Graph Reasoning Module for Alzheimer's Disease Diagnosis: A Plug-and-Play Method

Chen-Chen Fan, Hongjun Yang, *Member, IEEE*, Chutian Zhang, Liang Peng, Xiaohu Zhou, *Member, IEEE*, Shiqi Liu, Sheng Chen, and Zeng-Guang Hou, *Fellow, IEEE*

***Abstract*— Recent advances in deep learning have led to increased adoption of convolutional neural networks (CNN) for structural magnetic resonance imaging (sMRI)-based Alzheimer's disease (AD) detection. AD results in widespread damage to neurons in different brain regions and destroys their connections. However, current CNN-based methods struggle to relate spatially distant information effectively. To solve this problem, we propose a graph reasoning module (GRM), which can be directly incorporated into CNN-based AD detection models to simulate the underlying relationship between different brain regions and boost AD diagnosis performance. Specifically, in GRM, an adaptive graph Transformer (AGT) block is designed to adaptively construct a graph representation based on the feature map given by CNN, a graph convolutional network (GCN) block is adopted to update the graph representation, and a feature map reconstruction (FMR) block is built to convert the learned graph representation to a feature map. Experimental results demonstrate that the insertion of the GRM in the existing AD classification model can increase its balanced accuracy by more than 4.3%. The GRM-embedded model achieves state-of-the-art performance compared with current deep learning-based AD diagnosis methods, with a balanced accuracy of 86.2%.**

***Index Terms*— Alzheimer's disease diagnosis, graph convolution network, plug-and-play, structural magnetic resonance imaging.**

Manuscript received 2 August 2023; revised 6 November 2023; accepted 26 November 2023. Date of publication 28 November 2023; date of current version 7 December 2023. This work was supported in part by the National Natural Science Foundation of China under Grant U1913601, Grant 62073319, Grant 62003343, and Grant U20A20224; in part by the Beijing Natural Science Foundation under Grant L212034; and in part by the Beijing Nova Program under Grant 20220484175. *(Corresponding authors: Hongjun Yang; Zeng-Guang Hou.)*

Chen-Chen Fan is with the Medical Big Data Research Center, Chinese PLA General Hospital, Beijing 100853, China, and also with the State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China.

Hongjun Yang, Liang Peng, Xiaohu Zhou, and Shiqi Liu are with the State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: hongjun.yang@ia.ac.cn).

Chutian Zhang is with the Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology, Macau 999078, China.

Sheng Chen is with the State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China.

Zeng-Guang Hou is with the State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the CASIA-MUST Joint Laboratory of Intelligence Science and Technology, Macau Institute of Systems Engineering, Macau University of Science and Technology, Macau 999078, China (e-mail: zengguang.hou@ia.ac.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3337533

I. Introduction

AD IS the most common neurodegenerative disease, affecting millions of people worldwide. While there are no specific remedies for AD, early and accurate diagnoses can effectively improve patients' living quality [1]. Structural magnetic resonance imaging (sMRI) is a widely used imaging technique in research as well as in clinical practice [2]. The sMRI is an extensively utilized imaging modality in AD detection because of its non-invasive, high resolution, and moderate cost characteristics [3]. While sMRI can capture brain structure and connection abnormalities caused by AD, the clinician's competence often affects the diagnosis accuracy based on sMRI. Fortunately, computer-aided diagnosis technology offers a possible way to detect AD through sMRI quantitatively and effectively. Fortunately, computer-aided diagnosis technology has been widely used in the medical field [4], [5], [6], [7], which provides a possible method for quantitatively and effectively detecting AD through sMRI.

Current sMRI-based AD detections mostly adopt CNN [8]. Studies have tried to adopt 2D CNNs for AD detection with input composed of 2D slices from 3D sMRI [9], [10]. The conversion to 2D images can avoid overfitting as it expands the dataset and makes it possible to transfer existing pre-trained CNNs with outstanding performance on image classification tasks to AD detection. However, the absence of 3D information impairs performance. To better take advantage of spatial information, 3D CNNs have been adopted. Some studies selected 3D patches from the whole sMRI and trained independent CNNs for each patch [11], [12]. Though the 3D structure was preserved, the patches selected may not be informative as they may contain disease-irrelevant information, thus causing the model to be computationally inefficient.

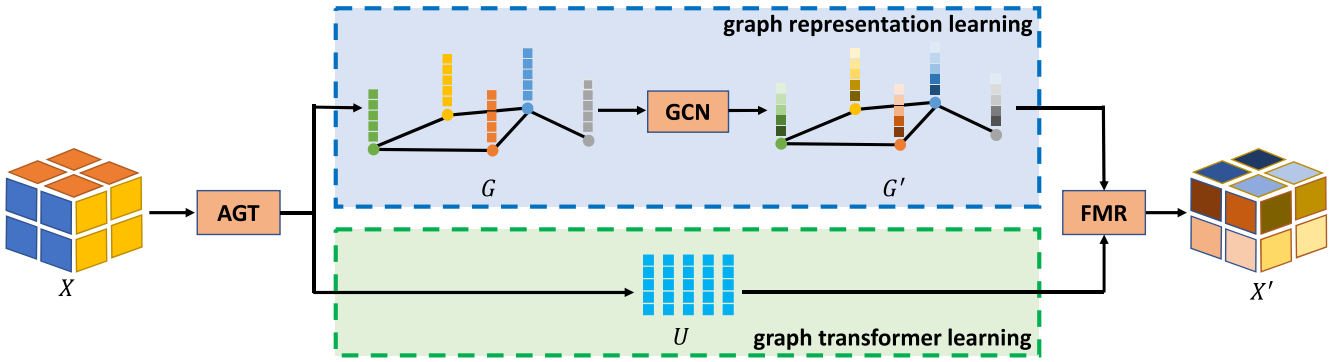

Fig. 1. The framework of GRM. GRM consists of an AGT block, a GCN block, and a FMR block, and can be divided into a graph representation learning branch and a graph transformer learning branch. The AGT block adaptively constructs the graph representation G from the input feature map X and generates an update matrix U for graph transformer learning. A GCN block is adopted in the graph representation learning branch for graph feature extraction. The outputs of the two branches are combined and processed by the FMR block to reconstruct a new feature map X' that shares the same shape with X.

Region of interest (ROI)-based methods focus on informative disease-related regions, effectively decreasing the complexity of the framework. The hippocampus is commonly used as an ROI as it has been affected since the early stages of AD [13], [14]. Data extracted from the hippocampus is sent to CNN for classification [15]. However, as AD also affects the connections between multiple brain regions, focusing only on selected ROIs fails to cover all possible pathological locations in the whole brain. Some studies have tried to use the whole sMRI as input directly, performing a subject-level classification [16], [17]. It has been shown that different 3D CNNs can achieve similar performances, which are better than the 2D slice approaches [8].

Although CNN has attracted significant attention in AD detection, its convolutional formulation is limited to data structured in an ordered grid-like fashion [18]. It has trouble relating spatially distant information as its feature extraction depends on receptive fields that operate in constrained local neighborhoods [19]. Approaches like deep-stacked convolutional operations, increased kernel sizes, and new operations like non-local attention layers have been used to tackle this problem. However, they increased the computational complexity and could only mitigate the problem to some extent [20].

Therefore, it is not easy for CNN to directly capture the potential relationship between different brain regions related to AD. The intrinsic nature of AD patient's brain structure and connectivity abnormalities motivates the exploration of brain network representations based on graph theory to quantitatively evaluate the connectivity properties of brain networks. The graph convolutional network (GCN) generalizes the convolution operation from grid data to graph representations [21] and makes it possible to model the relationship between local features extracted by CNN. To date, only a few studies explored AD diagnosis by GCN with sMRI data. Their graph representation constructions relied on manually selected features like cortical thickness and geometry from MRI scans [22], [23]. Constructing an adaptive graph representation may contribute to a better representation of the MRI data in a data-driven way, and might provide richer information for the reasoning of the graph network.

In this work, we design a graph reasoning module (GRM), which can be directly incorporated into a CNN-based AD detection model to boost performance. The feature representation extracted by CNN is processed by GCN to generate a relation-aware representation. As shown in Fig. 1, GRM consists of an adaptive graph Transformer (AGT) block, a GCN block, and a feature map reconstruction (FMR) block. The GRM can be divided into two branches, graph representation learning and graph transformer learning. The AGT block adaptively transforms the input feature map into a corresponding graph representation, which is then sent to the GCN block in the graph representation learning branch for feature extraction. Also, an update matrix U is constructed to update the parameters in the AGT. The outputs of the two branches are then merged and sent to the FMR block to reconstruct a new feature map that shares the same size as the input.

Our main contributions can be summarized as follows:

- We propose a plug-and-play module, GRM, which can be efficiently inserted into CNN-based AD detection models to capture the underlying relationship between spatially distant regions.
- We design an AGT block in GRM to adaptively convert the input feature map into a graph representation without requiring time-consuming feature extractions.
- We embed the GRMs into ResNet18 and ResNet34, which increase their balanced AD classification accuracy by more than 4.3%, achieving state-of-the-art performance compared with current deep learning-based AD diagnosis methods.

II. Methods

In this section, we describe the proposed GRM, as shown in Fig. 1. The AGT block adaptively converts the input feature map to a graph representation, which is then sent to the graph representation learning branch and processed by the GCN block for feature extraction. As for the graph transformer

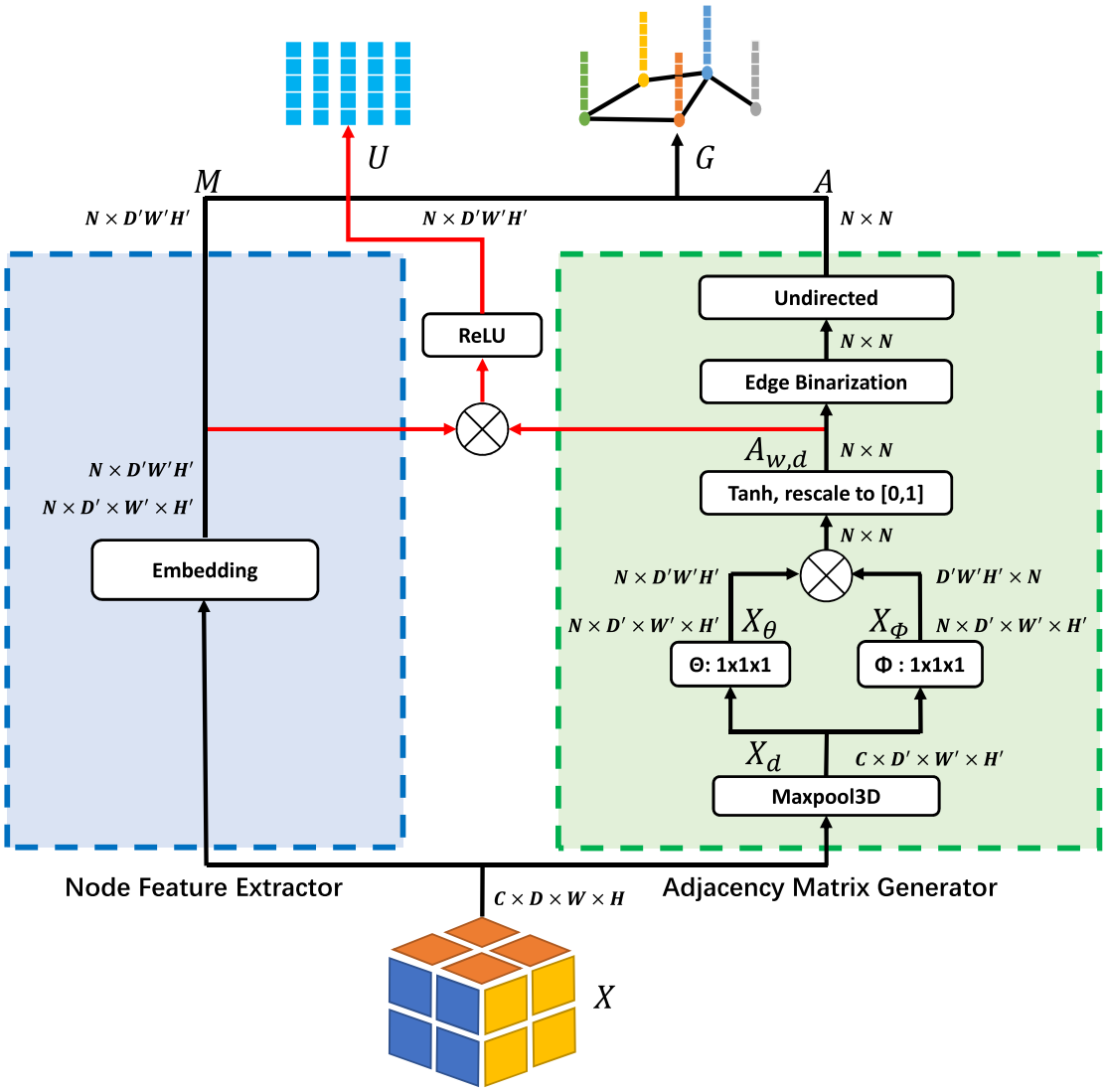

Fig. 2. The framework of the proposed AGT block. The feature maps are shown as the shape of their tensors."⊗" denotes matrix multiplication.

learning branch, an update matrix U is constructed to update the parameters of the AGT block during training. The FMR block is adopted to reconstruct corresponding feature maps by combining the GCN block's output and update matrix.

A. AGT Block

The AGT block transforms the input feature map $X \in \mathbf{R}^{C\times D\times W\times H}$ (where C, D, H, W are the channel, depth, height, and width, respectively) into a graph $G = (V, E)$ with N nodes $v_i \in V$ and edges $(v_i, v_j) \in E$. The graph is characterized by its adjacency matrix A and node feature matrix M. As shown in Fig. 2, A and M are obtained from the adjacency matrix generator and node feature extractor, respectively. An update matrix U is constructed to update the parameters of the AGT block.

1) Node Feature Extraction: The input feature map X is processed by a 3D convolution layer, denoted as Emb$(\cdot)$, with kernel size $1 \times 1 \times 1$ and stride 2 to extract a new feature embedding, which is then reshaped to obtain the node feature matrix $M \in \mathbf{R}^{N\times \frac{DWH}{8}}$. Each node is associated with a $\frac{DWH}{8}$-dimensional feature vector.

$$M = R[\text{Emb}(X)] \tag{1}$$

where R represents the reshape operation which converts the embedded feature with shape $\mathbf{R}^{N\times D'\times W'\times H'}$ to $\mathbf{R}^{N\times D'W'H'}$ $(D' = D/2, W' = W/2, H' = H/2)$.

2) Adjacency Matrix Generation: As illustrated in Fig. 2, for a given feature map X, we first downsample it by a 3D maxpooling layer, denoted as Maxpool$(\cdot)$, to reduce its dimensionality and computational complexity. A smaller feature map X_d is obtained:

$$X_d = \text{Maxpool}(X) \tag{2}$$

where $X \in \mathbf{R}^{C\times D\times W\times H}$ and $X_d \in \mathbf{R}^{C\times D'\times W'\times H'}$. Then, X_d is fed into two 3D convolution layers, $\Theta(\cdot)$ and $\Phi(\cdot)$, to generate new feature vectors, denoted as X_θ and X_ϕ:

$$X_\phi = R[\Phi(X_d)]^T \tag{3}$$

$$X_\theta = R[\Theta(X_d)] \tag{4}$$

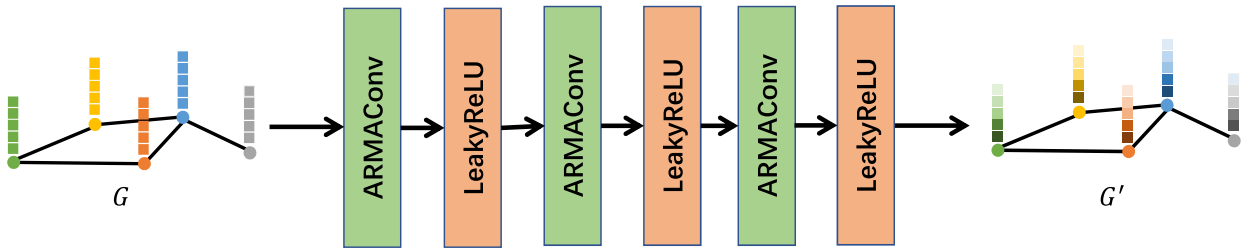

Fig. 3. The framework of GCN block. GCN block adopts three ARMAConv [24] layers with LeakyReLU activation function to aggregate and update the graph representation.

X_ϕ and X_θ is then reshaped into $\mathbf{R}^{D'W'H'\times N}$ and $\mathbf{R}^{N\times D'W'H'}$, respectively. We perform matrix multiplication between X_θ and X_ϕ. The result is further processed by Tanh activation function and then rescaled to range [0, 1] to get the weighted directed adjacency matrix $A_{w,d}$:

$$A_{w,d} = \mathrm{Tanh}(X_\theta X_\phi) \tag{5}$$

$A_{w,d}$ is processed by edge binarization, where a threshold t is set, and only those edge values that exceed t are retained and set to one. Then we convert the graph to an undirected graph. The GCN block takes the adaptive graph represenataion generated by the AGT block as input, which is characterized by node feature matrix M given in Eq.(1) and adjacency matrix A.

3) Update Matrix Construction: An additional branch $U \in \mathbf{R}^{N\times D'W'H'}$ is constructed by merging the generated node feature matrix M given in Eq.(1) and the weighted directed adjacency matrix $A_{w,d}$ given in Eq.(5) to update the parameters of the AGT block during training:

$$U = \mathrm{ReLU}[A_{w,d}M] \tag{6}$$

B. GCN Block

The GCN block receives the node feature matrix M and adjacency matrix A generated by the AGT block and processes them by three graph convolution layers with ReLU activation functions to aggregate and update the representation of node features, as shown in Fig. 3.

We adopt the graph method from [25]. For input signal $M \in \mathbf{R}^{N\times D'W'H'}$ where each node is denoted by a $D'W'H'$-dimensional feature vector, the convolutional operator can be denoted as:

$$M' = \hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}M\Psi \tag{7}$$

where $\Psi \in \mathbf{R}^{D'W'H'\times D'W'H'}$ is the matrix of graph filter parameters, $\hat{A} = A + I$ denotes the adjacency matrix with inserted self-loops where I is the identity matrix, and $\hat{D}_{ii} = \sum_j \hat{A}_{i,j}$ represents the degree matrix.

C. FMR Block

The FMR block is designed to reconstruct the feature map from the node feature matrix M' given by the GCN block (Eq.(7)) and the update matrix U given by the AGT block (Eq.(6)). As shown in Fig. 4, M' and U are added and then reshaped into $\mathbf{R}^{N\times D'\times W'\times H'}$. After that, a 3D transpose convolution operator is applied to reconstruct a feature map X' that shares the same shape as the original input X.

III. Experiments

A. Dataset

Data used in this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progressions of mild cognitive impairment (MCI) and early AD.

The original dataset passes through the t1-linear pipeline of Clinica [8], [26]. More precisely, bias field correction is applied using the N4ITK method [27]. Next, an affine registration is performed using the SyN algorithm [28] from ANTs [29] to align each image to the MNI space with the ICBM 2009c nonlinear symmetric template [30].

We adopt the same dataset setting as Wen et al. [8] and follow its data split method. Specifically, sMRI data of 330 CN subjects and 336 AD patients are used in our experiments. This total record is split into training (466) and testing (200) set at subject-level. The testing set consists of 100 randomly chosen subjects for each diagnostic class (i.e., 100 CN subjects and 100 AD patients). During training, we perform five-fold cross-validation. For each fold, the model with the highest balanced accuracy (BA) [31] on the validation set is saved and further tested on the testing set.

B. Implementation Details

The models are implemented using Python 3.7.9 and PyTorch on a workstation with NVIDIA Tesla V100 graphic processing unit. For all experiments, we set the batch size to 8 and carry out exhaustive grid searches for optimal learning rate and weight decay parameter combinations. The optimizer used in training is adamw. We convert the original 2D ResNet [32] to a 3D version and embed GRM in ResNet18 and ResNet34, denoted as ResNet18_GRM and ResNet34_GRM. Data are resized to $128 \times 128 \times 128$ for further processing. Their edge binarization threshold t is both set to 0.7, the weight decay is both set to 5e-4, and the learning rate is set to 5e-4 and 9e-4, respectively. The models are trained for 50 epochs, and the early-stop strategy is used. If the BA on the validation set does not increase for 10 epochs, the current experiment will be stopped. BA given in Eq.(8), sensitivity given in Eq.(9), specificity given in Eq.(10), positive predictive

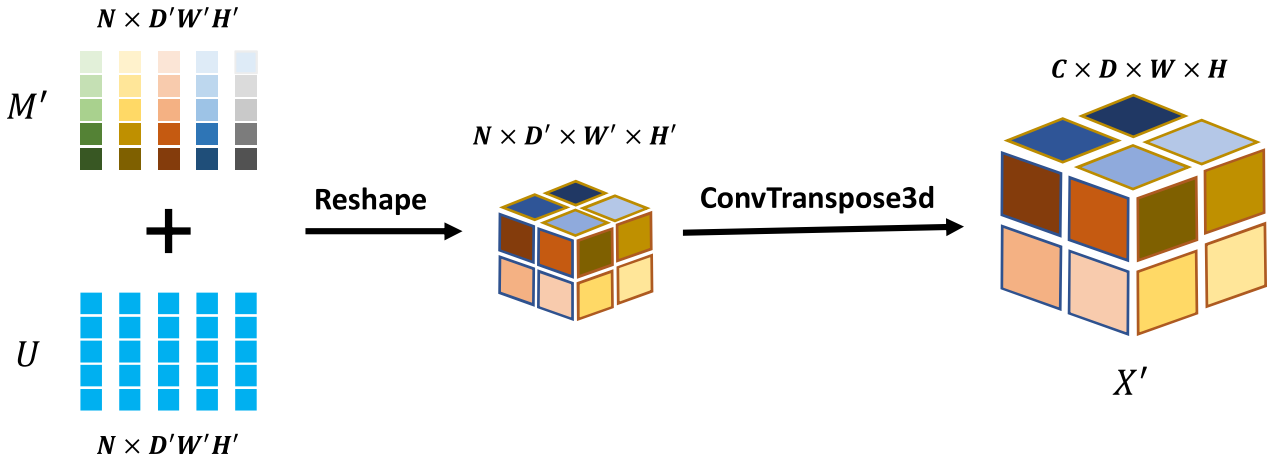

Fig. 4. The framework of the FMR module. FMR takes the output of the GCN block (M') as well as the update matrix (U) as input, and reconstruct the feature map that shares the same shape with the input of GRM.

TABLE I
ARCHITECTURES OF RESNET18_GRM AND RESNET34_GRM

layer name	output size	ResNet18_GRM	ResNet34_GRM
conv1	$64 \times 64 \times 64$	$7 \times 7 \times 7$, 64, stride 2	
max pool	$32 \times 32 \times 32$	$3 \times 3 \times 3$ max pool, stride 2	
layer 1	$16 \times 16 \times 16$	$\begin{bmatrix} 3 \times 3 \times 3, 64 \\ 3 \times 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 \times 3, 64 \\ 3 \times 3 \times 3, 64 \end{bmatrix} \times 3$
layer 2	$8 \times 8 \times 8$	$\begin{bmatrix} 3 \times 3 \times 3, 128 \\ 3 \times 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 \times 3, 128 \\ 3 \times 3 \times 3, 128 \end{bmatrix} \times 4$
GRM 1	$8 \times 8 \times 8$	feature 64, nodes 64	feature 64, nodes 64
layer 3	$8 \times 8 \times 8$	$\begin{bmatrix} 3 \times 3 \times 3, 256 \\ 3 \times 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 \times 3, 256 \\ 3 \times 3 \times 3, 256 \end{bmatrix} \times 6$
GRM 2	$8 \times 8 \times 8$	feature 64, nodes 128	feature 64, nodes 128
layer 4	$8 \times 8 \times 8$	$\begin{bmatrix} 3 \times 3 \times 3, 512 \\ 3 \times 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3 \times 3, 512 \\ 3 \times 3 \times 3, 512 \end{bmatrix} \times 3$
avg pool	512	adaptive average pool ($1 \times 1 \times 1$), dropout, flatten	
fc	2	fully connected layer (with Softmax activation function)	

value (PPV) given in Eq.(11), and negative predictive value (NPV) given in Eq.(12) are adopted to evaluate the classification performance. The average and standard deviation of each metric are recorded to compare the performance of different models.

$$\text{BA} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \tag{8}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{10}$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{11}$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \tag{12}$$

where TP represents true positive, TN represents true negative, FP represents false positive, FN represents false negative, P represents the total number of positives, and N represents the total number of negatives.

C. Results

As shown in Table I, two GRMs, located in the shallow and deep layers, are both inserted into ResNet18 and ResNet34, denoted as ResNet18_GRM and ResNet34_GRM, respectively. Comparisons have been performed with three existing state-of-the-art methods: ConvNet3D [16], DeepCNN [33], and VoxCNN [34]. The different classification experiments and the results on the validation set (validation BA during five-fold cross-validation) are detailed in

TABLE II
COMPARISON OF DIFFERENT WORK ON THE VALIDATION SET

Model	BA	Sensitivity	Specificity	PPV	NPV
ConvNet3D	56.07 ± 2.20	41.07 ± 9.32	71.07 ± 10.06	63.14 ± 4.51	51.08 ± 1.46
DeepCNN	82.47 ± 6.94	**80.07 ± 5.68**	84.87 ± 9.42	86.22 ± 8.27	**78.59 ± 5.95**
VoxCNN	78.36 ± 14.79	66.01 ± 33.13	**90.71 ± 8.49**	71.63 ± 36.29	74.37 ± 14.20
ResNet18_GRM	82.55 ± 2.25	75.40 ± 4.68	89.71 ± 3.23	**89.56 ± 2.92**	76.11 ±3.15
ResNet34_GRM	**84.91 ± 1.82**	79.63 ± 7.01	90.18 ± 5.63	89.41 ± 4.31	78.19 ± 4.51

TABLE III
COMPARISON OF DIFFERENT WORK ON THE TESTING SET

Model	BA	Sensitivity	Specificity	PPV	NPV
ConvNet3D	56.60 ± 0.86	45.40 ± 11.91	67.80 ± 12.75	60.22 ± 5.78	55.55 ± 1.11
DeepCNN	84.15 ± 3.02	**85.70 ± 3.00**	82.60 ± 4.18	83.30 ± 3.61	**85.70 ± 3.04**
VoxCNN	78.10 ± 14.10	67.40 ± 33.76	88.80 ± 6.52	68.73 ± 34.46	77.66 ± 13.89
ResNet18_GRM	**86.20 ± 2.22**	82.80 ± 2.42	89.60 ± 2.42	88.88 ± 2.42	83.98 ± 2.83
ResNet34_GRM	**86.20 ± 1.60**	82.60 ± 3.44	**89.80 ± 1.60**	**89.04 ± 1.44**	83.87 ± 2.73

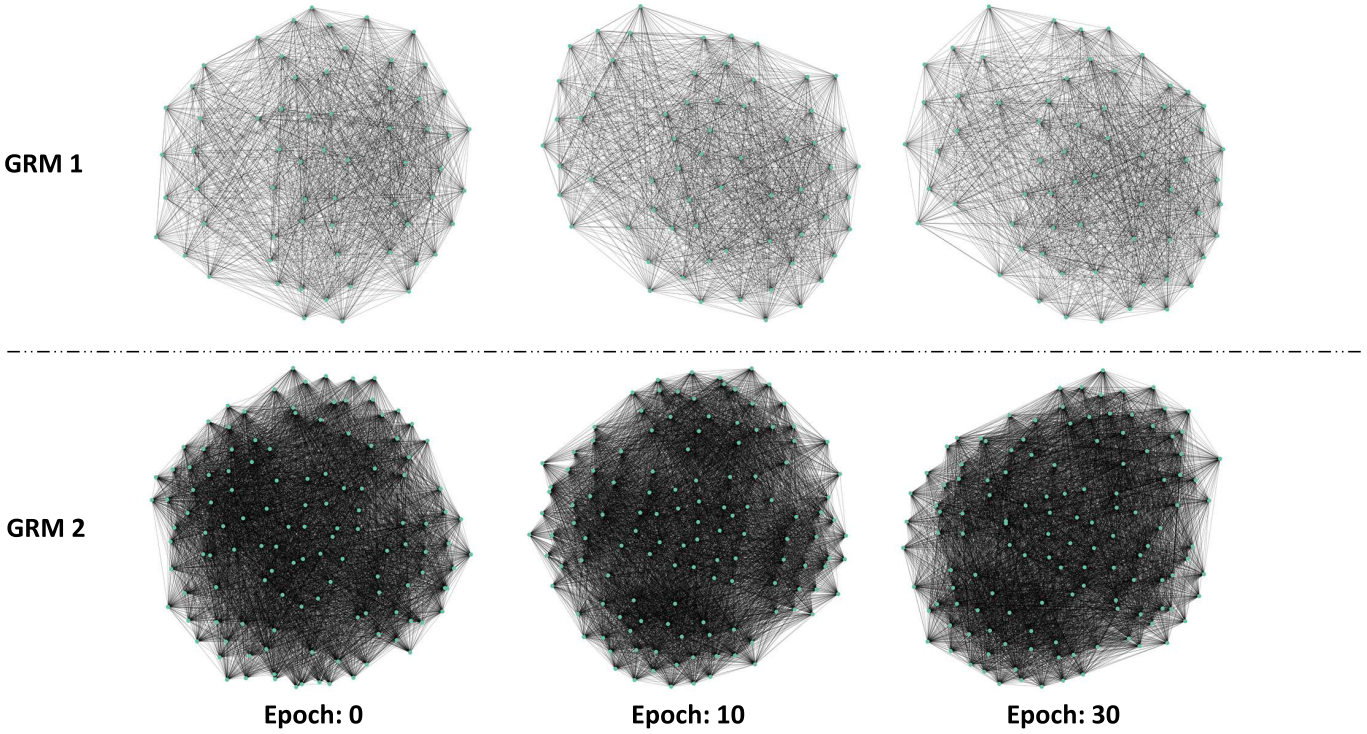

Fig. 5. Visualization results of graph data learned by the two GRMs in ResNet18_GRM.

Table II, where GRM-embedded models achieve better performances. Specifically, ResNet34_GRM has the highest BA and the second-best PPV. Compared with DeepCNN, the best-performing baseline model, ResNet34_GRM improves BA by 0.97% while decreasing the standard deviation by 5.12%. As for ResNet18_GRM, it achieves the best PPV and ranks second on BA. While VoxCNN did perform well on specificity, its standard deviation is relatively high, and BA is only 78.36%. The performance of ConvNet3D is relatively poor, which is close to random guessing.

Results on the testing set are presented in Table III. Resnet18_GRM and ResNet34_GRM receive the top BA. Meanwhile, ResNet34_GRM also has the highest specificity and PPV. Compared with the best baseline model, DeepCNN, GRM-embedded models' BA is 1.8% higher while maintaining a smaller standard deviation.

D. Ablation Study

We conduct ablation studies on the testing set to evaluate the effectiveness of the proposed GRM. Here the baseline models are referred to ConvNet3D, DeepCNN, VoxCNN, and the 3D version of ResNet18 and ResNet34. As shown in Table IV, the adoption of GRMs can significantly boost the performance of the five baseline models on the testing set while maintaining a relatively small standard deviation. Specifically, compared with the performance of corresponding baselines, BA of ResNet18_GRM and ResNet34_GRM is improved by 4.40% and 4.30%, respectively, while reducing the standard deviation

TABLE IV
ABLATION STUDY FOR GRM. PERFORMANCES ARE EVALUATED ON THE TESTING SET

Model	BA	Sensitivity	Specificity	PPV	NPV
ConvNet3D	56.60 ± 0.86	45.40 ± 11.91	67.80 ± 12.75	60.22 ± 5.78	55.55 ± 1.11
ConvNet3D_GRM	62.95 ± 4.01 (↑ 6.35)	69.60 ± 11.88	56.30 ± 15.76	65.17 ± 3.38	63.04 ± 4.98
DeepCNN	84.15 ± 3.02	**85.70 ± 3.00**	82.60 ± 4.18	83.30 ± 3.61	**85.70 ± 3.04**
DeepCNN_GRM	85.00 ± 1.27 (↑ 0.85)	82.00 ± 3.30	88.00 ± 2.90	87.36 ± 2.54	83.11 ± 2.10
VoxCNN	78.10 ± 14.10	67.40 ± 33.76	88.80 ± 6.524	68.73 ± 34.46	77.66 ± 13.89
VoxCNN_GRM	82.40 ± 1.59 (↑ 4.30)	81.40 ± 1.02	83.40 ± 2.65	83.11 ± 2.33	81.75 ± 1.08
Resnet18	81.80 ± 3.60	77.00 ± 9.57	86.60 ± 3.50	85.42 ± 2.61	79.66 ± 5.94
Resnet18_GRM	**86.20 ± 2.22 (↑ 4.40)**	82.80 ± 2.42	89.60 ± 2.42	88.88 ± 2.42	83.98 ± 2.83
Resnet34	81.90 ± 2.06	83.80 ± 1.60	80.00 ± 5.12	80.96 ± 3.77	83.18 ± 0.91
Resnet34_GRM	**86.20 ± 1.60 (↑ 4.30)**	82.60 ± 3.44	**89.80 ± 1.60**	**89.04 ± 1.44**	83.87 ± 2.73

by 1.38% and 0.46%, respectively. On the other three baseline models, the BA of ConvNet3D_GRM, DeepCNN_GRM, and VoxCNN_GRM is increased by 6.35%, 0.85%, and 4.30% respectively. Among them, the improvement of DeepCNN_GRM is smaller, which may be because the original model has achieved good results, limiting the potential for performance improvement. In terms of variance, the variance of DeepCNN_GRM and VoxCNN_GRM is 1.75% and 12.51% lower than the baseline model respectively. The variance of ConvNet3D_GRM is larger than the baseline model, which may be because the ConvNet3D model performs poorly and has abnormally low variance. The adoption of GRMs also improves the performance on other evaluation metrics like sensitivity, specificity, PPV and NPV.

E. Discussion

To better understand how the GRM encodes the graph information, we visualize the graph data learned by the two GRMs in ResNet18_GRM. As shown in Fig. 5, the two rows represent the result of GRM 1 and GRM 2 according to Table I, and different columns denote the results obtained at epochs 0, 10, and 30, respectively. GRM 1 and GRM 2 receive input feature maps with shape $128\times8\times8\times8$ and $256\times8\times8\times8$, respectively, and transform them to graph representations with 64 and 128 nodes, respectively. Each node is represented by a 64-dimensional feature vector. It can be seen that connections between nodes are evenly distributed at the beginning. As the training goes further, some nodes gradually become densely connected, and some become sparsely related, indicating that the AGT block is able to continuously adjust the construction of the graph to obtain a more discriminative graph representation.

IV. CONCLUSION

In this paper, the GRM is proposed, which can be directly embedded into existing AD detection models to boost its performance by simulating the latent relationship between spatially distant regions. Experimental results demonstrate that GRMs can be directly embedded into existing CNN-based diagnosis models and increase the balanced accuracy of AD classification by more than 4.3%. Compared with the current deep learning-based diagnosis baseline, GRM-embedded models achieve state-of-the-art performance with a balanced accuracy of 86.2%.

REFERENCES

[1] C. R. Jack et al., "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 685–691, 2008.

[2] M. Symms, H. Jäger, K. Schmierer, and T. Yousry, "A review of structural magnetic resonance neuroimaging," *J. Neurol., Neurosurg. Psychiatry*, vol. 75, no. 9, pp. 1235–1244, Sep. 2004.

[3] W. Lin et al., "Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment," *Frontiers Neurosci.*, vol. 12, p. 777, Nov. 2018.

[4] D. Bai, T. Liu, X. Han, and H. Yi, "Application research on optimization algorithm of sEMG gesture recognition based on light CNN+LSTM model," *Cyborg Bionic Syst.*, vol. 2021, Jan. 2021, Art. no. 9794610.

[5] M. Hashizume, "Perspective for future medicine: Multidisciplinary computational anatomy-based medicine with artificial intelligence," *Cyborg Bionic Syst.*, vol. 2021, Jan. 2021, Art. no. 9160478.

[6] C.-C. Fan et al., "Group feature learning and domain adversarial neural network for aMCI diagnosis system based on EEG," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 9340–9346.

[7] C.-C. Fan, H. Yang, Z.-G. Hou, Z.-L. Ni, S. Chen, and Z. Fang, "Bilinear neural network with 3-D attention for brain decoding of motor imagery movements from the human EEG," *Cognit. Neurodynamics*, vol. 15, no. 1, pp. 181–189, Feb. 2021.

[8] J. Wen et al., "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101694.

[9] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain Informat.*, vol. 5, no. 2, pp. 1–14, Dec. 2018.

[10] A. Valliani and A. Soni, "Deep residual nets for improved Alzheimer's diagnosis," in *Proc. 8th ACM Int. Conf. Bioinf., Comput. Biol., Health Informat.*, Aug. 2017, p. 615.

[11] D. Cheng, M. Liu, J. Fu, and Y. Wang, "Classification of MR brain images by combination of multi-CNNs for AD diagnosis," in *Proc. SPIE*, vol. 10420, Jul. 2017, Art. no. 1042042.

[12] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020.

[13] B. C. Dickerson et al., "MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease," *Neurobiol. Aging*, vol. 22, no. 5, pp. 747–754, 2001.

[14] N. Schuff et al., "MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers," *Brain*, vol. 132, no. 4, pp. 1067–1077, May 2008.

[15] K. Aderghal, A. Khvostikov, A. Krylov, J. Benois-Pineau, K. Afdel, and G. Catheline, "Classification of Alzheimer disease on imaging modalities with deep CNNs using cross-modal transfer learning," in *Proc. IEEE 31st Int. Symp. Computer-Based Med. Syst. (CBMS)*, Jun. 2018, pp. 345–350.

[16] K. Bäckström, M. Nazari, I. Y. Gu, and A. S. Jakola, "An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 149–153.

[17] F. Li, D. Cheng, and M. Liu, "Alzheimer's disease classification based on combination of multi-model convolutional networks," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Oct. 2017, pp. 1–5.

[18] D. Ahmedt-Aristizabal, M. Ali Armin, S. Denman, C. Fookes, and L. Petersson, "Graph-based deep learning for medical diagnosis and analysis: Past, present and future," 2021, *arXiv:2105.13137*.

[19] B. Wu et al., "Visual transformers: Token-based image representation and processing for computer vision," 2020, *arXiv:2006.03677*.

[20] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.

[21] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.

[22] V. R. Sampathkumar, "ADiag: Graph neural network based diagnosis of Alzheimer's disease," 2021, *arXiv:2101.02870*.

[23] C.-Y. Wee, C. Liu, A. Lee, J. S. Poh, H. Ji, and A. Qiu, "Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations," *NeuroImage, Clin.*, vol. 23, Jan. 2019, Art. no. 101929.

[24] F. M. Bianchi, D. Grattarola, L. Livi, and C. Alippi, "Graph neural networks with convolutional ARMA filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3496–3507, Jul. 2022.

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

[26] A. Routier et al. (Mar. 2021). *Clinica: An Open Source Software Platform for Reproducible Clinical Neuroscience Studies*. [Online]. Available: https://hal.inria.fr/hal-02308126

[27] N. J. Tustison et al., "N4ITK: Improved n3 bias correction," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010.

[28] B. Avants, C. Epstein, M. Grossman, and J. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, Feb. 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841507000606

[29] B. B. Avants, N. J. Tustison, M. Stauffer, G. Song, B. Wu, and J. C. Gee, "The insight ToolKit image registration framework," *Frontiers Neuroinform.*, vol. 8, p. 44, Apr. 2014. [Online]. Available: https://www.frontiersin.org/article/10.3389/fninf.2014.00044

[30] V. Fonov, A. Evans, R. McKinstry, C. Almli, and D. Collins, "Unbiased nonlinear average age-appropriate brain templates from birth to adulthood," *NeuroImage*, vol. 47, p. S102, Jul. 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811909708845

[31] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3121–3124.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[33] D. Cheng and M. Liu, "CNNs based multi-modality classification for AD diagnosis," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2017, pp. 1–5.

[34] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 835–838.