Received 16 January 2024, accepted 30 January 2024, date of publication 8 February 2024, date of current version 20 February 2024.
Digital Object Identifier 10.1109/ACCESS.2024.3364545

RESEARCH ARTICLE

Classification Study of Alzheimer's Disease Based on Self-Attention Mechanism and DTI Imaging Using GCN

YILIN SANG AND WAN LI
School of Computing Science and Engineering, Beijing Technology and Business University, Beijing 100048, China
Corresponding author: Wan Li (wanli@btbu.edu.cn)

ABSTRACT Alzheimer's disease (AD) is a neurodegenerative disorder. Diffusion tensor imaging (DTI) provides information about the integrity of white matter fiber bundles that are related to the neuropathological mechanisms, and it is one of the commonly used techniques in AD research. In this study, we first divided each subject's brain into 90 regions based on the automated anatomical labeling (AAL) brain atlas. The average fractional anisotropy (FA) values between each pair of regions were applied to construct a brain network. We utilized the number of voxels with fibers passing through each brain region as the node features. The brain networks and node features were input into a novel graph convolutional neural network (GCN) structure involving the self-attention pooling mechanism proposed in this study to classify AD and normal controls (NC). The classification performance was compared among different preprocessed brain networks and node features. The final classification result achieved an accuracy of 87.5%.

INDEX TERMS Alzheimer's disease, diffusion tensor imaging, image classification, brain networks, graph convolutional neural networks.

I. INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disorder that causes irreversible deterioration of neurological function. It is primarily characterized by a decline in cognitive function, significantly impacting the everyday lives of patients and their families [1]. In 2021, over 55 million people worldwide were diagnosed with this disease, and the number of AD patients is estimated to reach 78 million by 2030 [2]. With the development of neuroimaging techniques, various neuroimaging modalities have shown potential for improving the diagnosis of AD from different perspectives.

Diffusion tensor imaging (DTI) is a non-invasive magnetic resonance imaging (MRI) technique that captures water molecules' degree of anisotropic diffusion along axons in the white matter. It can identify abnormal diffusion patterns in various neurological disorders and provide information about the integrity of white matter fiber tracts related to neurobiological mechanisms [3]. So far, DTI is the only neuroimaging technique that can describe white matter fiber pathways and is highly sensitive to microstructural white matter damage within fiber bundles. Therefore, DTI is typically used to specify anatomical connectivity impairments that cannot be detected by structural MRI (sMRI). The two most frequently used features to characterize white matter integrity are fractional anisotropy (FA) and mean diffusivity (MD) [4]. FA provides information about fiber density, axon diameter, and myelination, with decreased values indicating a loss of fiber tract integrity. MD measures the average diffusivity of water molecules in non-collinear directions, with increased values indicating increased free diffusion of water molecules and compromised anisotropy. The main pathological features of AD include neuritic plaques or amyloid plaques (extracellular deposits) and neurofibrillary tangles (intracellular aggregates of hyperphosphorylated tau proteins), which can be revealed as decreased FA and increased MD in the cingulate, corpus callosum, and hippocampus regions [5], [6].

Graph neural network (GNN) is a general type of graph neural network that can handle various types of graph data, including directed, undirected, and weighted graphs [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Roberta Palmeri.

Compared to the convolutional neural network (CNN), which is primarily used for grid-like data such as images arranged in a 2D or 3D grid structure (e.g., pixels in an image or frames in a video), GNN operates on graph-structured data where nodes have relationships and connections. Graph data can be characterized as a collection of nodes and edges, such as social networks, knowledge graphs, and brain networks. GNN is a universal framework for processing graph-structured data. Its core idea is to capture interactions between nodes by propagating messages among them.

The graph convolutional network (GCN) is a GNN structure that involves convolutional operations. GCN models the relationships between nodes and edges to learn from the graph layout [8]. Each layer of GCN performs convolution operations among the features of nodes and their neighboring nodes, allowing for richer representations. GCN typically consists of multiple graph convolutional layers and nonlinear activation functions. Through comparison, it has been found that GCN is more suitable for brain network classification and can better explore the connections between various brain regions. Therefore, we have determined to develop a brain network classification model based on GCN.

II. RELEVANT RESEARCH

AD-related GCN research was first conducted in 2019. Using fiber tractography, Song et al. generated a fiber graph based on brain regions. The obtained adjacency matrix was then input into GCN for the four-class AD classification. The node degree and clustering coefficient were also input into a support vector machine (SVM) classifier for performance comparison with GCN. The results indicated that GCN outperformed SVM [9]. Kong et al. suggested a generative model for structural brain networks based on adversarial learning. They directly learned the structural connections from DTI images and input the generated connectivity matrix into GCN for AD classification, achieving satisfactory classification results [10]. Yang et al. proposed a method for extracting features from graph-structured data. They optimized the features using maximum mutual information and then employed the brain network to classify AD subjects. This approach exceeded other feature extraction methods in classification accuracy [11].

Not only single-modality DTI images but GCN has also been widely applied in recent years to classify using a combination of DTI and other modality images. Especially the functional MRI (fMRI)-DTI combination since it is straightforward to get functional and structural brain networks [12], [13], [14], [15], [16]. Although structural MRI (sMRI)-DTI fusion classification research can consider gray and white matter features, the CNN model mainly accomplishes the classification [17], [18], [19], [20]. One study has endeavored to combine images of three modalities for classification. 3D sMRI images were input into the multi-channel ResNet network model, while the brain networks constructed by DTI and fMRI were input into the GCN model. Finally, multi-channel ResNet and GCN were combined for multi-modality classification to obtain pleasing results [21].

Other than the limited applications of GCN in AD-related research, the currently favored AD-classification approaches utilizing DTI images are briefly introduced as follows.

GCN-excluded classification studies can be organized into three categories: voxel-based, brain region-based, and network-based classification studies. The research on voxel-based classification is to select the most representative AD voxels from the whole brain, calculate their DTI parameter values, such as FA and MD, and then classify them by various classifiers [22], [23], [24], [25]. Brain region-based research predominantly focuses on the AD-sensitive brain regions, such as the parietal lobe, hippocampus, amygdala, and middle temporal lobe [26], [27]. The white matter features are extracted from the abovementioned regions and then input into various classifiers [28], [29]. Due to the influence of AD, some connections of neurons would be damaged, resulting in information transmission barriers and corresponding symptoms [30]. Therefore, the network-based classification research can be categorized into classification by extracting DTI parameters from fiber bundles [31], [32] and classification by analyzing brain networks [33], [34], [35], [36].

More and more studies have been conducted in recent years on using GCN networks to classify mental diseases. Besides Alzheimer's disease, GCN has also applied to classify other diseases such as Parkinson's disease [37], [38], [39], autism [40], [41], [42], major depression [43], [44], [45], schizophrenia [46], [47], [48], attention deficit hyperactivity disorder [49], [50], and bipolar disorder [51], [52].

In recent years of research, the studies using GCN for classification can account for about 70%. In other words, the current research has begun to pay more attention to the essential characteristics of DTI images: the white matter fiber bundle can be abstracted into a structural brain network. With the unveiling potential of GCN algorithms, researchers have tried to employ GCN on white matter connectivity for AD classification to reach satisfactory performance. Hence, more attempts need to be conducted. Moreover, a well-known finding is that AD attacks specific white matter fiber bundles [53], [54]. However, the current GCN-based AD classifications treat all the network connections equally.

Moreover, GCN is more suitable for the classification of brain networks. Firstly, GCN can make full use of the structure information of brain networks in DTI images for classification, and can extract and utilize the connection pattern and topology of brain networks from DTI data, so as to better capture the changes of brain networks in Alzheimer's disease. The second point is that GCN is able to classify both node features and connection information, combining them as input features. This helps to more fully characterize brain networks in Alzheimer's disease and improve classification performance. The third point is that in the task of classifying Alzheimer's disease, the global structure of the entire brain network is critical to understanding and identifying disease features. GCN can aggregate global information layer by layer, and can capture a broader context of the brain network across connections of different distances, helping to improve classification performance.

Therefore, we applied the GCN on the brain networks abstracted from the DTI image for classification. Notably, we endeavored to add the self-attention mechanism to the original GCN structure in this study to realize better AD classification.

III. EXPERIMENT

Our study utilizes the white matter features of DTI images and employs GCN with the self-attention mechanism for classification. The network takes structural brain networks based on DTI as input to generate cognitive state category labels and uses these labels as output to obtain the final classification accuracy. Our experiments use two labels: AD and normal control (NC). The DTI data processing will be introduced in the first part of this section, followed by a description of the GCN framework in the second part. The results of the experiments will be presented in the third part. Finally, we summarized our investigation and provided the expectations for future research.

A. EXPERIMENTAL DATA

The data applied in this study is sourced from the ADNI database (https://adni.loni.usc.edu/), from which we selected 70 AD patients and 70 NC individuals. The preprocessing experiments were conducted using FSL (FMRIB Software Library) and the FSL-based PANDA (Pipeline for Analyzing braiN Diffusion imAges) software. PANDA is a Linux-based software that is running within MATLAB. The preprocessing workflow begins with converting the downloaded data from ADNI into the nii.gz format using FSL. Subsequently, skull stripping and eddy current correction are performed, and then DTI parameters such as FA and MD are calculated via FSL. Next, PANDA's deterministic fiber tracking technique was employed to construct white matter fiber bundles based on the white matter trajectories. Finally, the automated anatomical labeling (AAL) brain atlas is utilized to segment the brain into a 90×90 brain network. Each brain region can be considered a node in the network, with features encompassing the number of voxels in each brain region. After preprocessing, three structural brain networks are obtained: (1) the FA brain network, constructed based on the average FA values between each brain region according to the brain atlas; (2) the FN brain network, constructed based on the number of fibers between each brain region according to the brain atlas; (3) the LEN brain network, constructed based on the average fiber length between each brain subdivision according to the brain atlas. The node features include (1) ROIS (ROISurfaceSize), denoting the number of voxels traversed by fibers in each brain region; (2) ROIV (ROIVoxelSize), representing the number of voxels in each brain region.

B. EXPERIMENTATION

This section primarily introduces the overall workflow of the study, starting with the GCN framework used in this study. The calculation formula of GCN and the process of information propagation are explained. Following that, the general design of the model is presented, and finally, an overview of the working process of the self-attention mechanism is provided.

D represents the degree matrix of the graph, and H denotes the feature matrix of the nodes. Adding self-connections to the adjacency matrix is crucial, implying that the diagonal numbers are all set to 1. This definition is used in the graph convolution formula, although in other cases, the adjacency matrix may not include self-loops [55]. The adjacency matrix with self-connections allows the preservation of the features of each node and their propagation through the network. Calculating the degree matrix D involves summing each row of the adjacency matrix and assigning the resulting values to the diagonal. The added diagonal values are then inverted and square-rooted to obtain the degree matrix with a value of $-1/2$. Multiplying the adjacency matrix on both sides by the degree matrix effectively adds weights to the edges. This adjustment leverages the varying node degrees to control the amount of information transmitted.

$$H^{(l+1)} = \sigma(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}). \tag{1}$$

The significance of incorporating the degree matrix is that, after multiplying the self-connection matrix and the node feature matrix, if a node is connected to many edges, the resulting feature of that node may become exceptionally large. This is because it needs to be summed with the feature vectors of multiple nodes. In our study, we selected voxel values as node features for each brain region. As these features are propagated, they can become particularly significant. Therefore, it is necessary to normalize the nodes by applying normalization techniques to confine all features within a reasonable range. This normalization helps balance the importance of nodes with more significant degrees. Incorporating the degree matrix is akin to performing this normalization operation.

The purpose of GCN is to perform feature extraction, and according to the rules of matrix multiplication, multiplying the adjacency matrix and the feature matrix is the information propagation process. If the information is propagated to the l+1 layer, it is necessary to obtain the feature matrix of the l+1 layer using the feature matrix of the l layer. As shown in Figure 1, Node 0 is connected to Node 1 and Node 2, and each node has its feature matrix. So, the feature matrix of Node 0 in the l+1 layer is obtained by adding the feature vector of Node 0 in the l layer to the feature vector of Node 1 and the feature matrix of Node 2 in the l layer. The propagation of graph convolution is the process of information transmission. Each node first receives information from its neighboring nodes and then gathers information from all nodes in the graph in a layer-by-layer propagation process. In our study, we treat each brain region as a node. Through preprocessing, we obtain a 90×90 adjacency matrix. In this adjacency matrix, the association between each brain region can be regarded as an edge between two nodes. Therefore, the information propagation in the brain network within GCN is the "communication" between each brain region through

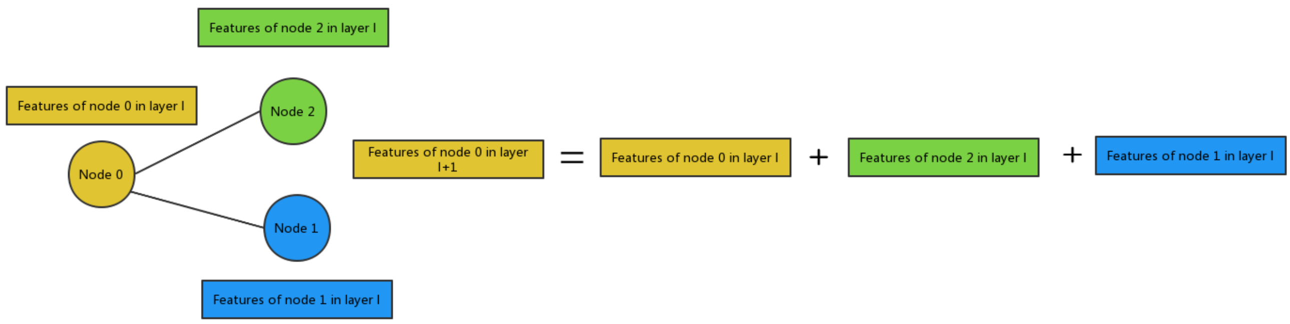

FIGURE 1. **Illustration of Graph Convolution Information Propagation Process.**

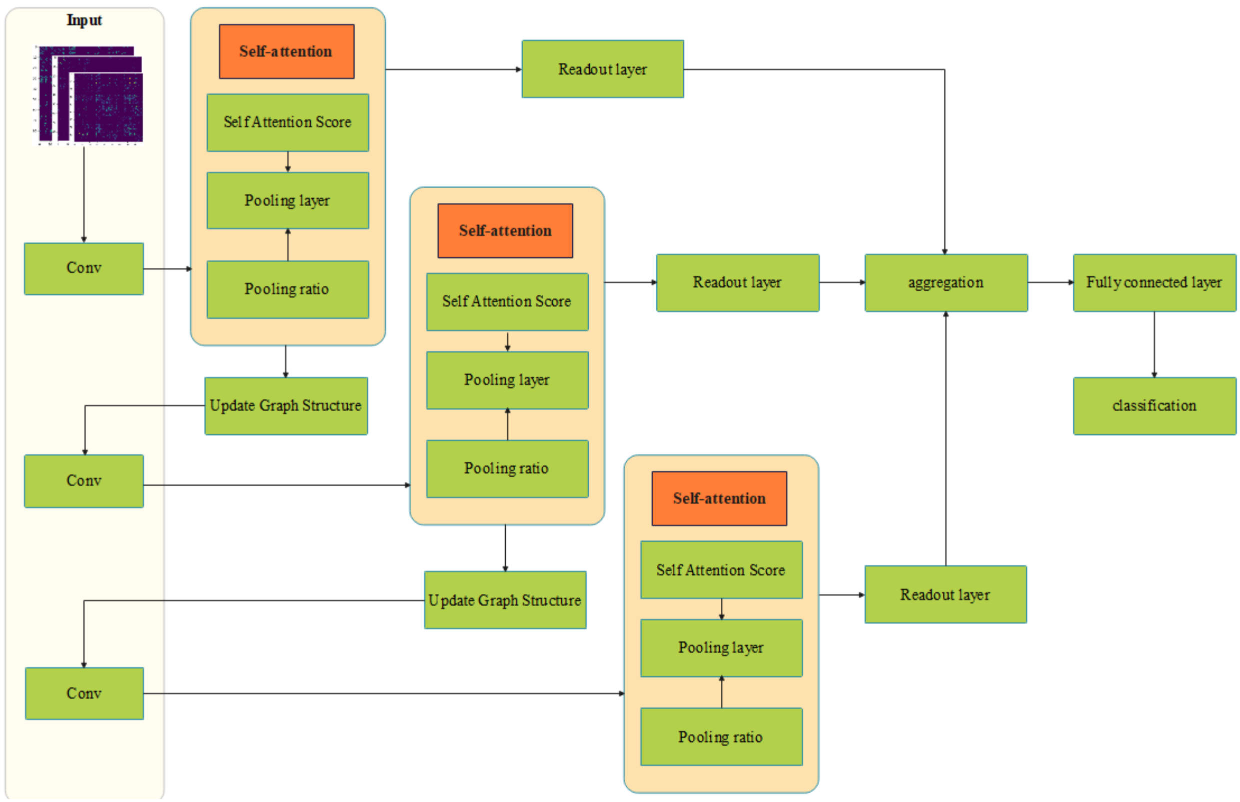

FIGURE 2. **Illustration of AD Classification Model Based on Self-Attention Mechanism.**

their node features, ultimately obtaining the features of each node in the entire network for final training.

Our experimental model consists of a three-layer GCN hierarchical pooling model, as shown in Figure 2. It consists of three modules, each consisting of a graph convolutional layer and a graph pooling layer. The outputs of each module are aggregated in the readout layer, which is responsible for aggregating node features to generate a fixed-size representation. The sum of the outputs from each readout layer is then passed through a linear layer for classification. The model's key aspect lies in the pooling layer's design, which incorporates a self-attention mechanism. This mechanism filters the nodes by using self-attention pooling to eliminate irrelevant nodes. We take the DTI brain network and node features as input. After one layer of convolution, in the pooling layer, we first compute the self-attention scores Z for the input brain network. Once the self-attention scores are obtained, we use the top-rank function to sort the self-attention scores of each node in the brain network. Then, we define a pooling ratio k to select the desired number of nodes to retain. Finally, we output the new graph structure, adjacency matrix, and feature matrix. After three rounds of convolution, the newly obtained node features from each convolution are outputted through the readout layer. Then, they are aggregated and passed through a fully connected layer for classification, resulting in the final classification results.

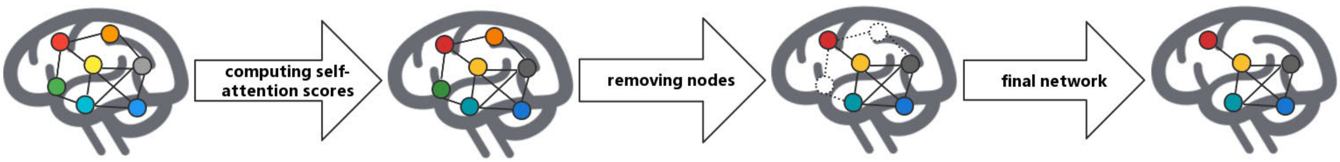

FIGURE 3. Operation process of the self-attention mechanism, where circles of different colors represent different brain regions, connecting lines represent the connections between brain regions, and dotted lines and dotted circles represent nodes with lower rankings to be removed.

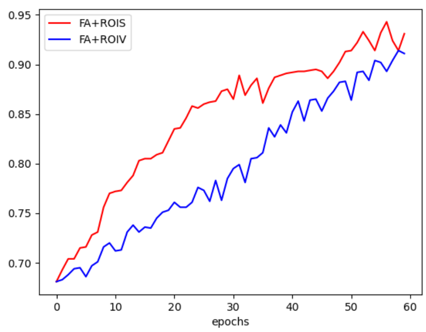

FIGURE 4. Accuracy Comparison between FA+ROIS/ROIV.

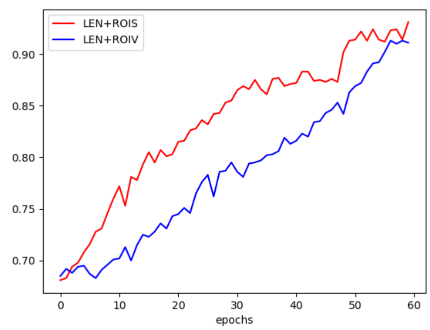

FIGURE 6. Accuracy Comparison between LEN+ROIS/ROIV.

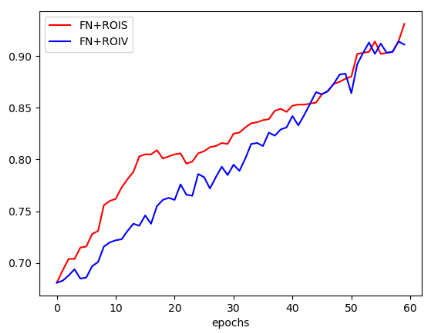

FIGURE 5. Accuracy Comparison between FN+ROIS/ROIV.

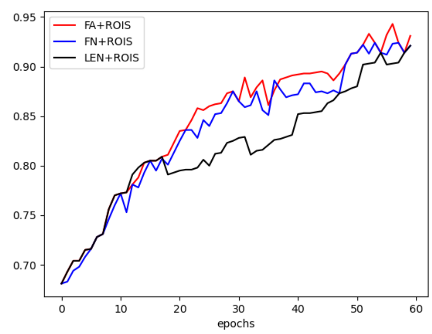

FIGURE 7. Accuracy Comparison between FA/FN/LEN+ROIS.

The self-attention pooling mechanism we incorporated allows learning hierarchical representations with relatively fewer parameters in an end-to-end manner. It takes into account both node features and the topological structure. Additionally, it utilizes a self-attention mechanism to distinguish between nodes that should be removed and retained [45]. The calculation process of the self-attention mechanism is displayed in Figure 3. Firstly, it computes the self-attention scores for each node in each brain network. Then, it sorts the nodes based on these scores. Finally, setting a pooling ratio removes the nodes to obtain the final graph. Nodes that are removed can be determined based on prior knowledge of deleting nodes that are irrelevant to AD classification. Next, we will provide a detailed explanation of the calculation formula and process of the self-attention mechanism.

The calculation of self-attention scores and the top-rank function is depicted in equations (2) and (3). The formula for calculating the self-attention scores Z uses parameters similar to those used in graph convolution. Here, X represents the nodes' feature matrix and the convolutional weights of the input feature space. Once the self-attention scores Z are obtained, the top-kN nodes can be selected based on the values of Z. After sorting, the top-rank function is used to obtain the indices of the top-kN values. The parameter k is a hyperparameter ranging from 0 to 1, representing the pooling ratio determining the number of nodes retained. After obtaining the indices of the retained nodes, a new feature matrix and adjacency matrix can be obtained. In simple terms, the self-attention scores are calculated using the formula, and based on these scores, the nodes are sorted while irrelevant nodes are removed. The pooling ratio k determines the number of nodes to be removed.

$$Z = \sigma(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}X\theta_{att}) \tag{2}$$

$$idx = top - rank(Z, [kN]) \tag{3}$$

Our research combines the self-attention mechanism with GCN to classify AD. We utilize DTI to abstract the brain networks into traits based on the number of white matter fibers and combine them with GCN for classification. The experimental process begins with obtaining preprocessed brain networks, including FN, FA, LEN, and node features

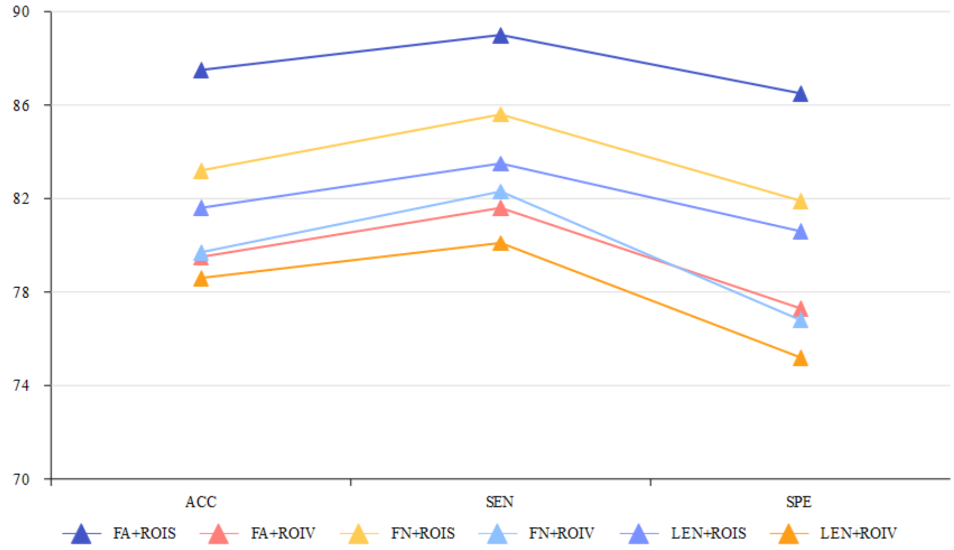

FIGURE 8. Accuracy, sensitivity and specificity using different feature combinations.

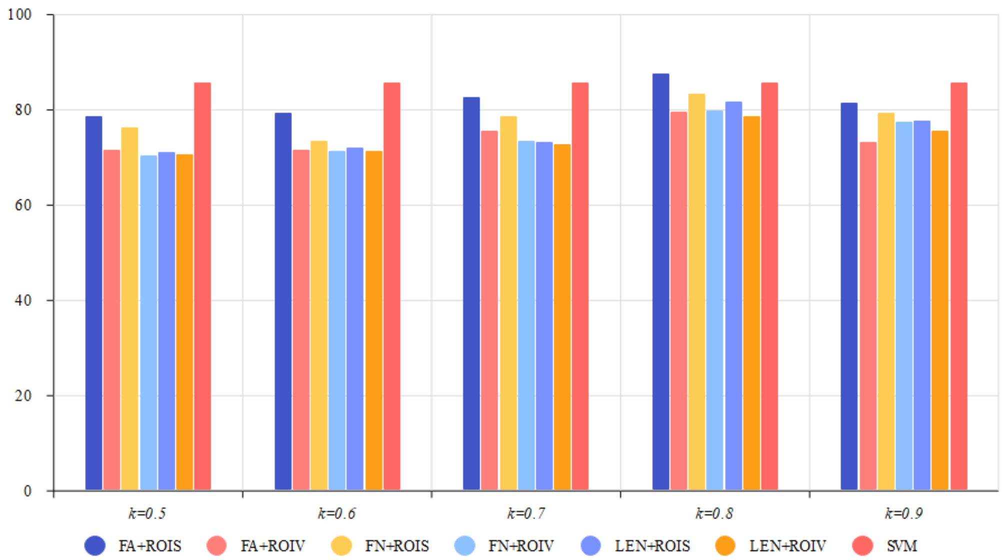

FIGURE 9. The accuracy of different feature combinations combined with different *k* values was compared with that of SVM.

ROIS and ROIV. The connections between each brain region are abstracted as relationships between nodes and edges in the network. ROIS and ROIV serve as node features, and each node's graph membership and graph labels are used as inputs to the GCN model. Next, the model is designed with three convolutional layers. In the pooling layers following each convolutional layer, the self-attention mechanism is incorporated to filter the nodes. By removing irrelevant nodes from the entire brain network, the accuracy of the classification is improved.

C. EXPERIMENTAL RESULTS

In this experiment, six combination features were tested in the ADNI dataset: FN brain network as network feature and ROIS as node feature, FN brain network as network feature and ROIV as node feature, FA brain network as network feature and ROIS as node feature, FA brain network as network feature and ROIV as node feature, LEN brain network as network feature and ROIS as node feature, LEN brain network is the network feature and ROIV is the node feature. To compare the impact of different k values on accuracy, the k value was scaled from 0.5 to 0.9. The final result showed that using the FA brain network as network features, ROIS as node features, and setting the k value to 0.8 achieved an accuracy of 87.5%. Figures 4-7 show the accuracy of different combinations when $k = 0.8$, and Figure 8 shows the various combinations' accuracy, sensitivity, and specificity. Figure 9 compares the accuracy of each k value and SVM.

TABLE 1. Table 1. Accuracy of different matrix and node feature combinations with different k values.

Matrix + Node Features	K=0.5	K=0.6	K=0.7	K=0.8	K=0.9
FA+ROIS	78.6%	79.2%	82.6%	87.5%	81.3%
FA+ROIV	71.5%	71.6%	75.4%	79.5%	73.1%
FN+ROIS	76.3%	73.4%	78.6%	83.2%	79.3%
FN+ROIV	70.3%	71.2%	73.4%	79.7%	77.3%
LEN+ROIS	71.1%	71.9%	73.2%	81.6%	77.6%
LEN+ROIV	70.6%	71.3%	72.7%	78.6%	75.5%

TABLE 2. Comparison of accuracy with other literature.

Literature	Classifier	Features	Accuracy
[46]	GCN	Connectivity Matrix	83.3%
[47]	GCN	Graph Data	77.3%
Ours	GCN	FA+ROIS	87.5%

This study considers the combination of different adjacency matrices representing brain networks and various node features and tests the influence of different k values on accuracy. The accuracy graph indicates that using the FA brain network as network features and ROIS as node features yields the best classification results. Moreover, using ROIS as node features generally outperforms using ROIV as node features for classification. By analyzing the node feature data, this result may be because ROIS represents the number of voxels with fibers passing through them. In contrast, ROIV represents the number of voxels in each brain region, with significant differences in the number of fibers in each brain region.

Regarding the choice of k value, the best result is achieved at $k = 0.8$, with slightly better results for $k = 0.7$ compared to $k = 0.9$. This may be because there are still more redundant nodes retained at $k = 0.9$. The lower accuracy at $k = 0.5$ and $k = 0.6$ may be due to the deletion of overmuch nodes, resulting in incomplete information propagation. The accuracy results of this experiment compared to other studies are listed in Table 2. This experiment achieved good accuracy results in classification research conducted using GCN. In addition, this experiment is compared with the traditional machine learning algorithm SVM, and it can be seen that the accuracy of FA+ROIS is higher than that of SVM when $k = 0.8$. At the same time, other combinations are lower than that of SVM.

This experiment's advantages lie in using different brain networks combined with varying node features for experimentation. Furthermore, integrating the self-attention mechanism with GCN allows for effective information propagation between each node feature. Using the k value enables the removal of redundant nodes, ensuring that each node receives more important information during information propagation. Additionally, the experiment considers the impact of different k values. However, there are certain limitations in this study. The current approach does not allow for autonomously setting the k value for each layer during the training process. Currently, the experiment assesses the impact of fixed k values on the results. However, in the training process of GCN, if it were possible to dynamically set the k value to match the current state at each convolution, it could lead to improved results. This aspect remains an area for further research.

There have been relatively few experiments using GCN for classification in the current research landscape. Moreover, researchers commonly treat DTI as auxiliary images and combine them with fMRI, inputting structural and functional brain networks into GCN for classification. However, this study shifts its focus to standalone DTI images. It achieves promising results in binary classification and multi-modality experiments conducted using GCN.

IV. SUMMARIZE

Our study combines the self-attention mechanism with GCN to classify AD from NC, using structural brain networks constructed based on DTI images. The purpose is to investigate the classification performance of the unique white matter network derived from DTI images when the self-attention mechanism is included with GCN. The advantage lies in avoiding complex preprocessing and feature extraction steps. Instead, only the DTI brain network is input to GCN to obtain classification accuracy. Most studies focus on innovative feature extraction methods in the current research landscape. They extract features from voxels or brain regions using various feature extraction techniques and then employ traditional classifiers such as SVM for classification. However, due to the primary role of DTI images in AD's white matter regions and its image clarity limitations, relatively few studies utilize DTI images for classification research using CNN models. Therefore, our study directs its attention to DTI brain networks.

The self-attention mechanism is involved because specific brain regions are essential in AD classification research, and others are irrelevant. After abstracting DTI images into brain networks, the importance of each brain region can be determined based on its degree within the network. The self-attention mechanism can eliminate irrelevant nodes during the training process, equivalent to removing brain regions unrelated to AD in the brain network. Additionally, it can rank each node based on its self-attention score, allowing for integration with AD-related brain regions and improving classification accuracy.

This study has room for improvement regarding the automatic selection of the k value. The k value is used in the self-attention pooling model to determine the number of nodes to retain. However, it cannot be set as a self-learning parameter that autonomously selects the optimal number of nodes for each network during classification training. This issue requires incorporating prior medical knowledge to determine the appropriate number of brain regions to include in the brain network for optimal classification. In future research, particular attention will be given to this problem to achieve autonomous learning of the k value.

This study successfully integrates the self-attention mechanism with GCN to classify DTI brain networks and achieves promising results. It demonstrates the feasibility of directly using DTI brain networks through GCN for classification, providing a new approach for future research. Different

GCN models can be employed for AD classification studies, thus filling the gap in deep learning classification research utilizing DTI images. Previously, the application of DTI images in deep learning primarily served as a supplementary role, combined with gray matter features from sMRI images to enhance classification accuracy.

Future research on AD classification using DTI images can focus on the unique structural brain networks derived from DTI images. Especially with the rise of GCN, more studies aim to simplify the classification process by using GCN, eliminating the need for complex and tedious preprocessing steps. This approach is more conducive to practical implementation in the future. Traditional machine learning algorithms consume significant time for training when dealing with large brain images. However, by processing them into brain networks, not only can a substantial amount of time be saved, but there is also no need for additional feature extraction. DTI brain networks can be directly trained. This time-saving aspect becomes particularly beneficial in practical use in the future.GCN is a relatively new network with plenty of space for development. Therefore, future research on AD classification using DTI images, functional brain networks derived from fMRI, and the fusion of these two networks should pay more attention to this aspect of utilizing GCN.

REFERENCES

[1] A. Collie and P. Maruff, "The neuropsychology of preclinical Alzheimer's disease and mild cognitive impairment," *Neurosci. Biobehavioral Rev.*, vol. 24, no. 3, pp. 365–374, May 2000.

[2] S. Gauthier, P. Rosa-Neto, J. A. Morais, and C. Webster, "World Alzheimer report 2021: Journey through the diagnosis of dementia," Alzheimer's Disease Int., London, U.K., Tech. Rep., 2021.

[3] D. Le Bihan, J. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat, "Diffusion tensor imaging: Concepts and applications," *J. Magn. Reson. Imag.*, vol. 13, no. 4, pp. 534–546, Apr. 2001.

[4] C. Pierpaoli, P. Jezzard, P. J. Basser, A. Barnett, and G. Di Chiro, "Diffusion tensor MR imaging of the human brain," *Radiology*, vol. 201, no. 3, pp. 637–648, Dec. 1996.

[5] Y. Zhang, N. Schuff, G.-H. Jahng, W. Bayne, S. Mori, L. Schad, S. Mueller, A.-T. Du, J. H. Kramer, K. Yaffe, H. Chui, W. J. Jagust, B. L. Miller, and M. W. Weiner, "Diffusion tensor imaging of cingulum fibers in mild cognitive impairment and Alzheimer disease," *Neurology*, vol. 68, no. 1, pp. 13–19, Jan. 2007.

[6] M. Bozzali, S. E. MacPherson, M. Cercignani, W. R. Crum, T. Shallice, and J. Rees, "White matter integrity assessed by diffusion tensor tractography in a patient with a large tumor mass but minimal clinical and neuropsychological deficits," *Funct. Neurol.*, vol. 27, no. 4, pp. 239–246, Oct. 2012.

[7] F. Scarselli, A. C. Tsoi, M. Gori, and M. Hagenbuchner, "Graphical-based learning environments for pattern recognition," in *Structural, Syntactic, and Statistical Pattern Recognition* (Lecture Notes in Computer Science), Aug. 2004, pp. 42–56.

[8] T. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[9] T.-A. Song, S. R. Chowdhury, F. Yang, H. Jacobs, G. E. Fakhri, Q. Li, K. Johnson, and J. Dutta, "Graph convolutional neural networks for Alzheimer's disease classification," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Venice, Italy, Apr. 2019, pp. 414–417.

[10] H. Kong and S. Wang, "Adversarial learning based structural brain-network generative model for analyzing mild cognitive impairment," 2022, *arXiv:2208.08896*.

[11] J. Yang, S. Wang, and T. Wu, "Maximum mutual information for feature extraction from graph-structured data: Application to Alzheimer's disease classification," *Int. J. Speech Technol.*, vol. 53, no. 2, pp. 1870–1886, Jan. 2023.

[12] Y. Qiu, S. Yu, Y. Zhou, D. Liu, X. Song, T. Wang, and B. Lei, "Multi-channel sparse graph transformer network for early Alzheimer's disease identification," in *Proc. IEEE 18th Int. Symp. Biomed. Imag. (ISBI)*, Nice, France, Apr. 2021, pp. 1794–1797.

[13] X. Song, F. Zhou, A. F. Frangi, J. Cao, X. Xiao, Y. Lei, T. Wang, and B. Lei, "Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101947.

[14] J. Pan, B. Lei, Y. Shen, Y. Liu, Z. Feng, and S. Wang, "Characterization multimodal connectivity of brain network by hypergraph GAN for Alzheimer's disease analysis," *Pattern Recognit.*, vol. 13021, pp. 467–478, Oct. 2021.

[15] J. Pan and S. Wang, "Cross-modal transformer GAN: A brain structure-function deep fusing framework for Alzheimer's disease," 2022, *arXiv:2206.13393*.

[16] B. Lei, Y. Zhu, S. Yu, H. Hu, Y. Xu, G. Yue, T. Wang, C. Zhao, S. Chen, P. Yang, X. Song, X. Xiao, and S. Wang, "Multi-scale enhanced graph convolutional network for mild cognitive impairment detection," *Pattern Recognit.*, vol. 134, Feb. 2023, Art. no. 109106.

[17] M. Fang, Z. Jin, F. Qin, Y. Peng, C. Jiang, and Z. Pan, "Re-transfer learning and multi-modal learning assisted early diagnosis of Alzheimer's disease," *Multimedia Tools Appl.*, vol. 81, no. 20, pp. 29159–29175, Aug. 2022.

[18] L. Houria, N. Belkhamsa, A. Cherfa, and Y. Cherfa, "Multi-modality MRI fusion for Alzheimer's disease detection using deep learning," *Phys. Eng. Sci. Med.*, vol. 45, no. 4, pp. 1043–1053, Feb. 2022.

[19] S. Srivishagan, L. Kumaralingam, K. Thanikasalam, U. A. J. Pinidiyaarachchi, and N. Ratnarajah, "Discriminative patterns of white matter changes in Alzheimer's," *Psychiatry Res., Neuroimag.*, vol. 328, Jan. 2023, Art. no. 111576.

[20] S. Kolahkaj and H. Zare, "A connectome-based deep learning approach for early MCI and MCI detection using structural brain networks," *Neurosci. Informat.*, vol. 3, no. 1, Mar. 2023, Art. no. 100118.

[21] X. Tian, Y. Liu, L. Wang, X. Zeng, Y. Huang, and Z. Wang, "An extensible hierarchical graph convolutional network for early Alzheimer's disease identification," *Comput. Methods Programs Biomed.*, vol. 238, Aug. 2023, Art. no. 107597.

[22] C. Luo, M. Li, R. Qin, H. Chen, L. Huang, D. Yang, Q. Ye, R. Liu, Y. Xu, H. Zhao, and F. Bai, "Long longitudinal tract lesion contributes to the progression of Alzheimer's disease," *Frontiers Neurol.*, vol. 11, 2020, Art. no. 503235.

[23] E. Lella, A. Pazienza, D. Lofù, R. Anglani, and F. Vitulano, "An ensemble learning approach based on diffusion tensor imaging measures for Alzheimer's disease classification," *Electronics*, vol. 10, no. 3, p. 249, Jan. 2021.

[24] N. Xia, Y. Li, Y. Xue, W. Li, Z. Zhang, C. Wen, J. Li, and Q. Ye, "Intravoxel incoherent motion diffusion-weighted imaging in the characterization of Alzheimer's disease," *Brain Imag. Behav.*, vol. 16, no. 2, pp. 617–626, Apr. 2022.

[25] A. De and A. S. Chowdhury, "DTI based Alzheimer's disease classification with rank modulated fusion of CNNs and random forest," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114338.

[26] A. Demirhan, T. M. Nir, A. Zavaliangos-Petropulu, C. R. Jack, M. W. Weiner, M. A. Bernstein, P. M. Thompson, and N. Jahanshad, "Feature selection improves the accuracy of classifying Alzheimer disease using diffusion tensor images," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Brooklyn, NY, USA, Apr. 2015, pp. 126–130.

[27] T. Maggipinto, R. Bellotti, N. Amoroso, D. Diacono, G. Donvito, E. Lella, A. Monaco, M. A. Scelsi, and S. Tangaro, "DTI measurements for Alzheimer's classification," *Phys. Med. Biol.*, vol. 62, no. 6, pp. 2361–2375, Mar. 2017.

[28] J. L. D. da Rocha, I. Bramati, G. Coutinho, F. T. Moll, and R. Sitaram, "Fractional anisotropy changes in parahippocampal cingulum due to Alzheimer's disease," *Sci. Rep.*, vol. 10, no. 1, 2020, Art. no. 2660.

[29] B. Bigham, S. A. Zamanpour, and H. Zare, "Features of the superficial white matter as biomarkers for the detection of Alzheimer's disease and mild cognitive impairment: A diffusion tensor imaging study," *Heliyon*, vol. 8, no. 1, Jan. 2022, Art. no. e08725.

[30] L. Cao, B. R. Schrank, S. Rodriguez, E. G. Benz, T. W. Moulia, G. T. Rickenbacher, A. C. Gomez, Y. Levites, S. R. Edwards, T. E. Golde, B. T. Hyman, G. Barnea, and M. W. Albers, "Aβ alters the connectivity of olfactory neurons in the absence of amyloid plaques in vivo," *Nature Commun.*, vol. 3, no. 1, 2012, Art. no. 1009.

[31] X. Dou, H. Yao, D. Jin, F. Feng, P. Wang, B. Zhou, B. Liu, Z. Yang, N. An, X. Zhang, and Y. Liu, "Characterizing white matter connectivity in Alzheimer's disease and mild cognitive impairment: Automated fiber quantification," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Venice, Italy, Apr. 2019, pp. 117–121.
[32] D. B. Stone, S. G. Ryman, A. P. Hartman, C. J. Wertz, and A. A. Vakhtin, "Specific white matter tracts and diffusion properties predict conversion from mild cognitive impairment to Alzheimer's disease," *Frontiers Aging Neurosci.*, vol. 13, 2021, Art. no. 711579.
[33] C. Ye, S. Mori, P. Chan, and T. Ma, "Connectome-wide network analysis of white matter connectivity in Alzheimer's disease," *NeuroImage, Clin.*, vol. 22, Feb. 2019, Art. no. 101690.
[34] J. P. J. Savarraj, R. Kitagawa, D. H. Kim, and H. A. Choi, "White matter connectivity for early prediction of Alzheimer's disease," *Technol. Health Care*, vol. 30, no. 1, pp. 17–28, Dec. 2021.
[35] F. He, Y. Li, C. Li, J. Zhao, T. Liu, L. Fan, X. Zhang, and J. Wang, "Changes in the connection network of whole-brain fiber tracts in patients with Alzheimer's disease have a tendency of lateralization," *NeuroReport*, vol. 32, no. 14, pp. 1175–1182, Oct. 2021.
[36] W. Huang, X. Li, X. Li, G. Kang, Y. Han, and N. Shu, "Combined support vector machine classifier and brain structural network features for the individual classification of amnestic mild cognitive impairment and subjective cognitive decline patients," *Frontiers Aging Neurosci.*, vol. 13, 2021, Art. no. 687927.
[37] T. Lyu and H. Guo, "BGCN: An EEG-based graphical classification method for Parkinson's disease diagnosis with heuristic functional connectivity speculation," in *Proc. 11th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Baltimore, MD, USA, Apr. 2023, pp. 1–4.
[38] L. Huang, X. Ye, M. Yang, L. Pan, and S. H. Zheng, "MNC-Net: Multi-task graph structure learning based on node clustering for early Parkinson's disease diagnosis," *Comput. Biol. Med.*, vol. 152, Jan. 2023, Art. no. 106308.
[39] J. Zhang, J. Lim, M.-H. Kim, S. Hur, and T.-M. Chung, "WM-STGCN: A novel spatiotemporal modeling method for parkinsonian gait recognition," *Sensors*, vol. 23, no. 10, p. 4980, May 2023.
[40] Y. Wang, H. Long, Q. Zhou, T. Bo, and J. Zheng, "PLSNet: Position-aware GCN-based autism spectrum disorder diagnosis via FC learning and Rois sifting," *Comput. Biol. Med.*, vol. 163, Sep. 2023, Art. no. 107184.
[41] L. Li, G. Wen, P. Cao, X. Liu, O. R. Zaiane, and J. Yang, "Exploring interpretable graph convolutional networks for autism spectrum disorder diagnosis," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 18, no. 4, pp. 663–673, Nov. 2022.
[42] M. R. Lamani, P. J. Benadit, and K. Vaithinathan, "Multi-atlas graph convolutional networks and convolutional recurrent neural networks-based ensemble learning for classification of autism spectrum disorders," *SN Comput. Sci.*, vol. 4, no. 3, 2023, Art. no. 213.
[43] M. Zhu, Y. Quan, and X. He, "The classification of brain network for major depressive disorder patients based on deep graph convolutional neural network," *Frontiers Hum. Neurosci.*, vol. 17, 2023, Art. no. 1094592.
[44] E. N. Pitsik, V. A. Maximenko, S. A. Kurkin, A. P. Sergeev, D. Stoyanov, R. Paunova, S. Kandilarova, D. Simeonova, and A. E. Hramov, "The topology of fMRI-based networks defines the performance of a graph neural network for the classification of patients with major depressive disorder," *Chaos, Solitons Fractals*, vol. 167, Feb. 2023, Art. no. 113041.
[45] P. Dai, D. Lu, Y. Shi, Y. Zhou, T. Xiong, X. Zhou, Z. Chen, B. Zou, H. Tang, Z. Huang, and S. Liao, "Classification of recurrent major depressive disorder using a new time series feature extraction method through multisite RS-fMRI data," *J. Affect. Disorders*, vol. 339, pp. 511–519, Oct. 2023.
[46] X. Chen, J. Zhou, P. Ke, J. Huang, D. Xiong, Y. Huang, G. Ma, Y. Ning, F. Wu, and K. Wu, "Classification of schizophrenia patients using a graph convolutional network: A combined functional MRI and connectomics analysis," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104293.
[47] X. Chen, P. Ke, Y. Huang, J. Zhou, H. Li, R. Peng, J. Huang, L. Liang, G. Ma, X. Li, Y. Ning, F. Wu, and K. Wu, "Discriminative analysis of schizophrenia patients using graph convolutional networks: A combined multimodal MRI and connectomics analysis," *Frontiers Neurosci.*, vol. 17, Mar. 2023.
[48] S. Ghosh, E. Bhargava, C.-T. Lin, and S. S. Nagarajan, "Graph convolutional learning of multimodal brain connectome data for schizophrenia classification," *bioRxiv*, 2023, doi: 10.1101/2023.01.05.522960.
[49] G. Jayawardena, S. Jayarathna, and Y. He, "ADHD prediction through analysis of eye movements with graph convolution network," Ph.D. dissertation, Dept. Comput. Sci., College Sci., Old Dominion Univ., Norfolk, VA, USA, 2023.
[50] Y. Tang, D. Chen, J. Wu, W. Tu, J. J. M. Monaghan, P. Sowman, and D. Mcalpine, "Functional connectivity learning via siamese-based SPD matrix representation of brain imaging data," *Neural Netw.*, vol. 163, pp. 272–285, Jun. 2023.
[51] L. Liu, G. Wen, P. Cao, T. Hong, J. Yang, X. Zhang, and O. R. Zaiane, "BrainTGL: A dynamic graph representation learning model for brain network analysis," *Comput. Biol. Med.*, vol. 153, Feb. 2023, Art. no. 106521.
[52] G. Wen, P. Cao, L. Liu, J. Yang, X. Zhang, F. Wang, and O. R. Zaiane, "Graph self-supervised learning with application to brain networks analysis," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 8, pp. 4154–4165, Aug. 2023.
[53] X. Ouyang, K. Chen, L. Yao, X. Wu, J. Zhang, K. Li, Z. Jin, and X. Guo, "Independent component analysis-based identification of covariance patterns of microstructural white matter damage in Alzheimer's disease," *PLoS One*, vol. 10, no. 3, Mar. 2015, Art. no. e0119714.
[54] C. D. Mayo, E. L. Mazerolle, L. Ritchie, J. D. Fisk, and J. R. Gawryluk, "Longitudinal changes in microstructural white matter metrics in Alzheimer's disease," *NeuroImage, Clin.*, vol. 13, pp. 330–338, Jan. 2017.
[55] T. Kipf, "Deep learning with graph-structured representations," Ph.D. dissertation, Informat. Inst., University of Amsterdam, Amsterdam, The Netherlands, 2020.
[56] H. Kong, J. Pan, Y. Shen, and S. Wang, "Adversarial learning based structural brain-network generative model for analyzing mild cognitive impairment," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*, Aug. 2022, pp. 361–375.

YILIN SANG received the B.S. degree from the Henan University of Engineering, in 2020, and the M.S. degree from Beijing Technology and Business University, in 2024, where he is currently pursuing the M.S. degree in computer technology. His research interests include medical image classification and deep learning.

WAN LI received the B.S. degree from Zhengzhou University and the Ph.D. degree from the Beijing University of Technology. She is currently an Assistant Professor with Beijing Technology and Business University. Her research interests include medical image processing and deep learning.