chicago, 1601 W. Taylor Street, Chicago 60612, IL, USA

Abstract

Multimodal brain network analysis enables a comprehensive understanding of neurological disorders by integrating information from multiple neuroimaging modalities. However, existing methods often struggle to effectively model the complex structures of multimodal brain networks. In this paper, we propose a novel tensor-based graph convolutional network (TGNet) framework that combines tensor decomposition with multi-layer GCNs to capture both the homogeneity and intricate graph structures of multimodal brain networks. We evaluate TGNet on four datasets—HIV, Bipolar Disorder (BP), and Parkinson's Disease (PPMI), Alzheimer's Disease (ADNI)—demonstrating that it significantly outperforms existing methods for disease classification tasks, particularly in scenarios with limited sample sizes. The robustness and effectiveness of TGNet highlight its potential for advancing multimodal brain network analysis. The code is available at <https://github.com/rongzhou7/TGNet>.

Keywords: Multimodal brain networks, Tensor, Graph convolutional network, Disease classification

Introduction

In recent years, brain network analysis has attracted considerable interest in the fields of neuroscience and related sciences, as it plays an important role in the understanding of biologically fundamental mechanisms of brain function, such as how the brain sustains cognition, what signals the connections convey and how these signals affect brain regions [1, 2]. In particular, brain network analysis has been found useful in the detection of cognitive impairment and early diagnosis of several neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, and HIV-dementia [3–5]. Multimodal brain network analysis, which integrates graph-structured information from multiple neuroimaging modalities such as structural magnetic resonance imaging (sMRI), diffusion tensor imaging (DTI), and functional magnetic resonance imaging (fMRI), has emerged as a powerful approach for understanding complex brain functions [6, 7]. This integration enhances the overall understanding of brain network structures and pathologies by leveraging the strengths of each modality, thereby enabling more effective diagnosis and biomarker discovery.



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

In the analysis of multimodal brain networks, various machine learning methods have been investigated for representation learning and disease prediction from shallow to deep models, such as canonical correlation analysis (CCA) [8, 9], multiple graph kernel learning [10], tensor decomposition [11–13], and convolutional neural networks (CNNs) [14–19]. Although significant progress has been made in this field, there still lacks a general and effective model to capture multimodal aspects of brain networks. In particular, the brain network has sophisticated and non-linear structures, which may not be well captured by shallow models. CNN-based approaches specialize in processing data that has a grid-like topology such as images, which may not sufficiently capture the graph structure of brain networks. Besides, most of the traditional CNN-based approaches treat all the modalities (or channels) equitably and cannot exploit the correlation among multiple modalities.

Recently, graph convolutional networks (GCNs) have emerged as a promising approach for integrating multimodal data and learning powerful representations from graph-structured information, and they have been successfully applied to multimodal brain network analysis [20–27]. For instance, MVGCN [20] introduces a multi-view GCN to integrate brain network data from multiple neuroimaging modalities, significantly improving predictive accuracy in Parkinson's disease. Similarly, MVS-GCN [21] employs a prior brain structure learning-guided multi-view GCN framework to fuse different brain network modalities, enhancing autism spectrum disorder diagnosis. However, these approaches rely heavily on prior knowledge to define a common feature space, which limits their applicability across broader scenarios. Approaches like Mask-GNN [22] attempt to simplify cross-modal feature interactions through an edge-masking strategy, but they fall short in capturing the intricate dependencies between different brain network modalities. Similarly, while SGCN [23] integrates multimodal regional neuroimaging data into a single graph and enhances GCN interpretability by employing both node-masking and edge-masking strategies to emphasize key nodes and edges, it still struggles to fully capture the complex interactions inherent in multimodal data. Additionally, methods that combine shallow and deep learning models, such as GCN-SVM [24], which pairs GCN for node embedding with SVM for classification, have shown potential in improving performance for brain disease diagnosis. However, these approaches often rely on predefined graph structures or K-nearest neighbors, which limits their ability to generalize to more complex multimodal datasets. In light of these limitations, fully leveraging multimodal brain network data remains a challenge, particularly in addressing data alignment errors and reducing dependence on manual feature selection. Therefore, developing a more generalized GCN model for multimodal brain network analysis, capable of effectively capturing the inherent graph structures of multimodal data without relying on predefined knowledge, is crucial for enhancing adaptability, robustness, and the overall effectiveness of brain network analysis.

In this paper, we introduce TNet, a general tensor-based GCN framework for multimodal brain network analysis. Our approach provides a novel way to model multimodal brain networks globally by leveraging tensor representations and multiplex GCNs. First, we stack multimodal brain networks from all subjects into a high-dimensional tensor and apply Higher-Order Singular Value Decomposition (HOSVD) for multilinear tensor projection, which captures both the homogeneity and structural equivalence of

multimodal networks while preserving essential information. Next, the projected tensor is used to construct a graph structure through a K-Nearest Neighbor (KNN) algorithm, defining relationships between nodes. This graph structure, along with the tensor features, is then fed into the multi-GCNs for learning representations across modalities. Finally, a modality pooling layer integrates these representations by assigning weights to each modality, and pooled features are passed through a fully connected network for disease classification. This process effectively combines tensor projection with GCN to achieve comprehensive multimodal integration.

The main contributions of this paper are summarized as follows:

- We introduce a novel tensor-based graph convolutional network (TGNet) for multimodal brain network analysis. TGNet leverages tensor representation to formulate multiplex GCNs, capturing both the homogeneity of multimodal data and the inherent properties of graphs. By learning the graph structure through projection pursuit, TGNet simplifies the process using a straightforward multilinear tensor projection.
- Our model extends GCNs to multimodal data using tensor representation, addressing the challenge of predefining the population graph structure. This structure is often ambiguous due to the inherent complexity of intra-graph and inter-graph connections.
- We evaluate the effectiveness of the proposed TGNet model on three challenging real datasets (HIV, Bipolar disorder, Parkinson's disease, and Alzheimer's disease) for disease classification tasks. The results show that TGNet delivers highly competitive performance compared to existing thirteen methods.

The remainder of the paper is organized as follows. In “[Related work](#)” section, we review related work on multimodal brain network analysis. In “[Methodology](#)” section, we first describe the problem formulation and then present our model along with the corresponding learning algorithm. In “[Experiments](#)” section, we conduct a comprehensive experimental analysis to validate and justify the effectiveness of the proposed method. Finally, in “[Conclusions](#)” section, we conclude the paper based on our findings.

Related work

Tensor-Based Multimodal Brain Network Analysis. Multimodal brain network data can be effectively represented using tensors, which capture complex, multi-dimensional relationships across modalities. Tensor-based methods often employ tensor decomposition and factorization techniques to extract informative features from such high-dimensional data. For example, MPCA [28] is a general multilinear principal component analysis approach for feature extraction from tensor objects. It has been applied to concatenate multimodal brain networks into a single unified tensor, enabling effective extraction of features for each subject across both modalities and individuals [29]. MIC [30] first used the kernel-based similarity matrices to form an initial tensor across multiple modalities, followed by CP decomposition to extract feature representation for each subject. In [31], the authors proposed a multi-view clustering method using t -product tensor factorization with sparse and low-rank constraints to capture high-order correlations in multi-view data. Moreover, [32] introduced a multi-view clustering approach

with graph embedding (MCGE) by modeling multi-view graph data as tensors, using tensor factorization to jointly learn graph embeddings and improve clustering on multi-modal brain networks. In M2E [33], partial symmetric tensor decomposition is explored to learn the multi-view multi-graph embeddings for multimodal brain network analysis. Recent developments have advanced these approaches further. For example, [34] proposed a multi-view functional brain network (FBN) fusion strategy for brain disease identification, which stacks adjacency matrices from multiple FBNs estimation methods into a third-order tensor and applies tensor factorization to derive joint embeddings, capturing relationships across modalities [35] proposed a slow-thinking module that constructed a knowledge graph using tensor decomposition to further refine the multi-modal integration for brain disease diagnosis.

Despite these advancements, a common limitation of existing tensor-based methods is their inability to fully exploit the deep graph structures present in brain networks. While tensor decomposition can effectively model high-dimensional relationships, these approaches often fail to capture the non-linear dependencies and complex topological features that are crucial for brain network analysis. Specifically, non-linear dependencies refer to the intricate, non-linear interactions across different data modalities, which are often oversimplified by conventional tensor methods. Additionally, complex topological features describe the hierarchical and multi-level connectivity patterns within brain networks that are essential for accurately modeling neural structure and function. However, conventional tensor-based multimodal brain network methods often struggle to adequately model these intricate relationships due to their reliance on multilinear assumptions and the limited ability to capture complex topological information. This results in a lack of sensitivity to subtle yet critical interactions between different modalities, potentially leading to an incomplete understanding of brain connectivity and function.

GCN-based Multimodal Brain Network Analysis. In recent years, GCNs [15, 36, 37] have received growing attention and demonstrated to be useful in multimodal brain network analysis. For instance, [20] devised a multi-view GCN (MVGCN) method, which requires prior knowledge of geometric coordinates to define a common feature representation space, yet this kind of information is not always available and is often difficult to obtain in multimodal cases [38] introduced the attention mechanism to combine the multimodal features acquired by GCN, but it requires vectorized features as input, which may result in exceptionally long vectors and ignore the graph structural information of brain networks. More recently, [39] proposed a cross-modal distillation method to capture inter-modal dependencies through graph learning and mutual learning mechanisms. However, it depends on high-quality multi-modal graph mapping and cross-distillation, which may cause low-quality learning in a few modalities to affect the overall framework performance [40] merged the multi-modal data processed by their respective GNN branches at the level of node vectors and adjacency matrices. Yet, it didn't utilize the correlations during the feature extraction process across different modalities, thereby constraining the extent of multi-modal information integration. Additionally, [26] introduced an interpretable GNN model for multimodal brain network analysis, focusing on identifying disorder-specific biomarkers in connectome data. However, its reliance on small neuroimaging datasets limits scalability, potentially affecting performance on larger multimodal

datasets [25] integrated sMRI and fMRI data by using features from the structural Graph to enhance edge weights in the functional Graph, enabling more accurate neuropsychiatric disorder classification through a multi-layer GCN, but it depends on handcrafted features and predefined structural graphs, limiting its scalability [21] proposed a multi-view GCN framework for brain disorder diagnosis, integrating graph structure learning and multi-task embedding to unify graph representations across views. However, its effectiveness depends on well-aligned multi-view data, which can limit performance in datasets with higher variability or incomplete views [21] proposed a multi-view GCN framework for brain disorder diagnosis, which integrates graph structure learning and multi-task embedding to capture functional subnetworks across modalities. However, it relies heavily on geometric prior knowledge, limiting their applicability. Similarly, [24] introduced a multi-modal GCN method that jointly embeds fMRI and DTI data for improved brain disorder diagnosis, leveraging prior knowledge of brain structures. However, its dependency on structured geometric priors limits flexibility. Moreover, [23] developed a sparse interpretable GCN model for multi-modal brain disease diagnosis by learning importance probabilities for brain regions and connections across imaging modalities. Nonetheless, it simply combines multimodal ROI-based features into a single brain network and falls short in handling complex interactions between modalities. Lastly, [22] developed an interpretable MaskGNN model for multimodal brain connectivity analysis, integrating fMRI, DTI, and sMRI to enhance brain structure-function understanding. However, the model relies on consistent neuroimaging alignment, which may constrain its applicability to multimodal datasets with higher cross-modality variance.

Recent research has also explored combining tensor and GCN models for various tasks [41–47]. For example, [41] discovered overlapping functional brain networks by using tensor decomposition before feeding into GCNs. The Kronecker sum operation was utilized in [42], and t -product was employed in [43, 44] to handle dynamic graphs and multi-relational graphs, respectively. More recently, [48] utilized low-rank tensor approximation to optimize the functional connectivity network generated from an fMRI image [47] leverages DTI as the data preprocessing method for GCN with the self-attention mechanism. Additionally, [27] employed adversarial decomposed-VAE to fuse brain structure and function for analyzing cognitive impairments, although its reliance on accurate modality alignment can hinder its effectiveness when data is misaligned [49] introduces tensor-based complex-valued graph neural network for dynamic coupling multimodal brain networks.

While existing GCN-based multimodal methods have made notable progress, they continue to face significant challenges, such as reliance on geometric prior knowledge, limited ability to handle complex cross-modal interactions, and dependence on high-quality multimodal data alignment. To overcome these limitations, we introduce a framework that combines tensor representations with GCNs, allowing for the simultaneous modeling of multimodal relationships and the capture of graph structural information. This approach offers a more flexible and scalable solution for complex multimodal brain network data, leading to improved integration and performance across diverse modalities.

Methodology

Problem formulation

A brain network usually leverages a graph structure to describe interconnections between brain regions, which can be represented by a weighted graph $G = \{V, E, \mathbf{X}\}$, where $V = \{v_i\}_{i=1}^N$ is the node set indicating brain regions, $E = \{e_{ij}\}_{i,j=1}^N$ is the edge set between nodes and $\mathbf{X} \in \mathbb{R}^{N \times N}$ is the weighted connectivity matrix where x_{ij} is the corresponding edge weight. In the multimodal scenario, assume that each subject consists of M -modal brain networks $\{G_1, \dots, G_M\}$, where each one is extracted from a specific imaging modality (or measure) such as fMRI and DTI. These networks share the same set of nodes, *i.e.*, using an identical definition of brain regions, but may differ in network topology and edge weights. We consider a multimodal brain network dataset \mathcal{D} from S subjects with M different modalities. Specifically, $\mathcal{D} = \{(\{G_{1s}, \dots, G_{ms}, \dots, G_{Ms}\}, y_s)\}_{s=1}^S$, where G_{ms} represents the brain network data for the m -th modality of the s -th subject, and y_s is the corresponding class label. The primary objective of multimodal brain network data analysis is to explore the interconnections among various modalities.

The goal of multimodal brain network data analysis is to probe the interrelationships between different modalities and obtain from low-level or raw relational data higher-level descriptions of brain-behavior states to facilitate disease diagnosis or treatment monitoring.

Architecture of TGNNet

Figure 1 provides an overview of the proposed tensor-based multiplex graph convolutional network (TGNNet) model, which consists of three major components. Briefly, the first part performs cross-modality bridging with tensor decomposition, which is used to efficiently extract latent structures across modalities and individuals. The second part is the multi-GCN aggregator of TGNNet model, which adopts GCN to capture intrinsic data graph structures, and then encode relationships of different modalities. The third part involves modality pooling and prediction, where the model integrates the information from multiple modalities using trainable modality importance weights and then performs classification with a fully connected network. TGNNet can be used in both spectral and spatial domains, as the graph convolution process is equivalent regardless of the specific domain [50]. In this study, we consider a general multi-layer GCN model with the following propagation rule [36]:

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^{(l)} \mathbf{W}^{(l)}) \quad (1)$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ is the adjacency matrix of the undirected graph \mathbf{G} with self-connections \mathbf{I}_N , $\tilde{d}_{ii} = \sum_j \tilde{a}_{ij}$ and $\mathbf{W}^{(l)}$ is a layer-specific trainable weight matrix. $\sigma(\cdot)$ denotes an activation function, such as the $\text{ReLU}(\cdot) = \max(0, \cdot)$, and $\mathbf{H}^{(l)} \in \mathbb{R}^{N \times D}$ is the feature matrix of the l -th layer.

In the following, we discuss each component in detail and discuss how the form of this propagation rule can be effectively extended to multiplex models with tensor representation, thereby enabling the formation of a multimodal learning environment.

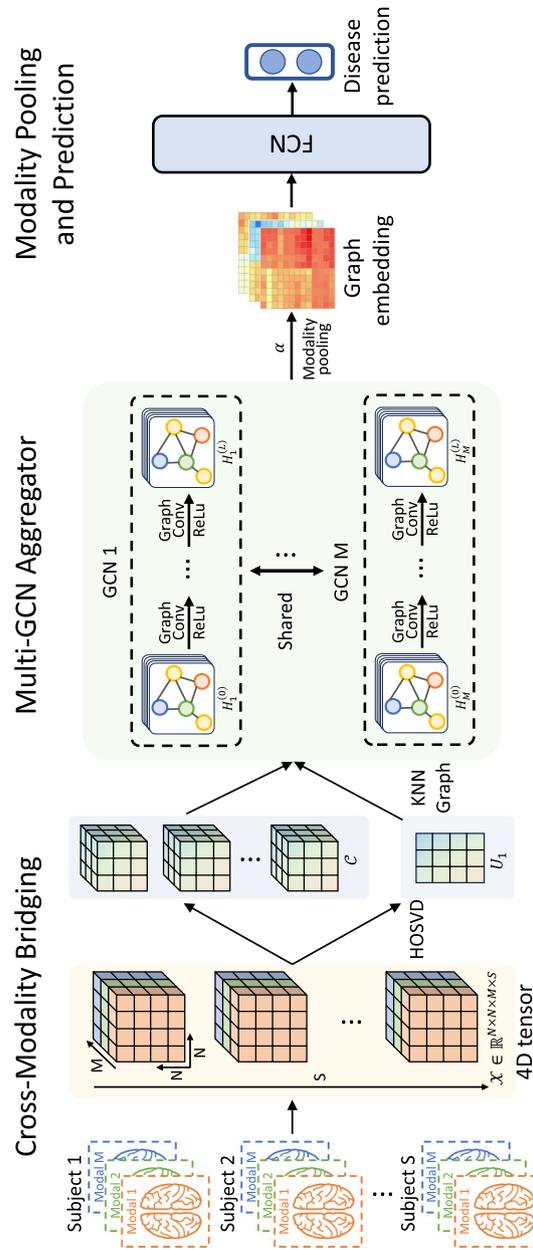


Fig. 1 An overview of our proposed TGNet framework. The framework consists of three main components: Cross-Modality Bridging, Multi-GCN Aggregator, and Modality Pooling and Prediction. In the Cross-Modality Bridging section, multimodal brain network data from S subjects and M modalities are stacked into a 4D tensor $\mathcal{X} \in \mathbb{R}^{N \times M \times M \times S}$. Higher-Order Singular Value Decomposition (HOSVD) is then applied to decompose the tensor and obtain node-level embeddings U_1 , which are used to construct a K-Nearest Neighbor (KNN) graph. In the Multi-GCN Aggregator, the KNN graph serves as input to multiple Graph Convolutional Network (GCN) layers, where graph convolution and ReLU activation are applied iteratively to learn multimodal representations of brain networks. In the final component, Modality Pooling and Prediction, the learned feature embeddings are weighted by trainable modality weights α for quality-aware fusion. The weighted features are then passed through a Fully Connected Network (FCN) for disease prediction

Cross-modality bridging

It is nontrivial to design a graph-based model to combine multimodal brain networks as these networks are extremely heterogeneous within and across individuals and modalities. Motivated by the fact that tensor analysis can effectively model a family of multi-relational data and capture the inherent heterogeneity across modalities and individuals [51–53], we propose a tensor-based projection method to map all subjects in each modality into a common space, facilitating the efficient fusion of multimodal brain networks.

Specifically, given a multimodal brain network dataset $\mathcal{D} = \{\{G_{1s}, \dots, G_{ms}, \dots, G_{Ms}\}, y_s\}_{s=1}^S$, we construct a single 4D tensor $\mathcal{X} \in \mathbb{R}^{N \times N \times M \times S}$, where N is the number of brain regions (nodes), M is the number of modalities, and S is the number of subjects. The weighted connectivity matrices is $\mathbf{X}_{ms} = \mathcal{X}(:, :, m, s) \in \mathbb{R}^{N \times N}$, where m refers to the modality index, and s refers to the subject index. To explore the uniformity of multimodal brain networks and also capture most data variation, we adopt the feature extraction method in [54] and introduce common feature projection matrices \mathbf{U}_1 and \mathbf{U}_2 by minimizing the following problem:

$$\begin{aligned} \min_{\mathbf{C}_{ms}, \mathbf{U}_1, \mathbf{U}_2} \sum_{m=1}^M \sum_{s=1}^S \|\mathbf{X}_{ms} - \mathbf{U}_1 \mathbf{C}_{ms} \mathbf{U}_2^T\|_F^2 \\ \text{s.t. } \mathbf{U}_1^T \mathbf{U}_1 = \mathbf{I} \quad \text{and} \quad \mathbf{U}_2^T \mathbf{U}_2 = \mathbf{I} \end{aligned} \quad (2)$$

where $\mathbf{U}_1 \in \mathbb{R}^{N \times d}$ characterizes node-level relationship, $\mathbf{U}_2 \in \mathbb{R}^{N \times d}$ is used for feature extraction and \mathbf{C}_{ms} is the coefficient matrix of \mathbf{X}_{ms} obtained via $\mathbf{C}_{ms} = \mathbf{U}_1^T \mathbf{X}_{ms} \mathbf{U}_2$.

The projection matrices \mathbf{U}_1 and \mathbf{U}_2 are obtained by performing higher-order singular value decomposition (HOSVD) on the tensor \mathcal{X} . Specifically, the optimization problem in Eq. (2) can be written as:

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{U}_1, \mathbf{U}_2} \|\mathcal{X} - \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2\|_F^2 \\ \text{s.t. } \mathbf{U}_1^T \mathbf{U}_1 = \mathbf{I} \quad \text{and} \quad \mathbf{U}_2^T \mathbf{U}_2 = \mathbf{I} \end{aligned} \quad (3)$$

where \times_k denotes the tensor k -mode product ($k = 1, 2$). We notice that $\mathbf{U}_1 = \mathbf{U}_2$ due to the symmetric property of \mathbf{X}_{ms} , thus we focus our analysis on the effectiveness of \mathbf{U}_1 . Intuitively, the left projection matrix \mathbf{U}_1 in Eq. (2) captures the global node-level relationship. One benefit of obtaining \mathbf{U}_1 with Eq. (2) is that it does not require any prior knowledge such as node and graph labels, thus all data can be utilized efficiently. Also, when new subjects are available, \mathbf{U}_1 can be updated in an online fashion [55].

Multi-GCN aggregator

An essential component of the propagation rule in Eq. (1) is to define a graph convolution filter in the spatial or spectral domain based on an aggregator, *e.g.*, normalized adjacency matrix $\tilde{\mathbf{A}}$. Unfortunately for brain network analysis under the multimodal environments, the common ROIs used to calculate $\tilde{\mathbf{A}}$ [20] are not always available in real-world cases. To address this issue, we notice that the i -th row of \mathbf{U}_1 encodes the weights for the i -th node in the projection $\mathbf{U}_1^T \mathbf{X}_{ms}$. Hence, by selecting the important columns of \mathbf{U}_1 according to the singular values [28], we can use the truncated \mathbf{U}_1 and

K -Nearest Neighbor (KNN) graph to define an undirected adjacency matrix to facilitate the cross-modality learning for multiplex GCNs. To be specific, we identify the set of nodes N_i that are neighbors to the node v_i using KNN and connect v_i and v_j if $v_i \in N_j$ or if $v_j \in N_i$. Mathematically, we define the adjacency matrix \mathbf{A} [20] as:

$$a_{ij} = \begin{cases} \exp(-\frac{\|\mathbf{u}_i - \mathbf{u}_j\|^2}{2\sigma^2}) & \text{if } v_i \in N_j \text{ or } v_j \in N_i, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

where \mathbf{u}_i is the i -th row of truncated \mathbf{U}_1 , and σ is the kernel width parameter. Then we can substitute $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ with $\tilde{d}_{ii} = \sum_j \tilde{a}_{ij}$ in Eq. (1).

For simplicity, let us consider the first layer of Eq. (1) with $\mathbf{H}_{ms}^{(0)} = \mathbf{C}_{ms}$ as input based on Eq. (2), and let $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}$, then in our case, the propagation rule of each input feature matrix \mathbf{C}_{ms} is:

$$\mathbf{H}_{ms}^{(1)} = \sigma(\hat{\mathbf{A}} \mathbf{C}_{ms} \mathbf{W}^{(0)}) = \sigma(\hat{\mathbf{A}} \mathbf{U}_1^T \mathbf{X}_{ms} \mathbf{U}_2 \mathbf{W}^{(0)}) \quad (5)$$

where both \mathbf{U}_2 and the shared weight matrix $\mathbf{W}^{(0)}$ are used for feature extraction, and $\mathbf{W}^{(0)}$ is obtained in an end-to-end fashion, thus in practice, \mathbf{U}_2 and $\mathbf{W}^{(0)}$ can be combined to save some computing time. According to the symmetry of $\hat{\mathbf{A}}$, we can rewrite Eq. (5) as:

$$\mathbf{H}_{ms}^{(1)} = \sigma(\mathbf{C}_{ms} \times_1 \hat{\mathbf{A}}^T \times_2 \mathbf{W}^{(0)T}) \quad (6)$$

where \times_i denotes the i -th mode product. Rearranging Eq. (6) in the tensor form, the propagation rule for all graphs at the l -th layer is formulated as

$$\mathcal{H}^{(l+1)} = \sigma(\mathcal{H}^{(l)} \times_1 \hat{\mathbf{A}}^T \times_2 \mathbf{W}^{(l)T}) \quad (7)$$

where $\mathcal{H}^{(0)} = \mathcal{C}$ and $\mathcal{H}(:, :, m, s) = \mathbf{H}_{ms}$.

Modality pooling and prediction

For subject s with M modalities, the final output of Eq. (7) are M feature matrices $\{\mathbf{H}_{1s}^{(L)}, \dots, \mathbf{H}_{ms}^{(L)}, \dots, \mathbf{H}_{Ms}^{(L)}\}$, where $\mathbf{H}_{ms}^{(L)} \in \mathbb{R}^{N \times D_{out}}$ corresponds to the m -th modality of subject s in the L -th layer, and D_{out} represents the output feature size. To integrate information of all M modalities, we add an additional modality pooling layer by introducing a 1D trainable modality importance weight $\alpha = \{\alpha_1, \dots, \alpha_m, \dots, \alpha_M\}$. The final feature embedding matrix $\mathbf{F}_s \in \mathbb{R}^{N \times D_{out}}$ for the s -th subject is calculated as a weighted combination of all modalities by

$$\mathbf{F}_s = \sum_{m=1}^M \alpha_m \mathbf{H}_{ms}^{(L)} \quad (8)$$

Notice that when all elements α are the same, then Eq. (8) boils down to the simple average pooling strategy used by the multi-view GCN (MVGCN) [20]. Leveraging tensor representation and the propagation rule in Eq. (7), we can write Eq. (8) as a straightforward and self-explanatory equation

$$\mathcal{F}_{out} = \sigma(\mathcal{H}^{(L-1)} \times_1 \hat{\mathbf{A}}^T \times_2 \mathbf{W}^{(L-1)T}) \times_3 \boldsymbol{\alpha}^T \quad (9)$$

where $\mathcal{F}_{out} \in \mathbb{R}^{N \times D_{out} \times S}$ is the output feature embeddings for all S subjects with $\mathcal{F}_{out}(:, :, s) = \mathbf{F}_s$. From the perspective of transform-domain techniques [56], $\hat{\mathbf{A}}$, \mathbf{W} and $\boldsymbol{\alpha}$ are transform matrices along the first, second and third dimension of \mathcal{H} , and the activation function in Eq. (10) can be regarded as a threshold operator that filters out small and unimportant coefficients. Furthermore, if our model contains only one GCN layer with $L = 1$, then \mathcal{F}_{out} in Eq. (9) can be obtained by

$$\mathcal{F}_{out} = \sigma(\mathcal{C} \times_1 \hat{\mathbf{A}}^T \times_2 \mathbf{W}^T) \times_3 \boldsymbol{\alpha}^T = \sigma(\mathcal{X} \times_1 (\mathbf{U}_1 \hat{\mathbf{A}})^T \times_2 (\mathbf{U}_2 \mathbf{W})^T) \times_3 \boldsymbol{\alpha}^T \quad (10)$$

Our model is efficient because it avoids directly operating on long vectors and contains only a few trainable parameters. Therefore, it can be effortlessly extended to large brain network datasets with more modalities.

Algorithm 1 Tensor-based multi-GCN (TGNet)

Input: 4D tensor $\mathcal{X} \in \mathbb{R}^{N \times N \times M \times S}$ representing multimodal brain network data.

Step 1: (Cross-Modality Bridging): Calculate the first mode projection matrix \mathbf{U}_1 of \mathcal{X} according to Eq. (2).

Step 2: (Multi-GCN Aggregator): Compute the aggregated adjacency matrix $\tilde{\mathbf{A}}$ by Eq. (4) using \mathbf{U}_1 . Perform multi-layer GCNs according to the propagation rule in Eq. (7).

Step 3: (Modality Pooling and Prediction): Conduct modality pooling to extract feature embeddings for input subjects, which are then fed to FCN for classification.

Output: Predicted class label y .

Finally, a fully connected network (FCN) with *softmax* is applied to the feature embeddings \mathcal{F}_{out} for classification. It computes the probability distribution over the labels:

$$p(y_s = j | \mathbf{f}_s) = \frac{\exp(\mathbf{w}_j^T \mathbf{f}_s)}{\sum_{k=1}^K \exp(\mathbf{w}_k^T \mathbf{f}_s)}, \quad (11)$$

where \mathbf{w}_k is the weight vector of the k -th class, and \mathbf{f}_s is the vectorized output feature embedding of subject s obtained from $\mathcal{F}_{out}(:, :, s)$. Algorithm 1 summarizes the main steps of the proposed TGNet model.

Experiments

Datasets and preprocessing

Our method is evaluated on four real-world neuroimaging datasets containing multiple modalities. Potential confounders such as age and gender were carefully addressed during the collection and preprocessing stages, ensuring balanced distributions across groups. Furthermore, consistent image acquisition procedures were applied across all datasets to maintain uniformity. Table 1 provides the statistics for the three real-world datasets, which are briefly introduced below. Additional experiments conducted on the ADNI dataset are provided in the supplementary materials for further validation, due to space constraints.

Table 1 Details of three datasets used in the experiments

| Dataset | Feature size | Modality | Class | Sample size |
|---------|--------------|-----------------------------|---------|-------------|
| HIV | 90 × 90 × 2 | fMRI & DTI | Healthy | 35 |
| | | | Patient | 35 |
| BP | 82 × 82 × 2 | fMRI & DTI | Healthy | 45 |
| | | | Patient | 52 |
| PPMI | 84 × 84 × 3 | PiCo & Hough & Probtracx | Healthy | 149 |
| | | | Patient | 569 |

Human Immunodeficiency Virus Infection (HIV): This dataset was collected from the Early HIV Infection Study at Northwestern University, across both fMRI and DTI modalities [57]. We follow the standard procedure to preprocess the dataset [58]: Here, we use the DPARSF toolbox¹ to process the fMRI data. We realigned all images to the first volume, performed the slice timing correction, and normalized them to the standard MNI template. The normalized images were spatially smoothed with an 8-mm Gaussian kernel. The final whole-brain networks for each subject were created individually by parcellating the brain into 90 cerebral regions (excluding 26 cerebellar regions) and computing pairwise connectivity over correlation coefficients.

Bipolar Disorder (BP): This dataset was collected from the UCLA Ahmanson-Lovelace Brain Mapping Center and includes 52 bipolar I subjects in euthymia and 45 healthy controls with matched age and gender, across both fMRI and DTI modalities [59]. The resting-state fMRI data was acquired on a 3T Siemens Trio scanner using a T2* echo planar imaging (EPI) gradient-echo pulse sequence with integrated parallel acquisition technique (IPAT) and DTI data were acquired on a Siemens 3T Trio scanner. The brain networks were constructed using the functional connectivity (CONN) toolbox² [60]. The raw EPI images were first realigned and co-registered, after which we performed the normalization and smoothing. Then the confound effects from motion artifact, white matter, and CSF were regressed out of the signal. Finally, the brain networks were derived using the pairwise BOLD signal correlations based on the 82 labeled Freesurfer-generated cortical/subcortical gray matter regions.

Parkinson's Progression Markers Initiative (PPMI): This dataset was obtained from the Parkinson's Progression Markers Initiative (PPMI) database³ with raw MRI and DTI images. We preprocessed the MRI acquisitions on 718 subjects as follows. T1-weighted MRI data was acquired using the ADNI-2 sequence, and processed using the FreeSurfer⁴, followed by [61]. For DTI data, each subject's raw data were aligned to the b0 image using the FSL⁵ eddy-correct tool to correct for head motion and eddy current distortions. 84 ROIs is parcellated from T1-weighted MRI using Freesurfer. Based on these 84 ROIs, we reconstruct three types of brain connectivity matrices for each subject,

¹ <http://rfmri.org/DPARSF>

² <http://www.nitrc.org/projects/conn>

³ <http://www.ppmi-info.org/data>

⁴ <https://surfer.nmr.mgh.harvard.edu>

⁵ <http://www.fmrib.ox.ac.uk/fsl>

using the following three whole-brain probabilistic tractography algorithms: Probabilistic Index of Connectivity (PICO), Hough voting and Probtrackx [61].

Baselines and metrics

To demonstrate the effectiveness of our TGNNet model, we compare it against the following thirteen baseline methods for disease classification using the multimodal HIV, BP, and PPMI datasets.

- **M2E** [33]: It is a tensor-based method for multimodal feature extraction. We apply it to obtain the embeddings of all subjects and then perform classification with FCN.
- **MIC** [30]: It first uses the kernel matrices to form an initial tensor across multiple modalities, and then CP decomposition is employed to extract feature representation for each subject. We perform classification using the same settings as above.
- **MPCA** [28]: We concatenate all data information into a 4D tensor and then apply MPCA to extract feature embeddings for each subject across modalities and individuals. We then perform classification using these features.
- **MK-SVM** [62]: It is a multiple kernel learning method dependent on the SVM classifier, where the graph kernel is calculated as the weighted sum of single modality kernels.
- **3D-CNN** [63]: For each subject, we concatenate multimodal brain networks into 3D data. We then apply 3D-CNN for joint feature extraction and classification in an end-to-end manner.
- **GAT** [64]: Similar to GCN, we vectorize the input tensor data into a 2D feature matrix $\mathbf{X} \in \mathbb{R}^{MN^2 \times S}$, and apply the graph attention mechanism.
- **GCN** [36]: We reshape the 4D tensor data $\mathcal{X} \in \mathbb{R}^{N \times N \times M \times S}$ into a 2D feature matrix $\mathbf{X} \in \mathbb{R}^{MN^2 \times S}$, where each column corresponds to the vectorized representation of 3D multimodal data $\mathcal{X}(:, :, :, i)$, in this case, each vectorized graph can be viewed as a node, and the GCN model can be directly applied.
- **DiffPool** [65]: It is a hierarchical GCN method equipped with differentiable pooling for graph classification. Since it can only handle single modal data, we apply it to each modality independently and report the best result.
- **MVGCN** [20]: It is a multi-view GCN method, which requires prior knowledge of common geometric coordinate information to define shared feature space. Typically, such information is not available in multimodal data, thus we consider obtaining the shared feature space with the average of all brain networks across modalities and subjects, and then feed it into the MVGCN architecture.
- **MVS-GCN** [21]: It uses a shared graph convolutional layer to extract multi-view graph data features of subjects for disease classification. We use the different modal data of the subject as different views to apply MVS-GCN for feature extraction and classification.
- **GCN-SVM** [24]: It is a method that uses GCNs to extract features of each ROI of the subject and uses SVM for disease classification. For each subject, we combine multimodal data into the raw representation of ROI and apply GCN-SVM for feature extraction and classification.

- **MaskGNN** [22]: It is a multi-view GNN method that draws upon the connectivity information from each view. It integrates data from different views or modalities by amalgamating them at the node level. The combined data are then fed into a GNN, where a masking mechanism is applied to identify the connections most critical for the model's predictions.
- **SGCN** [23]: It is an interpretable GCN method that introduces the sparse regional and connective important probabilities in the brain network. We apply these sparse important probabilities to multimodal data and learn the graph-level embedding for graph classification.

In order to measure the performance of all compared methods, we use Accuracy and AUC (Area Under the ROC Curve) scores as indicators of classification quality, which are two widely used evaluation metrics for disease classification in medical fields. Typically, the larger the values, the better the classification performance.

Implementation details

For all our experiments, we use binary cross-entropy loss with Adam optimizer [66] to train the deep models. We empirically set the learning rate to 0.001 and the epoch to 50 iterations. We vary the dropout rate of the graph embedding layer from 0 to 0.5, and the number of TGNNet layers from {1, 2, 3}. In our proposed model, there are three major parameters, namely the batch size B in the training stage, the neighbor number of the K nearest neighbors when building the KNN graph, and the output feature size D_{out} in GCNs. We apply the grid search to determine the optimal values of these three parameters. In particular, we empirically select D_{out} from {10, 30, 50, 70, 90, 110}, and K and B from {2, 4, 6, 8, 10, 12}. We also carefully tune the parameters of all compared methods according to the authors' suggestions using the same data splits, in which the training, validation and testing set are set as the ratio of 8 : 1 : 1. To avoid randomness, the results are averaged based on ten independent runs. All experiments are performed on an 8-core machine with 16GB RAM. The deep learning backend is Tensorflow-GPU 2.2.0 with Python 3.6.

Results and discussions

To evaluate the effectiveness of our model, we conducted a series of experiments on multimodal brain network analysis. The experiments cover various aspects, including the performance comparison with state-of-the-art methods (“[Model comparison](#)” section), the effects of multimodal learning (“[Effectiveness of multimodal learning](#)” section), the ablation study on the impact of cross-modality bridging and weighted modality pooling strategies (“[Ablation study](#)” section), and sensitivity of the hyperparameters (“[Hyperparameter analysis](#)” section), the quality of graph embeddings (“[Embedding visualization](#)” section).

Model comparison

Table 2 presents the results of all compared methods on HIV, BP, and PPMI datasets. It is important to note that the first two datasets used in our experiments consist of

Table 2 Classification performance of different methods on HIV, BP and PPMI datasets in terms of average accuracy and AUC scores

| Type | Method | HIV | | BP | | PPMI | |
|-----------------|--------------|--------------------|--------------------|--------------------|--------------------|-------------------|-------------------|
| | | Accuracy | AUC | Accuracy | AUC | Accuracy | AUC |
| Shal-low | M2E | 50.61±15.84 | 51.53±13.68 | 57.78±12.61 | 53.63±11.82 | 76.98±8.65 | 71.53±8.34 |
| | MIC | 55.63±15.28 | 56.61±13.43 | 51.21±13.78 | 50.12±16.78 | 78.03±8.36 | 72.42±7.93 |
| | MPCA | 67.24±11.56 | 66.92±12.46 | 56.92±13.33 | 56.86±13.69 | 81.25±6.34 | 72.36±7.18 |
| Deep | MK-SVM | 67.71±13.18 | 69.89±10.36 | 60.12±10.83 | 56.78±12.86 | 81.68±5.06 | 75.96±6.89 |
| | 3D-CNN | 74.31±18.81 | 73.53±16.41 | 63.33±11.21 | 61.62±10.26 | 82.24±5.32 | 75.65±6.11 |
| | GAT | 68.58±13.51 | 67.31±14.32 | 61.31±15.04 | 59.93±13.54 | 82.28±5.95 | 75.19±5.57 |
| | GCN | 70.16±12.54 | 69.94±12.91 | 64.44±15.71 | 64.24±16.45 | 83.10±6.01 | 75.33±6.79 |
| | DiffPool | 71.42±14.78 | 71.08±15.12 | 62.22±12.83 | 62.54±13.41 | 82.19±6.23 | 72.75±2.18 |
| | MVGCN | 74.29±11.27 | 73.75±12.63 | 62.22±16.83 | 62.64±16.89 | 82.54±5.15 | 76.48±6.69 |
| | MVS-GCN | 78.57±16.66 | 77.50±12.94 | 65.00±16.03 | 63.70±15.31 | 80.28±1.41 | 76.54±8.03 |
| | MaskGNN | 75.71±11.41 | 79.20±12.43 | 64.95±10.34 | 64.71±10.31 | 80.43±5.68 | 72.24±5.14 |
| | SGCN | 77.14±14.09 | 80.47±12.67 | 63.89±17.24 | 65.18±16.73 | 81.09±5.17 | 74.78±7.69 |
| | GCN-SVM | 62.86±6.55 | 63.33±15.00 | 61.67±5.15 | 57.33±8.21 | 78.87±4.86 | 70.25±9.59 |
| Shal-low + Deep | TGNet (Ours) | 81.39±13.41 | 82.08±14.81 | 67.78±12.28 | 66.31±10.24 | 83.94±6.09 | 77.93±6.86 |

fewer than 100 subjects, resulting in high standard deviations in the outcomes. The key observations from Table 2 are as follows.

Overall, the proposed TGNet demonstrates notable improvements in classification accuracy across all datasets. Specifically, TGNet achieves average accuracy improvements of 2.82%, 1.61%, and 0.84% on the HIV, BP, and PPMI datasets, respectively, compared to the second-best methods. These results indicate that TGNet consistently outperforms other models, particularly on datasets with fewer samples where high variability in results is observed. To validate the significance of these performance differences, we conducted Wilcoxon signed-rank tests, comparing accuracy and AUC metrics between TGNet and other models. The tests yielded p-values less than 0.05 across all datasets, confirming that the improvements are statistically robust.

For the HIV dataset, TGNet shows a substantial accuracy improvement of 2.82% over MVS-GCN, demonstrating its effectiveness in integrating multimodal brain networks without relying on prior knowledge of geometric coordinates. Similarly, for the BP dataset, TGNet achieves a 1.61% increase in accuracy over SGCN, highlighting its superior capability to handle multimodal data integration and capture complex interactions between brain network modalities. On the PPMI dataset, TGNet provides a notable 0.84% improvement, which, although smaller, still indicates the robustness of the proposed method across different types of brain network data.

TGNet's advantages over shallow tensor-based methods like M2E and MPCA underline the importance of effectively modeling graph structures. The shallow methods fail to capture the rich, high-dimensional relationships inherent in brain network data, leading to lower performance. In contrast, TGNet leverages the strengths of both tensor decomposition and graph convolutional networks, ensuring a more comprehensive representation of the data.

Compared to straightforward GCN and GAT methods, TGNet's superiority lies in its joint modeling of node- and modality-level relationships. This joint modeling

avoids the pitfalls of vectorizing high-dimensional data, which can lead to the loss of important structural information, and preserves the multimodal tensor structure, ensuring that all relevant information is retained and utilized. Additionally, while MaskGNN enhances edge interpretability through a masking strategy, it oversimplifies the interactions between modalities, limiting its ability to capture intricate cross-modal dependencies. Similarly, GCN-SVM relies on predefined graph structures and K-nearest neighbors, which restricts its capacity to model complex multimodal interactions. TGNNet addresses these limitations by providing a more flexible and automated solution for integrating multimodal data.

Although the 3D-CNN model attempts to build multimodal representations using convolution and pooling layers, it falls short in explicitly considering multimodal relationships and graph structure information. This limitation leads to a noticeable performance gap compared to TGNNet. The 3D-CNN model flattens the graph representation, ignoring the inherent graph structure and the inter-modality connections. TGNNet, on the other hand, encodes both the multimodal characteristics and captures the graph structures of the input brain networks, providing a more effective solution for multimodal brain network analysis.

In summary, TGNNet demonstrates obvious improvements across various datasets by effectively combining the strengths of tensor decomposition and graph convolutional networks, making it a powerful tool for multimodal brain network analysis.

Effectiveness of multimodal learning

In the proposed TGNNet model, multi-modal information is integrated using the simple and effective modality-pooling strategy. To investigate how multiple modalities affect the graph representation learning ability of TGNNet and thus the classification quality, we compare TGNNet with its fine-tuned single-modality counterpart on three datasets.

As results shown in Table 3, we can see that the multimodal learning strategy leads to certain improvements in classification performance on all three datasets. This study is very encouraging and valuable for multimodal brain network analysis as it suggests that the TGNNet is able to integrate the multimodal information that is effective for classification in multi-GCN. Another interesting observation is that our model performs better

Table 3 Classification performance of TGNNet on unimodal and multimodal brain network datasets

| Method | Dataset | Modality | Accuracy | AUC |
|--------|---------|-------------------|--------------------|--------------------|
| TGNNet | HIV | fMRI | 53.61±11.88 | 52.31±13.63 |
| | | DTI | 75.71±15.12 | 76.08±15.61 |
| | | Both | 81.39±13.41 | 82.08±14.81 |
| | BP | fMRI | 57.78±12.61 | 51.83±10.84 |
| | | DTI | 64.44±13.59 | 63.34±15.91 |
| | | Both | 67.78±12.28 | 66.31±10.24 |
| | PPMI | PICo | 79.01±6.62 | 73.66±8.69 |
| | | Hough | 80.14±5.65 | 72.03±7.42 |
| | | Probtracx | 83.39±4.34 | 76.50±8.29 |
| All | | 83.94±6.09 | 77.93±6.86 | |

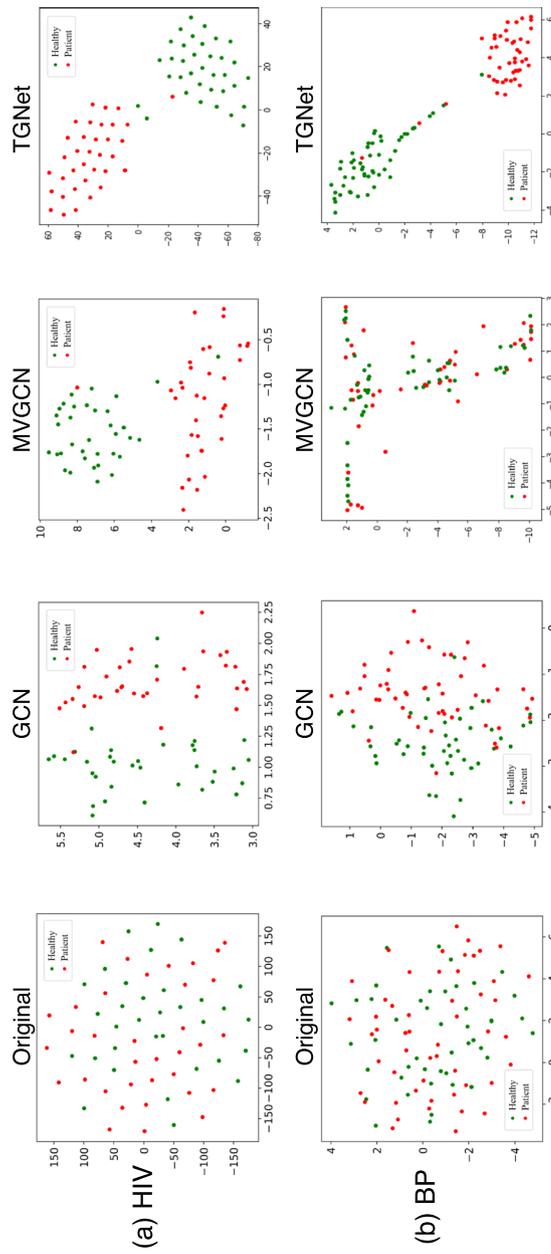


Fig. 2 a Visualization of modality weight α and (b-c) model accuracy vs. different hyperparameters of TGNet on HIV, BP and PPMI datasets

with the DTI modality compared to the fMRI modality. Consequently, the DTI modality is assigned a higher weight in the modality pooling operator α , as illustrated in Fig. 2a.

Ablation study

In TGNNet (Eq. (10)), there are two key components: U_1 , which captures node-level information through cross-modality bridging, and α , which characterizes the multimodal relationship via weighted modality pooling. To validate the contribution of these components to the superior performance of our TGNNet model, we conducted an ablation study. Table 4 compares TGNNet with two variations: TGNNet without U_1 (TGNNet- U_1) and TGNNet without weighted modality pooling (TGNNet- α). In TGNNet- U_1 , the graph is constructed directly from raw multimodal data without tensor decomposition. The results show that the introduction of the node projection matrix U_1 significantly boosts the performance of our model. For TGNNet- α , we replaced the weighted modality pooling with average pooling, which affected the model's ability to capture modality importance. Additionally, using an average pooling strategy ($\alpha = [\frac{1}{M}, \dots, \frac{1}{M}]$), as in TGNNet- α , results in slightly worse performance compared to automatic weighted modality pooling. This could be explained by the uniform distribution of weights in average pooling, which fails to capture the varying importance of different modalities. In contrast, automatic weighted pooling dynamically adjusts the weights based on the contribution of each modality, leading to a more accurate and nuanced representation of the multimodal data.

Hyperparameter analysis

We investigate the influence of three important parameters in our TGNNet model, namely the number of neighbors K when building the KNN graph, the dimensions of output features D_{out} produced by GCN and the batch size used for training. According to Figs. 2b-d, we notice that the performance of TGNNet is related to all three parameters, which should be carefully tuned based on site-specific conditions. For example, increasing the number of neighbors K does not guarantee improvements in classification performance, because most useful information is distributed in certain rows and columns, and adding more neighbors may introduce some noise.

Embedding visualization

The accuracy and quality of objective evaluations may be affected by the limited number of training and validation samples, which can make it difficult to produce reliable graph representations. Given this challenge, generating satisfactory representations for subsequent analysis is a critical task in graph-based problems. To qualitatively assess the effectiveness of TGNNet, we examine two types of embeddings in our

Table 4 Comparison of classification performance for three TGNNet variants

| Dataset | TGNNet- U_1 | TGNNet- α | TGNNet |
|---------|---------------|------------------|--------------------|
| HIV | 75.71±14.33 | 80.81±12.04 | 81.39±13.41 |
| BP | 63.33±12.52 | 66.68±10.53 | 67.78±12.28 |
| PPMI | 82.21±7.18 | 83.52±6.61 | 83.94±6.09 |

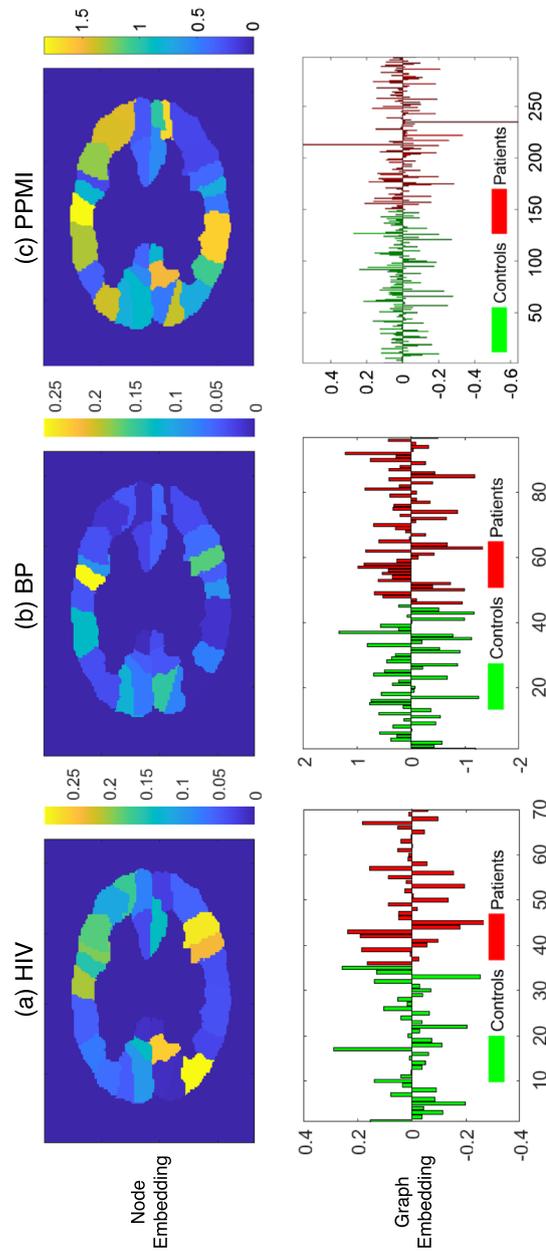


Fig. 3 Embedded node features (upper panels) and graph features (lower panels) of subjects on HIV, BP and PPMI datasets. The upper panels show the node embedding features obtained by tensor decomposition, where the coordinate system represents neuroanatomy and the color shows the activity intensity of the brain region. The lower panels show the graph embedding features which represent the factor strengths for healthy controls (green bars) and patients (red bars)

study: node embeddings (U_1) and graph embeddings (f_s). Node embedding refers to the representation of individual brain regions (nodes) within the networks, learned through tensor decomposition, capturing localized structural information. On the other hand, graph embedding refers to the representation of the entire brain network, learned through GCNs, encapsulating the relationships between all nodes to reflect the global properties of the network. Both types of embeddings provide valuable insights—node embeddings allow us to study the fine-grained brain connectivity patterns, while graph embeddings offer a more holistic view of brain network structure. Figure 3 illustrates these differences by visualizing node and graph embeddings obtained by TGNNet on HIV, BP, and PPMI datasets. The upper panels show the node embedding features for the entire population, where the color intensity reflects the activity levels in each brain region. The lower panels show the graph embedding features for healthy controls (green bars) and patients (red bars) separately. We have the following observations:

- From the upper panels, the embedded neuroanatomy learned from the HIV, BP, and PPMI datasets shows notable differences, with certain regions playing a more significant role in distinguishing the diseases. The intensity values in the node embeddings represent the strength of features derived from neuroimaging data, with higher values corresponding to stronger regional connectivity and lower values indicating weaker connectivity. For example, the highlighted yellow regions (e.g., left parietal lobes and right frontal lobes) in the HIV data suggest that these regions are crucial for characterizing brain activity in HIV patients. In contrast, the blue regions (e.g., postcentral cortex and occipital cortex) with low-intensity values indicate decreased connectivity in HIV-infected individuals, which aligns with clinical findings in the medical literature [67]. These decreased connectivity patterns align with the reduced communication efficiency often observed in neurodegenerative and neuropsychiatric disorders. For BP, we observe that the parietal lobes are impaired in bipolar disorder, which is also in line with previous studies [68, 69]. For PPMI, there are more highlighted yellow regions than HIV and BP, reflecting the complex underlying causes of Parkinson's disease.
- From the lower panels, it is evident that the graph embeddings for healthy controls (green bars) and patients (red bars) differ significantly, demonstrating that TGNNet effectively extracts discriminative representations for these two groups. Specifically, the graph embeddings for healthy controls are predominantly positive, suggesting stronger and healthier brain network structures. In contrast, the embeddings for patients display more negative values, indicating disrupted or weaker connectivity. The predominantly positive graph embeddings in healthy controls suggest greater global brain network integrity, which is consistent with well-established patterns of healthy brain connectivity. This reflects more cohesive and efficient communication between brain regions in healthy individuals. In contrast, the greater number of negative values in the graph embeddings of patients points to disrupted or weakened connectivity across the brain network, a feature commonly observed in neurodegenerative conditions such as HIV and Parkinson's disease. These disruptions in network structure may correspond to reduced com-

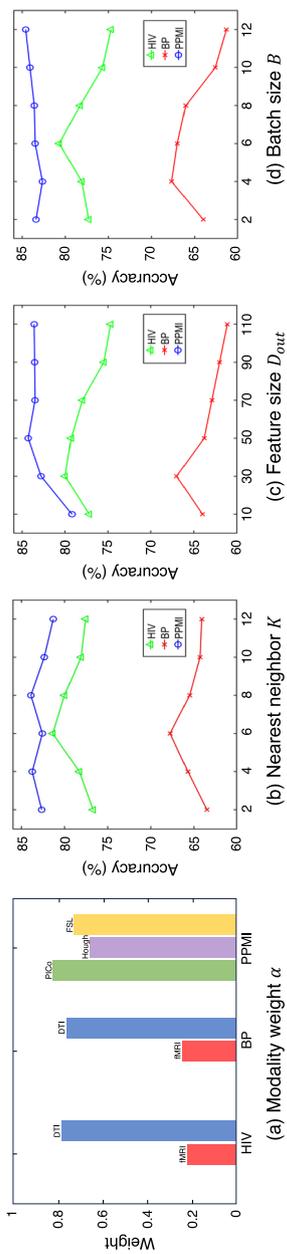


Fig. 4 A t-SNE visualization of the graph embeddings learned by GCN-based models (GCN, MVGCN and TGNet) on HIV and BP datasets

munication efficiency and impaired functional integration between brain regions, underscoring the pathological impact of these diseases.

Moreover, we apply t-SNE [70] to visualize the graph embeddings learned by TGNet and compare them with those from GCN and MVGCN on small-scale HIV and BP datasets. As shown in Fig. 4, it can be seen that our TGNet model learns a higher quality of graph embeddings where the graphs are well-clustered according to their labels.

From the visualizations in Figs. 3-4, it is evident that TGNet effectively captures both local and global properties of the brain networks, leading to clear separations between healthy controls and patients across all datasets.

Conclusions

In this paper, we have presented a novel tensor-based graph convolutional network (TGNet) framework for multimodal brain network analysis. It advances prior works by showing how tensor and GCN techniques can be combined together to effectively model multimodal graph-structured data for joint embedding and classification, without using any prior knowledge of the data. Experimental results on four challenging multimodal brain network datasets (HIV, Bipolar, PPMI and ADNI) showed that our approach achieves superior performance for feature embedding and classification, compared with state-of-the-art methods.

Despite these advantages, TGNet has certain limitations. First, TGNet assumes certain dependencies between modalities, which may limit its performance in cases where cross-modal associations are weak or inconsistent. This limitation becomes more evident in scenarios where modality complementarity is insufficient or where data modalities lack consistency. Furthermore, the complexity of tensor decomposition and GCNs may reduce the interpretability of the model. In applications where clear biological interpretations are required, TGNet's structure may make it challenging to extract straightforward explanations.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13040-024-00409-6>.

Supplementary Material 1.

Acknowledgements

We would like to extend our sincere thanks to the developers of the DPARSF, CONN, FreeSurfer, and FSL toolboxes. Their excellent software significantly facilitated our data processing and analysis, making this research possible. We also appreciate the valuable support and resources provided by various research platforms and databases that contributed to the success of this study. Our special thanks go to Houliang Zhou for his invaluable assistance in applying his proposed SGCN method to our experimental datasets, significantly enhancing our research. Finally, we are grateful to the reviewers for their insightful comments and constructive feedback, which greatly refined and improved the quality of this work.

Authors' contributions

Zhaoming Kong: Conceptualized the study, performed the majority of the experiments, analyzed the data, and wrote the initial draft of the manuscript. Rong Zhou: Contributed to rewriting significant portions of the manuscript, prepared Figures 1-4, assisted in data analysis, contributed to the interpretation of experimental results, and edited the manuscript. Xinwei Luo: Assisted in writing specific sections of the manuscript and contributed to the overall editing process. Songlin Zhao: Conducted part of the baseline experiments and assisted in data analysis. Ann B. Ragin: Provided the HIV dataset and clinical support necessary for the study. Alex D. Leow: Provided the BP dataset and clinical support necessary for the study. Lifang He: Supervised the entire research process, provided research direction, and reviewed the manuscript. All authors discussed the main findings of the study and reviewed and approved the final manuscript.

Funding

This work was supported in part by the National Institutes of Health (R01MH080636, R21EY034179), National Science Foundation (MRI-2215789, IIS-2319451), and Lehigh University (Accelerator-S00010293 and CORE-001250).

Data availability

The Parkinson's Progression Markers Initiative (PPMI) is a public dataset available at <http://www.ppmi-info.org/data>. HIV and BP are private datasets, with HIV collected from the Early HIV Infection Study at Northwestern University [57] and BP collected from the collected from the UCLA Ahmanson-Lovelace Brain Mapping Center [59]. The data processed is anonymous with no personally identifiable information. All studies are conducted according to Good Clinical Practice guidelines and U.S. 21 CFR Part 50 (Protection of Human Subjects) and have Institutional Review Board approval. These datasets have been studied in previous literature [20, 26, 32, 71, 72].

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Received: 1 August 2024 Accepted: 21 November 2024

Published online: 06 December 2024

References

1. Fornito A, Zalesky A, Breakspear M. Graph analysis of the human connectome: promise, progress, and pitfalls. *Neuroimage*. 2013;80:426–44.
2. Tang H, Ma G, Zhang Y, Ye K, Guo L, Liu G, et al. A comprehensive survey of complex brain network representation. *Meta-Radiology*. 2023:100046.
3. Sun H, Wang A, He S. Temporal and spatial analysis of alzheimer's disease based on an improved convolutional neural network and a resting-state fMRI brain functional network. *Int J Environ Res Public Health*. 2022;19(8):4508.
4. Zhou R, Zhou H, Shen L, Chen BY, Zhang Y, He L. Integrating Multimodal Contrastive Learning and Cross-Modal Attention for Alzheimer's Disease Prediction in Brain Imaging Genetics. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2023. pp. 1806–1811.
5. Myszczyńska MA, Ojamies PN, Lacoste AM, Neil D, Saffari A, Mead R, et al. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat Rev Neurol*. 2020;16(8):440–56.
6. Tulay EE, Metin B, Tarhan N, Arkan MK. Multimodal neuroimaging: basic concepts and classification of neuropsychiatric diseases. *Clin EEG Neurosci*. 2019;50(1):20–33.
7. Liu S, Cai W, Liu S, Zhang F, Fulham M, Feng D, et al. Multimodal neuroimaging computing: a review of the applications in neuropsychiatric disorders. *Brain Inform*. 2015;2(3):167–80.
8. Benton A, Khayrallah H, Gujral B, Reisinger DA, Zhang S, Arora R. Deep Generalized Canonical Correlation Analysis. In: Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019). Florence: Association for Computational Linguistics; 2019. pp. 1–6.
9. Zhou R, Zhou H, Chen BY, Shen L, Zhang Y, He L. Attentive deep canonical correlation analysis for diagnosing Alzheimer's disease using multimodal imaging genetics. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Lecture Notes in Computer Science. Springer; 2023;14221:681–91.
10. Salim A, Shiju S, Sumitra S. Design of multi-view graph embedding using multiple kernel learning. *Eng Appl Artif Intell*. 2020;90:103534.
11. Zhang Y, Xiao L, Zhang G, Cai B, Stephen JM, Wilson TW, et al. Multi-paradigm fMRI fusion via sparse tensor decomposition in brain functional connectivity study. *IEEE J Biomed Health Inform*. 2020;25(5):1712–23.
12. Belyaeva I, Gabrielson B, Wang YP, Wilson TW, Calhoun VD, Stephen JM, et al. Learning Spatiotemporal Brain Dynamics in Adolescents via Multimodal MEG and fMRI Data Fusion Using Joint Tensor/Matrix Decomposition. *IEEE Trans Biomed Eng*. 2024;71(7):2189–200.
13. He L, Chen K, Xu W, Zhou J, Wang F. Boosted sparse and low-rank tensor regression. In: NIPS. New York: Curran Associates, Inc.; 2018. p. 1009–18.
14. Wang S, He L, Cao B, Lu CT, Yu PS, Ragin AB. Structural deep brain network mining. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery; 2017. p. 475–84.
15. Kawahara J, Brown CJ, Miller SP, Booth BG, Chau V, Grunau RE, et al. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*. 2017;146:1038–49.
16. Demir U, Gharsallaoui MA, Rezik I. Clustering-based deep brain multigraph integrator network for learning connective brain templates. In: Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis: Second International Workshop, UNSURE 2020, and Third International Workshop, GRAIL 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 2. Springer; 2020. pp. 109–120.

17. Zhu Q, Yang J, Wang S, Zhang D, Zhang Z. Multi-Modal Non-Euclidean Brain Network Analysis With Community Detection and Convolutional Autoencoder. *IEEE Trans Emerg Top Comput Intell.* 2022;7(2):436–46.
18. Chen X, Ke P, Huang Y, Zhou J, Li H, Peng R, et al. Discriminative analysis of schizophrenia patients using graph convolutional networks: A combined multimodal MRI and connectomics analysis. *Front Neurosci.* 2023;17:1140801.
19. Zhang, K., Zhou, R., Adhikarla, E. et al. A generalist vision–language foundation model for diverse biomedical tasks. *Nat Med.* 2024;30:3129–41.
20. Zhang L, et al. Multi-view Graph Convolutional Network and Its Applications on Brain Network Analysis. *Neurocomputing.* 2018;312:354–68.
21. Wen G, Cao P, Bao H, Yang W, Zheng T, Zaiane O. MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. *Comput Biol Med.* 2022;142:105239.
22. Qu G, Zhou Z, Calhoun VD, Zhang A, Wang YP. Integrated Brain Connectivity Analysis with fMRI, DTI, and sMRI Powered by Interpretable Graph Neural Networks. 2024. arXiv preprint [arXiv:2408.14254](https://arxiv.org/abs/2408.14254).
23. Zhou H, He L, Chen BY, Shen L, Zhang Y. Multi-Modal Diagnosis of Alzheimer's Disease using Interpretable Graph Convolutional Networks. *IEEE Trans Med Imaging.* 2024. p. 1–12.
24. Ma Y, Zhang T, Wu Z, Mu X, Liang X, Guo L. Multi-view Brain Networks Construction for Alzheimer's Disease Diagnosis. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2023. pp. 889–892.
25. Liu L, Wang YP, Wang Y, Zhang P, Xiong S. An enhanced multi-modal brain graph network for classifying neuropsychiatric disorders. *Med Image Anal.* 2022;81:102550.
26. Cui H, Dai W, Zhu Y, Li X, He L, Yang C. Interpretable graph neural networks for connectome-based brain disorder analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2022. pp. 375–385.
27. Zuo Q, Zhong N, Pan Y, Wu H, Lei B, Wang S. Brain structure-function fusing representation learning using adversarial decomposed-VAE for analyzing MCI. *IEEE Trans Neural Syst Rehabil Eng.* 2023;31:4017–28.
28. Lu H, Platanotis KN, Venetsanopoulos AN. MPCA: Multilinear principal component analysis of tensor objects. *IEEE Trans Neural Netw Learn Syst.* 2008;19(1):18–39.
29. Zhu Y, Cui H, He L, Sun L, Yang C. Joint embedding of structural and functional brain networks with graph neural networks for mental illness diagnosis. In: 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society. IEEE; 2022. pp. 272–276.
30. Shao W, He L, Philip SY. Clustering on multi-source incomplete data via tensor modeling and factorization. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer; 2015. pp. 485–497.
31. Yin M, Gao J, Xie S, Guo Y. Multiview subspace clustering via tensorial t-product representation. *IEEE Trans Neural Netw Learn Syst.* 2018;30(3):851–64.
32. Ma G, He L, Lu CT, Shao W, Yu PS, Leow AD, et al. Multi-view clustering with graph embedding for connectome analysis. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. New York: Association for Computing Machinery; 2017. p. 127–36.
33. Liu Y, He L, Cao B, Philip SY, Ragin AB, Leow AD. Multi-view multi-graph embedding for brain network clustering analysis. In: AAAI. New Orleans: AAAI Press; 2018:32.
34. Wang C, Zhang L, Zhang J, Qiao L, Liu M. Fusing Multiview Functional Brain Networks by Joint Embedding for Brain Disease Identification. *J Personalized Med.* 2023;13(2):251.
35. Li G, Huang Q, Liu C, Wang G, Guo L, Liu R, et al. Fully Automated Diagnosis of Thyroid Nodule Ultrasound using Brain-Inspired Inference. *Neurocomputing.* 2024;582:127497.
36. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. 2016. arXiv preprint [arXiv:1609.02907](https://arxiv.org/abs/1609.02907).
37. Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. In: NIPS. New York: Curran Associates, Inc.; 2017. p. 1024–34.
38. Kazi A, Shekarforoush S, Krishna SA, Burwinkel H, Vivar G, Wiestler B, et al. Graph convolution based attention model for personalized disease prediction. In: MICCAI. Springer; 2019. pp. 122–130.
39. Yang Y, Ye C, Guo X, Wu T, Xiang Y, Ma T. Mapping multi-modal brain connectome for brain disorder diagnosis via cross-modal mutual learning. *IEEE Trans Med Imaging.* 2023;43(1):108–21.
40. Zhang Y, He X, Chan YH, Teng Q, Rajapakse JC. Multi-modal graph neural network for early diagnosis of Alzheimer's disease from sMRI and PET scans. *Comput Biol Med.* 2023;164:107328.
41. Li X, Dvornek NC, Zhou Y, Zhuang J, Ventola P, Duncan JS. Graph neural network for interpreting task-fMRI biomarkers. In: MICCAI. Springer; 2019. pp. 485–493.
42. Zhang T, Zheng W, Cui Z, Li Y. Tensor graph convolutional neural network. 2018. arXiv preprint [arXiv:1803.10071](https://arxiv.org/abs/1803.10071).
43. Malik OA, Ubaru S, Horesh L, Kilmer ME, Avron H. Tensor graph neural networks for learning on time varying graphs. In: Proceedings of NIPS Workshop, 2019.
44. Huang Z, Li X, Ye Y, Ng MK. MR-GCN: Multi-Relational Graph Convolutional Networks based on Generalized Tensor Product. In: IJCAI. Yokohama: International Joint Conferences on Artificial Intelligence Organization. 2020. p. 1258–64.
45. Liu X, You X, Zhang X, Wu J, Lv P. Tensor graph convolutional networks for text classification. In: AAAI. New Orleans: AAAI Press; 2020;34:8409–16.
46. Ioannidis VN, Marques AG, Giannakis GB. Tensor graph convolutional networks for multi-relational and robust learning. *IEEE Trans Signal Process.* 2020;68:6535–46.
47. Sang Y, Li W. Classification Study of Alzheimer's Disease Based on Self-Attention Mechanism and DTI Imaging Using GCN. *IEEE Access.* 2024.
48. Samanta A, Sarma M, Samanta D, ALERT: Atlas-Based Low Estimation Rank Tensor Approach to Detect Autism Spectrum Disorder. In: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2023. pp. 1–4.
49. Yang Y, Cai G, Ye C, Xiang Y, Ma T. Tensor-based Complex-valued Graph Neural Network for Dynamic Coupling Multimodal brain Networks. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2023. pp. 1–5.

50. Balcilar M, Renton G, Héroux P, Gauzere B, Adam S, Honeine P. Bridging the Gap Between Spectral and Spatial Domains in Graph Neural Networks. 2020. arXiv preprint [arXiv:2003.11702](https://arxiv.org/abs/2003.11702).
51. Hung H, Wu P, Tu I, Huang S. On multilinear principal component analysis of order-two tensors. *Biometrika*. 2012;99(3):569–83.
52. Kolda TG, Bader BW. Tensor decompositions and applications. *SIAM Rev*. 2009;51(3):455–500.
53. Lee C, Wang M. Tensor denoising and completion based on ordinal observations. In: International conference on machine learning. PMLR; 2020. pp. 5778–5788.
54. Rajwade A, Rangarajan A, Banerjee A. Image denoising using the higher order singular value decomposition. *IEEE Trans Pattern Anal Mach Intell*. 2012;35(4):849–62.
55. Han L, Wu Z, Zeng K, Yang X. Online multilinear principal component analysis. *Neurocomputing*. 2018.
56. Beaufays F. Transform-domain adaptive filters: An analytical approach. *IEEE Trans Signal Process*. 1995;43(2):422–31.
57. Ragin AB, Du H, Ochs R, Wu Y, Sammet CL, Shoukry A, et al. Structural brain alterations can be detected early in HIV infection. *Neurology*. 2012;79(24):2328–34.
58. Cao B, Kong X, Zhang J, Philip SY, Ragin AB. Identifying HIV-induced subgraph patterns in brain networks with side information. *Brain Inform*. 2015;2(4):211–23.
59. Ajilore O, Vizueta N, Walshaw P, Zhan L, Leow A, Altshuler LL. Connectome signatures of neurocognitive abnormalities in euthymic bipolar I disorder. *J Psychiatr Res*. 2015;68:37–44.
60. Whitfield-Gabrieli S, Nieto-Castanon A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain connectivity*. 2012;2(3):125–41.
61. Zhan L, Zhou J, Wang Y, Jin Y, Jahanshad N, Prasad G, et al. Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease. *Front Aging Neurosci*. 2015;7:48.
62. Dyrba M, Grothe M, Kirste T, Teipel SJ. Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM. *Hum Brain Mapp*. 2015;36(6):2118–31.
63. Gupta A, Ayhan M, Maida A. Natural Image Bases to Represent Neuroimaging Data. In: ICML. Atlanta: JMLR. org; 2013:987–94.
64. Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y, et al. Graph attention networks. *Stat*. 2017;1050(20):10–48550.
65. Ying Z, You J, Morris C, Ren X, Hamilton W, Leskovec J. Hierarchical graph representation learning with differentiable pooling. In: NIPS. New York: Curran Associates, Inc.; 2018:4800–10.
66. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. In: ICLR (Poster). San Diego: International Conference on Learning Representations; 2015:5:6.
67. Li R, Wang W, Wang Y, Peters S, Zhang X, Li H. Effects of early HIV infection and combination antiretroviral therapy on intrinsic brain activity: a cross-sectional resting-state fMRI study. *Neuropsychiatr Dis Treat*. 2019;15:883.
68. Gandhi AB, Ifrah Kaleem JA, Hisbulla M, Kannichamy V, Antony I, Mishra V, et al. Neuroplasticity Improves Bipolar Disorder: A Review. *Cureus*. 2020;12(10):3129–41.
69. Ferro A, Bonivento C, Delvecchio G, Bellani M, Perlina C, Dusi N, et al. Longitudinal investigation of the parietal lobe anatomy in bipolar disorder and its association with general functioning. *Psychiatry Res Neuroimaging*. 2017;267:22–31.
70. Maaten Lvd, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9:2579–605.
71. Luo G, Li C, Cui H, Sun L, He L, Yang C. Multi-view brain network analysis with cross-view missing network generation. In: IEEE International Conference on Bioinformatics and Biomedicine. IEEE; 2022. pp. 108–115.
72. Cui H, Dai W, Zhu Y, Kan X, Gu AAC, Lukemire J, et al. Braingb: a benchmark for brain network analysis with graph neural networks. *IEEE Trans Med Imaging*. 2022;42(2):493–506.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.