



# Attention based multi-task interpretable graph convolutional network for Alzheimer's disease analysis

Shunqin Jiang<sup>a</sup>, Qiyuan Feng<sup>a</sup>, Hengxin Li<sup>a</sup>, Zhenyun Deng<sup>a,\*</sup>, Qinghong Jiang<sup>b,\*</sup>

<sup>a</sup> School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

<sup>b</sup> School of Electronics and Information Engineering, Guangxi Normal University, Guangxi, China

## ARTICLE INFO

### Keywords:

Alzheimer's disease diagnosis analysis  
Multi-task learning  
Attention unit  
Interpretability  
Graph convolutional network

## ABSTRACT

Alzheimer's Disease impairs the memory and cognitive function of patients, and early intervention can effectively mitigate its deterioration. Most existing methods for Alzheimer's analysis rely solely on medical images, ignoring the impact of some clinical indicators associated with the disease. Furthermore, these methods have thus far failed to identify the specific brain regions affected by the disease. To solve these limitations, we propose an attention-based multi-task Graph Convolutional Network (GNN) for Alzheimer's disease analysis. Specifically, we first segment brain regions based on tissue types and randomly assign a learnable weight for each region. Then, we introduce multi-task attention units to jointly capture the shared feature information between brain regions and across different tasks, achieving cross-interactions between medical images and clinical indicators. Finally, we design task-specific layers for each task, allowing the model to predict Alzheimer's Disease status and clinical scores. Experimental results on four Alzheimer's Disease datasets show that our approach not only outperforms the state-of-the-art in terms of accuracy, but also explicitly identifies brain regions associated with the disease as well as provides reliable clinical scores.

## 1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disease that can significantly affect the quality of life of older individuals [1,2]. Early intervention proves to be an effective way of preventing its deterioration [3–5]. The commonly used method is to predict the disease based on Magnetic Resonance Imaging (MRI) for the purpose of early intervention [6].

In recent years, MRI-based AD diagnostic techniques have been widely used in real-world medical diagnosis, but most of them treat AD diagnosis as a binary classification task and subsequently design various models to enhance its performance [7]. Representative methods, e.g., Graph-Neural-Networks-based (GCNs-based) methods [8], predict AD status by considering the structure and sample features between samples. However, GCNs-based methods can only make binary predictions of AD diagnosis, lacking corresponding explanations for their predictions, i.e., such methods rely solely on brain regions in MRI, potentially leading to real-world diagnostic errors, such as a one-sided diagnosis or even a completely incorrect diagnosis [9]. To improve the reliability of AD diagnosis, it is necessary to provide explanations for the results of AD diagnosis or other AD-related tasks, such as the Alzheimer's Disease Assessment Scale Cognitive subscale (ADAS-Cog)

and Mini-Mental State Examination (MMSE) [10,11]. ADAS-Cog and MMSE are two common clinical indicators that are often used to assist MRI for AD diagnosis.

In real-world medical diagnostic research, the acquisition of ADAS-Cog and MMSE is time-consuming and challenging due to the need for collaboration between medical professionals and patients. They are often affected by various factors such as the patient's education, cultural background, emotional state, and etc. [12]. The common method is to first design a regression task to analyze MRI, yielding these two metrics [13], and then combine them with the previous MRI-based binary classification results for AD diagnosis. However, this method often ignores the correlation between these two tasks, potentially leading to the failure of the overall optimal of AD diagnosis.

To address these two challenges, we propose the Attention-based Multi-Task Interpretable Graph Convolutional Network (AMTI-GCN). Specifically, we first combine the weight matrix and  $L_{2,1}$  norm sparsity technique to assign the weights of the corresponding brain regions of MRI for selecting the task-related brain regions. Then, we use the attention mechanism to learn the shared features between tasks by finding the optimal ratio of feature information sharing between tasks. Finally, we feed this shared feature into the task-specific multi-task prediction

\* Corresponding authors.

E-mail addresses: [202222280722@std.uestc.edu.cn](mailto:202222280722@std.uestc.edu.cn) (S. Jiang), [202222280723@std.uestc.edu.cn](mailto:202222280723@std.uestc.edu.cn) (Q. Feng), [202022081424@std.uestc.edu.cn](mailto:202022081424@std.uestc.edu.cn) (H. Li), [zdeng@uestc.edu.cn](mailto:zdeng@uestc.edu.cn) (Z. Deng), [qhjiang@mailbox.gxnu.edu.cn](mailto:qhjiang@mailbox.gxnu.edu.cn) (Q. Jiang).

<https://doi.org/10.1016/j.patrec.2024.02.016>

Received 23 August 2023; Received in revised form 13 January 2024; Accepted 19 February 2024

Available online 22 February 2024

0167-8655/© 2024 Published by Elsevier B.V.

layer to achieve the binary prediction of AD and its corresponding ADAS and MMSE. Since these two metrics are explicit and verifiable, they can be used as our interpretation of AD prediction.

We conclude the main contributions of our paper as follows:

- We propose a multi-task module for AD and the corresponding clinical indicators (ADAS-Cog, MMSE) prediction, making AD predictions more explainable.
- We designed an attention-based method to better leverage the correlation information between multiple tasks.
- We conduct comprehensive experiments on four AD datasets and demonstrate that our proposed method outperforms the SOTA in terms of accuracy.

## 2. Related work

**Interpretability research.** In medical applications, the interpretability of a model is crucial in terms of whether it earns the trust of physicians and patients. Various techniques have been explored to improve the transparency of models in AD diagnosis. For example, Oh and Yoon et al. [14] utilize the generation of a counterfactual map behind a diagnostic model to localize hypothetical abnormalities within a normal brain image, thereby guiding the diagnosis of the next enhanced diagnostic model. This approach not only improves diagnostic accuracy, but also provides clinicians with visual insights into the model's decision-making process. Meanwhile, Chen et al. [15] utilized forward selection and aspect consolidation to output accurate brain regions associated with an AD diagnosis, thus helping to identify key factors leading to an AD diagnosis.

**Multi-task research.** Multi-task learning has been used to predict multiple relevant clinical outcomes simultaneously [16], providing a way to capture shared information and improve overall prediction performance. Recent studies have been conducted in AD diagnosis using a multi-task framework to jointly predict disease progression and cognitive scores, showing promising results. For example, Liu et al. [17] proposed a deep multi-task multi-channel learning (DM2L) framework to simultaneously output AD classification and clinical score regression results with improved prediction accuracy compared to a single-task model. In addition, Liang et al. [18] proposed a multi-task learning framework that adaptively estimates missing values and predicts future progression of the condition over time based on subjects' historical measurements.

## 3. Method

In this section, we will introduce the proposed network called the Attention based Multi-task Interpretable Graph Convolutional Network (AMTI-GCN). As shown in Fig. 1, AMTI-GCN includes three modules, which are called the Interpretation module, the Feature sharing module, and the Task-specific module.

Remarkably, throughout the entire paper, we use boldface uppercase letters, boldface lowercase letters, and regular italic letters to represent matrices, vectors, and scalars, respectively. To be specific, in this paper,  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{n \times d}$  denotes the feature matrix of  $n$  samples,  $\mathbf{x}_{i,k}$  represents the  $i$ th sample of the  $k$ th task, and  $x_{ij,k}$  is the  $j$ th feature of the  $i$ th sample of the  $k$ th task.

### 3.1. Interpretation module

We select the brain regions significant to the specific task during training by generating the feature weight matrix and assigning the weights to the corresponding features of the samples. The concrete method is divided into the following three steps:

Firstly, to give different features weights corresponding to their importance to the task, we initialize a weight matrix. In particular, we consider the matrix  $\mathbf{W}_k = [\mathbf{w}_{1,k}, \mathbf{w}_{2,k}, \dots, \mathbf{w}_{d,k}] \in \mathbb{R}^{n \times d}$ ,  $k \in \{1, 2, \dots, m\}$

as the weight matrix of the  $k$ th task, which is a trainable matrix initialized randomly.

Secondly, since redundant features can negatively affect the final classification and regression results, we propose the feature sparsity approach to reduce the redundant features and obtain the important brain regions related to the specific task [19,20]. To sparse the features, we aim to make the corresponding weight  $w_{i,k}$  to be smaller or even zero when the  $i$ th feature is not important. So we utilize  $L_{2,1}$ -norm which allows the feature weights of each row to be as small as possible. Meanwhile, the mathematical expression of  $L_{2,1}$  is  $\|\mathbf{W}_k\|_{2,1}$ .

Thirdly, we need to assign the weights after sparseness in the second step to the corresponding features. Specifically, we perform the Hadamard product operation on the sparse weight matrix  $\mathbf{W}_k$  and the feature matrix  $\mathbf{X}$ , which can be expressed as follows:

$$\mathbf{X}_k^{(0)} = \mathbf{W}_k \odot \mathbf{X}, \quad (1)$$

where  $\odot$  represents Hadamard product, *i.e.*,  $\mathbf{W}_k$  and  $\mathbf{X}$  are multiplied element by element,  $\mathbf{X} \in \mathbb{R}^{n \times d}$  is the original feature matrix, all tasks are the same. By the Eq. (1) we can make each feature of each node on each task get the corresponding weight. Then, we feed the weighted feature matrix  $\mathbf{X}_k^{(0)}$  into the feature sharing module to learn information from other tasks that are beneficial for its own training.

### 3.2. Feature sharing module

We employ an attention mechanism to automatically find the optimal information sharing ratio during the training process and every task can use the attention mechanism to absorb the most useful information from other tasks.

For subsequent attention to better share feature information, we feed  $\mathbf{X}_k^{(0)}$  and adjacency matrix  $\mathbf{A}$  into the graph convolution layer to learn both feature and structure information of the node [21,22]. The specific formula is as follows:

$$\mathbf{X}_k^{(\ell+1)} = \sigma(\hat{\mathbf{D}}^{-1/2} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-1/2} \mathbf{X}_k^{(\ell)} \Theta_k^{(\ell)}), \quad (2)$$

where  $\ell \in \{0, \dots, L-1\}$ ,  $\mathbf{X}_k^{(\ell+1)}$  is the output after graph convolution,  $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$ ,  $\hat{\mathbf{D}} = \text{diag}(\hat{d}_1, \dots, \hat{d}_n)$  which is a diagonal matrix and in which  $\hat{d}_i = \sum_{j=1}^n (\hat{\mathbf{A}}_{ij})$ ,  $\Theta_k^{(\ell)} \in \mathbb{R}^{d_\ell \times d_{\ell+1}}$  is a trainable parameter matrix of the  $\ell$ -th layer of the  $k$ th task, and  $\sigma(\cdot)$  represents an activation function. After the  $L$  layers graph convolution, we obtain the embedded feature matrix  $\mathbf{X}_k^{(L)}$ .

Then we design an attention unit to share information between tasks, as shown in Fig. 2. The input of the attention unit is the embedded feature  $\mathbf{X}_k^{(L)}$  after  $L$  graph convolution layers. The output of the attention unit is the shared representation  $\tilde{\mathbf{X}}_k^{(L)}$  which is a linear combination of embedded feature matrices for  $m$  tasks. The attention-based feature sharing process can be described by the following equation:

$$\begin{bmatrix} \tilde{x}_{ij,1}^{(tL)} \\ \tilde{x}_{ij,2}^{(tL)} \\ \vdots \\ \tilde{x}_{ij,m}^{(tL)} \end{bmatrix} = \begin{bmatrix} \alpha_{11}^{(t)} & \alpha_{12}^{(t)} & \dots & \alpha_{1m}^{(t)} \\ \alpha_{21}^{(t)} & \alpha_{22}^{(t)} & \dots & \alpha_{2m}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1}^{(t)} & \alpha_{m2}^{(t)} & \dots & \alpha_{mm}^{(t)} \end{bmatrix} \times \begin{bmatrix} x_{ij,1}^{(tL)} \\ x_{ij,2}^{(tL)} \\ \vdots \\ x_{ij,3}^{(tL)} \end{bmatrix}, \quad (3)$$

where  $x_{ij,k}^{(tL)}$  is the  $j$ th feature of the  $i$ th sample for the  $k$ th task after  $t$  times feature sharing through attention unit,  $t \in \{1, 2, \dots, T\}$ ,  $\alpha_{ij}^{(t)}$ ,  $i, j \in \{1, 2, \dots, m\}$  are trainable parameters, denoting the ratio of information that the  $i$ th task obtains from the  $j$ th task and  $\alpha_{ij}^{(t)}$  satisfy the following equation:

$$\sum_{j=1}^m \alpha_{ij}^{(t)} = 1, i \in \{1, 2, \dots, m\}. \quad (4)$$

The larger the value of  $\alpha_{ij}^{(t)}$  means the higher the degree of sharing between task  $i$ th and task  $j$ th. If  $\alpha_{ij}^{(t)} = 0$ , it means the  $i$ th task does not absorb information from the  $j$ th task.

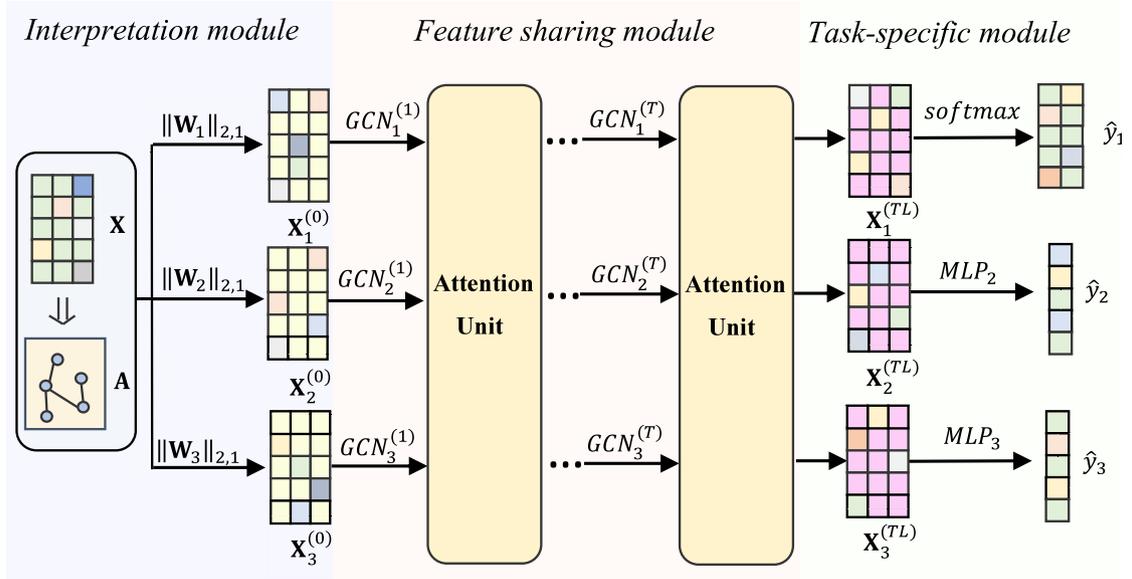


Fig. 1. The framework of our proposed AMTI-GCN. It consists of three modules: an interpretation module, a feature sharing module, and a task-specific module. The interpretation module aims to obtain sparse weight matrices (mask)  $W$  that indicate the importance of different features. The feature sharing module employs two attention units to enable feature exchange across different tasks. The task-specific module utilizes task-specific features to perform each task. These three modules are jointly trained in an end-to-end manner to achieve AD diagnosis results.

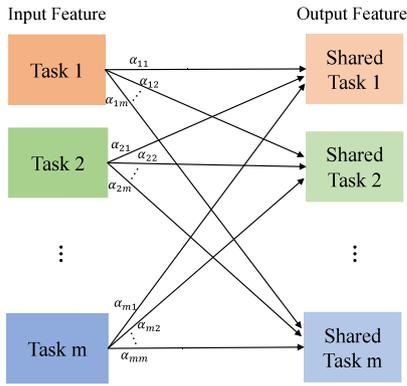


Fig. 2. Attention Unit: sharing feature information.

We place the attention unit after the graph convolution network. To be specific, after every feature extraction by graph convolution, we feed the embedded matrices into the attention unit to perform information interaction. After  $T \times L$  layers of graph convolution and  $T$  times of sharing, we feed the obtained features into the task-specific module.

### 3.3. Task specific module

After feature sharing module, we obtain the feature matrix  $\mathbf{X}^{(TL)} = [\mathbf{X}_0^{(TL)}, \mathbf{X}_1^{(TL)}, \dots, \mathbf{X}_m^{(TL)}]$  for  $m$  tasks.

If the  $k$ th task is the classification task, we input the embedded feature  $\mathbf{X}_k^{(TL)} \in \mathbb{R}^{n \times c}$  into a *softmax* layer to obtain the probability  $\hat{z}_{ij,k}$  that the  $i$ th sample belongs to the  $j$ th class. The specific expression is as follows:

$$\hat{z}_{ij,k} = \text{softmax}(\tilde{x}_{ij,k}^{(TL)}) = \frac{\exp(\tilde{x}_{ij,k}^{(TL)})}{\sum_{j=1}^c \exp(\tilde{x}_{ij,k}^{(TL)})} \quad (5)$$

For obtaining the corresponding probabilities of each sample's categories, we take the index of each sample's category with the highest probability as the prediction result  $\hat{y}_{i,k}$ . Then we can get the final prediction label  $\hat{y}_k$  of all samples.

$\mathbf{X}_k^{(TL)} \in \mathbb{R}^{n \times dT}$  for the task. Then, we input the matrix into the multi-layer perceptron (MLP) to output the prediction  $\hat{y}_k$ . The mathematical formula is expressed as follows:

$$\hat{y}_k = \text{MLP}_k(\tilde{\mathbf{X}}_k^{(TL)}), \quad (6)$$

### 3.4. Loss function

First, we design a loss function for each task  $k$  with the following mathematical expression:

$$L_k = L_{\text{task},k} + \beta_k L_{\text{sparse},k} \quad (7)$$

where  $L_{\text{task},k}$  denotes the task-specific loss function and  $L_{\text{sparse},k}$  denotes the constraint function used to sparse the features of the specific task.  $\beta_k$  is a tuning parameter to balance the magnitude of task loss and sparse learning.

For  $L_{\text{task},k}$ , if the  $k$ th task is a classification task, we use cross entropy to calculate its task-specific loss for cross entropy measures how well the model predicts the true labels of the data. The mathematical formula is expressed as follows:

$$L_{\text{classification},k} = - \sum_{i=1}^n y_{i,k} \ln \hat{y}_{i,k} \quad (8)$$

where  $y_{i,k}$  denotes the real label of the  $i$ th node and  $\hat{y}_{i,k}$  represents the prediction label of the  $i$ th node.

If the  $k$ th task is a regression task, we utilize mean square error which measures how well the model fits the data to obtain the task-specific loss of the  $k$ th task. The specific formula is expressed as follows:

$$L_{\text{regression},k} = \sum_{i=1}^n (y_{i,k} - \hat{y}_{i,k})^2 \quad (9)$$

where  $y_{i,k}$  denotes the real clinical score of the  $i$ th sample and  $\hat{y}_{i,k}$  represents the prediction score of the  $i$ th sample.

What is more, since we use the  $L_{2,1}$  norm to sparse the weight matrix for giving the insignificant features smaller weights which are even close to zero, the constraint function used to sparse the features of the specific task can be expressed as follows:

$$L_{\text{sparse},k} = \|\mathbf{W}_k\|_{2,1} = \sum_{i=1}^n \sqrt{\sum_{j=1}^d w_{ij,k}^2} \quad (10)$$

**Table 1**

The classification and regression performance of nine methods on AD-NC and AD-MCI datasets. We bold the best result in each setting.

Method	AD-NC							AD-MCI								
	ACC	SEN	SPE	AUC	ADAS-Cog		MMSE	ACC	SEN	SPE	AUC	ADAS-Cog		MMSE		
					CC	RMSE						CC	RMSE			
															CC	RMSE
RF	78.5	81.6	72.6	77.1	0.525	5.101	0.568	2.217	68.5	70.1	66.4	68.4	0.389	4.782	0.433	2.593
XGBoost	77.6	79.5	74.7	76.9	0.481	5.337	0.504	2.329	67.4	69.9	65.7	67.3	0.392	4.778	0.419	2.685
GCN	82.8	81.4	83.8	82.4	0.571	4.962	0.507	2.319	68.9	70.5	67.8	68.7	0.406	4.468	0.452	2.549
GAT	83.6	82.2	84.1	83.5	0.565	5.023	0.516	2.305	70.3	71.2	68.9	70.2	0.476	4.446	0.476	2.332
IDGCN	84.0	82.4	84.5	83.5	0.588	4.917	0.533	2.290	71.8	72.5	70.3	71.4	0.426	4.598	0.458	2.543
ST	84.8	82.7	86.7	83.2	0.596	4.857	0.541	2.276	71.2	73.7	67.9	70.8	0.440	4.532	0.466	2.468
HPS	84.6	83.8	85.3	84.5	0.628	4.709	0.623	2.009	72.5	73.5	70.3	72.1	0.471	4.460	0.471	2.382
CSN	85.3	83.6	86.1	84.7	0.637	4.641	0.665	1.926	73.0	75.3	71.3	73.2	0.495	4.242	0.512	1.926
AMTI-GCN	<b>86.2</b>	<b>85.4</b>	<b>87.0</b>	<b>85.9</b>	<b>0.649</b>	<b>4.513</b>	<b>0.681</b>	<b>1.889</b>	<b>74.2</b>	<b>76.6</b>	<b>72.4</b>	<b>73.8</b>	<b>0.512</b>	<b>4.015</b>	<b>0.547</b>	<b>1.886</b>

In summary, with  $L_{task,k}$  and  $L_{sparse,k}$ , we can calculate  $L_k$ . Because the significance of different tasks varies, we cannot directly add up their losses. Hence, the final loss  $L$  of our model AMTI-GCN can be expressed as a weighted sum of all task losses, i.e.,

$$L_{total} = \sum_{k=1}^3 \lambda_k L_k \quad (11)$$

According to Eqs. (7) and (11), there are six hyperparameters that need to be adjusted manually. Adjusting six hyperparameters is time-consuming. Therefore, we follow the idea of paper [23] by utilizing probability modeling to automatically learn the best weights for each task so as to reduce the time cost caused by hyperparameters. So we can simplify the loss function as follows:

$$L_{total} = \frac{1}{\sigma_1^2} L_1 + \frac{1}{2\sigma_2^2} L_2 + \frac{1}{2\sigma_3^2} L_3 \quad (12)$$

where  $\sigma_1$ ,  $\sigma_2$ , and  $\sigma_3$  are the noise scalars, respectively. They can be automatically adjusted during training. Therefore, we just need to manually adjust  $\beta_1, \beta_2, \beta_3$  in Eq. (7).

By minimizing Eq. (12), we can obtain the optimal parameters  $\Theta_k^t = \{\Theta_k^{t,0}, \Theta_k^{t,1}, \dots, \Theta_k^{t,L}\}$  of the graph convolution and the optimal parameters  $W_k$  of the weight matrix.

For clarity, the details of the AMTI-GCN are shown in Algorithm 1, where the EPOCHS means the number of training.

---

#### Algorithm 1 AMTI-GCN

---

**Input:**  $\mathbf{X} \in \mathbb{R}^{n \times d}$ , label information of the  $k$ -th task  $\mathbf{y}_k$  and hyperparameters  $\lambda_k$ , and  $\beta_k$ ,  $k \in \{1, 2, 3\}$ .

- 1: Initialization: GCN parameters  $\Theta_k^t$ , Weight Matrices  $W_k$ , attention unit parameters  $a_{ij}^t$ .
- 2: **while**  $epoch < EPOCHS$  **do**
- 3:  $\mathbf{X}_k^0 = W_k \odot \mathbf{X}$ ;
- 4: **for**  $t$  in range( $T$ ) **do**
- 5:  $\mathbf{X}_k^{(tL+1)} \leftarrow \{\mathbf{X}_k^{(tL)}, \mathbf{A}, \Theta_k^t\}$  by Eq. (2);  
 $\tilde{\mathbf{X}}_k^{(tL+1)} \leftarrow \mathbf{X}_k^{(tL+1)}$  by Eq. (3)
- 6: **end for**
- 7:  $\hat{\mathbf{y}}_k \leftarrow \tilde{\mathbf{X}}_k^{(TL)}$  by Eq. (5) and Eq. (6);
- 8:  $Loss \leftarrow \{\mathbf{y}_k, \hat{\mathbf{y}}_k\}$  by Eq. (11);
- 9: Back-propagate  $Loss$  to update model parameters ;
- 10: **end while**

**Output:**  $\hat{\mathbf{y}}_k, W_k$ .

---

## 4. Experiments

### 4.1. Experimental setup

**Datasets description.** For our experiments, we utilized raw digital images from the ADNI database and used 1.5T T1-weighted MRI data. Following the procedure in paper [24], we pre-processed the images by removing extraneous brain tissue, correcting for motion and time,

registering, filtering, and smoothing the images. Gray matter, white matter, and cerebrospinal fluid were the next three tissue types into which the images were segmented. We then warped them into the Jacob template [25] to obtain 93 brain regions. Finally, we obtained a 93-dimensional feature vector for each subject (patient) by extracting the gray matter volume of each region as a feature.

Our experiments involved 805 subjects, consisting of 186 AD patients, 393 Mild Cognitive Impairment (MCI) patients, and 226 normal controls (NC). Among the 393 MCI patients, 226 were MCI converters (MCIp) and 167 were MCI non-converters (MCI n). We also collected the scores of the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) and Mini Mental State Examination (MMSE), which are two commonly used clinical scores for AD diagnosis. We applied the following inclusion criteria for selecting the subjects: (1)NC subjects had MMSE scores ranging from 24 to 30 and Clinical Dementia Rating (CDR) of 0. (2)MCI subjects had MMSE scores ranging from 24 to 30 and CDR of 0.5. (3)AD subjects had MMSE scores ranging from 20 to 26 and CDR of either 0.5 or 1.0.

We then created four binary datasets by comparing different groups of subjects (i.e., AD-NC, AD-MCI, NC-MCI, and MCI n-MCIp). The sample proportion of each dataset is 186:226 for AD-NC, 186:393 for AD-MCI, 226:393 for NC-MCI, and 226:167 for MCI n-MCIp, respectively. We show the statistics information of samples in ADNI in Table 3.

**Graph construction.** Patients in the same class may share the same features. Therefore, we can use the connections between patients to construct the adjacency matrix to describe the relationships between samples following the paper [8]. The constructed adjacency matrix is  $\mathbf{A} \in \mathbb{R}^{n \times n}$ .

We use the reciprocal of the Euclidean distance to compute the similarity between any two distinct samples. For the  $i$ th and  $j$ th samples, we can calculate their distance as shown in the following equation:

$$a_{ij} = \frac{1}{\sqrt{\sum_{p=1}^d (x_{ip} - x_{jp})^2}} \quad (i \neq j), \quad (13)$$

where  $a_{ij}$  represents the similarity between the  $i$ th and  $j$ th samples.

In order to convert the matrix to the standard adjacency matrix form, we select the top  $K$  similar samples as neighbors for each sample. We set the weights of neighboring samples in the adjacency matrix to 1 and the weights of the remaining samples to 0, respectively. The mathematical expression is as follows:

$$a_{ij} = \begin{cases} 1 & \text{if } a_{ij} > \tilde{a}_{iK} \\ 0 & \text{otherwise,} \end{cases} \quad (14)$$

where  $\tilde{a}_{iK}$  is the  $K$ th largest similarity in the remaining  $d - 1$  samples for the  $i$ th sample. Then we convert matrix  $\mathbf{A}$  in symmetric form, by applying  $\mathbf{A} = (\mathbf{A} + \mathbf{A}^T)/2$ .

**Comparison methods.** We compare our proposed method with the following eight baseline methods: Random Forest (RF) [26], XGBoost [27], Graph Convolutional Network (GCN) [28], Graph Attention Network (GAT) [29], Interpretable Dynamic Graph Convolutional Networks

**Table 2**

The classification and regression performance of nine methods on NC-MCI and MCIIn-MCIP datasets. We bold the best result in each setting.

Method	NC-MCI							MCIIn-MCIP								
	ACC	SEN	SPE	AUC	ADAS-Cog		MMSE	ACC	SEN	SPE	AUC	ADAS-Cog		MMSE		
					CC	RMSE						CC	RMSE			
															CC	RMSE
RF	65.5	61.8	67.1	64.5	0.371	4.068	0.431	1.672	65.8	66.1	63.9	64.8	0.347	3.586	0.386	2.178
XGBoost	65.6	62.7	66.4	65.2	0.378	4.032	0.399	1.913	66.2	67.2	64.4	64.1	0.388	3.357	0.354	2.296
GCN	65.4	62.3	69.6	65.0	0.436	3.606	0.382	2.045	66.6	69.1	62.6	66.7	0.362	3.466	0.392	2.165
GAT	66.1	63.8	69.2	66.0	0.415	3.825	0.421	1.873	65.7	69.2	62.3	65.9	0.397	3.272	0.432	1.824
IDGCN	67.2	66.0	69.0	66.5	0.452	3.526	0.436	1.727	66.9	68.2	64.5	66.9	0.402	3.225	0.461	1.748
ST	69.2	66.8	<b>71.4</b>	68.5	0.443	3.587	0.438	1.638	70.7	71.6	69.5	70.0	0.419	3.049	0.486	1.626
HPS	67.9	65.7	70.6	67.3	0.456	3.407	0.472	1.582	69.6	70.7	68.4	70.0	0.446	3.112	0.458	1.776
CSN	68.6	67.8	70.5	68.7	0.469	3.363	0.489	1.489	71.2	<b>73.6</b>	69.5	71.7	0.458	3.016	0.493	1.518
AMTI-GCN	<b>70.1</b>	<b>69.3</b>	70.8	<b>70.6</b>	<b>0.477</b>	<b>3.264</b>	<b>0.498</b>	<b>1.400</b>	<b>71.9</b>	73.2	71.1	<b>72.5</b>	<b>0.485</b>	<b>2.872</b>	<b>0.522</b>	<b>1.415</b>

**Table 3**

Statistics information of samples in ADNI.

	AD	NC	MCIIn	MCIP
Number	186	226	226	167
Age	75.2 ± 7.4	75.3 ± 5.2	75.8 ± 6.8	74.8 ± 7.1
ADAS-Cog	18.3 ± 6.0	12.1 ± 3.8	12.9 ± 3.9	8.03 ± 3.8
MMSE	23.8 ± 2.0	29.0 ± 1.2	26.6 ± 1.7	28.4 ± 1.7

(IDGCN) [8], Single-task (ST) which use subnetwork of AMTI-GCN to perform single task without exchanging features across tasks, Hard Parameter Sharing (HPS) [30], Cross-stitch Networks (CSN) [31].

To evaluate the interpretability of AMTI-GCN, we compare it with three methods that are also interpretable: RF, IDGCN and ST.

**Experimental setting.** We used five-fold cross-validation and conducted the experiments 20 times for each method on the four datasets. We then reported the average results across the repetitions. We performed a total of three tasks in the experiments, namely patient classification, ADAS-Cog regression, and MMSE regression, and set  $m = 3$  as the number of tasks. We adopted different evaluation metrics for different tasks. For classification, we used accuracy (ACC), specificity (SPE), sensitivity (SEN), and AUC score (AUC). For regression, we used correlation coefficient (CC) and root mean squared error (RMSE). Moreover, for AD diagnosis, we also demonstrated the interpretability of AMTI-GCN by identifying the most significant features (brain regions) for the prediction.

**Implemental details.** We followed the paper recommendations to tune the hyper-parameters for each method and obtain their optimal results. For the proposed AMTI-GCN, we set the maximum number of epochs to 2000, the learning rate to 0.002, the number of graph convolutional layers to 2, and the times of feature sharing to 2, i.e.,  $L = 2$  in Eq. (2) and  $T = 2$  in Eq. (3).  $\beta_k \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$  in Eq. (7). We also list optimal  $\beta_k$  on four datasets, as shown in Table 6 in the appendix. We conducted the experiments on 8×NVIDIA RTX 3090.

## 4.2. Results and analysis

### 4.2.1. Classification results

We compare the classification performance of all methods on four datasets, as shown in Tables 1 and 2. The best results for each metric and dataset are in bold. The results show that: (1) Deep learning outperforms traditional machine learning. (2) Multi-task models often yield better results than analyzing a single task alone. (3) A soft parameter sharing model can obtain better results than a hard parameter sharing model. (4) Our proposed method, AMTI-GCN, almost outperformed the other eight methods on the four datasets in terms of ACC, SEN, SPE, and AUC. Specifically, AMTI-GCN outperformed the second highest method by 0.9%, 1.8%, 0.9%, and 1.2% in terms of ACC, SEN, SPE, and AUC, respectively.

### 4.2.2. Regression results

As we can see, the prediction of MMSE and ADAS-Cog clinical scores follows the same pattern as the classification results. The possible reasons for the above observations are as follows: (1) Deep learning models have an advantage over traditional machine learning in feature extraction. (2) Most deep learning methods do not consider the relationship between tasks and thus miss useful information from relevant tasks, which will lead to suboptimal results; (3) A hard parameter sharing mechanism can only use the underlying shared information indiscriminately, which lacks flexibility. Conversely, the soft parameter sharing mechanism can adjust the information sharing ratio between tasks. Hence, our method outperforms IDGCN and HPS. (4) Previous approaches either failed to effectively use relevant information between tasks or considered that different features are equally important to the task. By contrast, AMTI-GCN can automatically and flexibly adjust the ratio of information sharing among tasks based on the attention mechanism and, meanwhile, select significant features for tasks by weight matrices.

### 4.2.3. Interpretability

We evaluate the feature interpretability of the proposed AMTI-GCN by comparing it with three methods with interpretation ability: RF, IDGCN, and ST. We apply the four methods to the four datasets using five-fold cross-validation 20 times, resulting in 100 feature selection runs for each method. In each run, we record the 10 most significant features according to the method. Then, we count the frequency of occurrence of each significant feature across the 100 runs and select the 10 most frequent features as the representative features of the method. We show the indexes and importance ranks of the top 10 features for each method on each dataset in Table 4. We also visualize the brain regions selected by Random Forests, IDGCN, ST, and AMTI-GCN on the AD-NC dataset, as shown in Fig. 4 in Appendix.

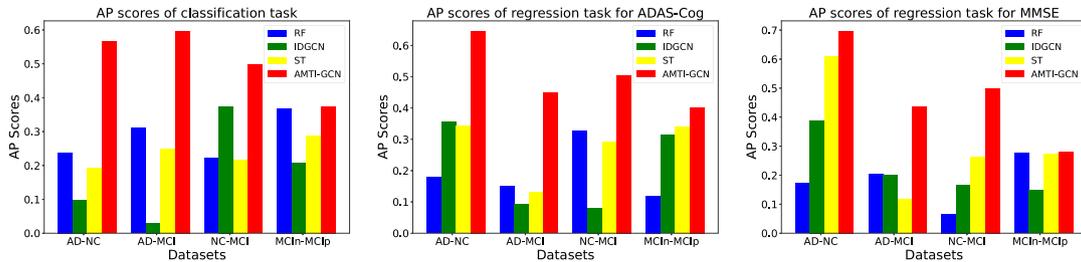
According to the paper [32], the hippocampal formation (30,69), amygdala (76,83), precentral gyrus (5,55), and parahippocampal gyrus (17,78) are the four brain areas most frequently linked to AD. We note that all four methods can choose all or most of these regions as significant features. However, the rankings of the regions are different according to the four methods. We utilize the average precision (AP) [33] as a metric to evaluate the ranking performance of the feature interpretability.

We consider hippocampal formation (30, 69), amygdala (76, 83), precentral gyrus (5, 55), and parahippocampal gyrus (17, 78) as important features. We compute the AP for each sequence in Fig. 3, and report the results of the four methods along with their sequences. We calculate the AP score for one classification metric and two regression metrics. We observe that AMTI-GCN has a higher AP score than the other three methods in both classification and regression tasks. We also notice that AMTI-GCN can select the most important features (hippocampal formation (30, 69) and amygdala (76, 83)). These outcomes demonstrate the potency of our approach, which outperforms the competition in terms of ranking performance.

**Table 4**

The indexes of top 10 important brain regions on AD-NC, AD-MCI, NC-MCI and MCIIn-MCIp. The bold numbers represent the number of the most relevant brain regions for AD diagnosis, corresponding to four different methods.

Method	AD-NC	AD-MCI	NC-MCI	MCIIn-MCIp
Classification				
RF	30,27,3,11, <b>83,69,8,52,84,80</b>	8,30, <b>83,5,11,85,69,80,47,4</b>	69,3,8,32,76,85,11, <b>83,92,46</b>	5,36,33, <b>83,1,30,46,59,55,17</b>
IDGCN	21,27,32,49, <b>69,60,30,58,41,83</b>	24,70,52, <b>69,84,20,22,23,12,11</b>	5,36,33, <b>83,1,30,46,55,39,17</b>	21, <b>69,30,22,64,78,62,70,52,48</b>
ST	30,64,35,41,21,75,91,17,63,5	28,5,89, <b>69,30,43,64,51,54,76</b>	40,10,5,25,17, <b>83,87,55,84,27</b>	69,5,75,87,11,10,63,2,62, <b>83</b>
AMTI-GCN	30,21, <b>69,83,78,82,17,64,62,5</b>	30, <b>78,22,83,5,53,33,17,38,69</b>	30,17, <b>78,55,67,76,74,47,5,70</b>	83, <b>30,74,49,17,17,65,48,61,2,69</b>
ADAS-Cog				
RF	27,46, <b>83,30,76,84,80,48,62,3</b>	83,27,20,24,8,1,84,36,44, <b>30</b>	27, <b>83,30,92,21,76,3,81,78,55</b>	58,36,22, <b>30,38,69,20,17,47,49</b>
IDGCN	69,47,76,21, <b>30,82,55,72,57,8</b>	68,23,5,79, <b>30,56,38,53,70,84</b>	9,23,21,5,78,63,50,82,67,46	63,21, <b>69,17,5,33,30,65,52,78</b>
ST	69,59,55,86,68, <b>30,76,62,18,91</b>	84,14,17,48,11,55,74, <b>30,28,4</b>	40,55,26,39, <b>69,17,48,50,83,76</b>	43,56, <b>83,5,76,69,48,76,79,84</b>
AMTI-GCN	69,41,76,27, <b>83,30,55,78,17,82</b>	28,55,60,76, <b>30,17,78,26,38,5</b>	83,5,41,69,47,55,44,78,22,63	30,74, <b>83,55,17,63,47,48,60,82</b>
MMSE				
RF	27, <b>83,46,30,84,80,62,76,89,3</b>	83,8,11,36,67,5,15,46,39, <b>69</b>	27,87,82,30,48,64,5,68,49,32	36, <b>83,75,47,17,57,21,30,55,78</b>
IDGCN	69,17,27,82,5,81,60,30,75,51	68,79,55,23,1,17, <b>69,30,82,4</b>	9,78,92,55,63,93,82,77,17,14	63,80,61,20,92,5,76,33,17,78
ST	69,30,19,78,47, <b>83,76,55,9,20</b>	84,65,49,17,78,93,50,33,80,76	15,40, <b>30,37,55,38,17,20,83,76</b>	83,60,90,82,67,78,55,77,30,74
AMTI-GCN	69,27,76, <b>83,78,5,41,30,17,47</b>	30,28,23,5,17,76, <b>83,29,68,92</b>	30,41,55,63,76,93,69,10,78,17	74, <b>30,76,65,26,78,83,47,21,87</b>

**Fig. 3.** AP scores of RF, IDGCN, ST, AMTI-GCN on four datasets.**Table 5**

Ablation study results on four datasets. We bold the best result in each dataset.

Module		AD-NC		AD-MCI		NC-MCI		MCIIn-MCIp					
Interpretation	Sharing	ACC	RMSE	ACC	RMSE	ACC	RMSE	ACC	RMSE				
		ADAS-Cog	MMSE	ADAS-Cog	MMSE	ADAS-Cog	MMSE	ADAS-Cog	MMSE				
✗	✗	82.9	5.019	2.586	68.5	4.762	2.649	67.6	3.626	2.610	68.8	3.157	2.648
✓	✗	84.8	4.857	2.276	71.2	4.532	2.468	68.2	3.587	1.638	70.7	3.094	1.626
✗	✓	85.1	4.786	2.059	72.1	4.287	2.146	69.6	3.475	1.602	70.0	2.978	1.582
✓	✓	<b>86.2</b>	<b>4.513</b>	<b>1.889</b>	<b>74.2</b>	<b>4.015</b>	<b>1.886</b>	<b>70.1</b>	<b>3.264</b>	<b>1.400</b>	<b>71.9</b>	<b>2.872</b>	<b>1.415</b>

### 4.3. Ablation study

To demonstrate the effectiveness of the interpretation module and the feature sharing module, we conducted the following ablation experiments. We take the single-task subnet of AMTI-GCN without the interpretation module as the baseline, and add the interpretation module and the feature sharing module on the baseline to prove that the two modules are effective respectively. Finally, the results are then compared with AMTI-GCN. We experimented on the four previous datasets. We use classification accuracy (ACC) and root mean square error (RMSE) as classification and regression metrics, respectively. We show the results in Table 5. In the first two columns, *Interpretation* represents the interpretation module, and *Sharing* represents the feature sharing module.

We can see from Table 5 that the performance of the model improves in both classification and regression tasks after adding the interpretation module to the baseline, which indicates that the interpretation module can reduce the redundant features and thus improve the performance of classification and regression. Meanwhile, the classification and regression performance can be improved by adding the

feature sharing module to the baseline. This indicates that the feature sharing module can exploit the correlation between tasks to improve the performance of each task. Finally, the performance of the model is further improved by adding two modules at the same time, indicating that the two modules can interact with each other to improve the overall performance of the model on each task.

## 5. Conclusion

The proposed AMTI-GCN shows superior performance and provides interpretability in both classification and regression tasks, as evidenced by the experimental results and the comparison with existing methods. In our work, we utilized a multi-task model with MRI as an input to predict the result of MRI, ADAS-Cog, and MMSE scores. However, we did not thoroughly investigate and analyze potential relationships between ADAS-Cog and MMSE scores and other factors, e.g., ADAS-Cog and MMSE cognitive scores may differ according to the level of education. In our future research, we plan to expand our analysis by considering additional factors, including education levels and mental health status, when predicting these two indicators.

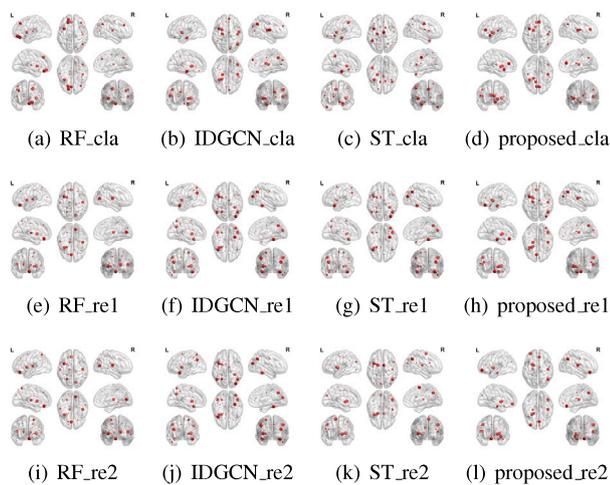


Fig. 4. Top 10 brain regions selected on the AD-NC dataset by RF, IDGCN, ST and AMTI-GCN on three tasks.

1	medial front-orbital gyrus right	47	middle occipital gyrus right
2	middle frontal gyrus right	48	middle temporal gyrus left
3	lateral ventricle left	49	lingual gyrus left
4	insula right	50	superior frontal gyrus left
5	precentral gyrus right	51	nucleus accumbens left
6	lateral front-orbital gyrus right	52	occipital lobe WM left
7	cingulate region right	53	postcentral gyrus left
8	lateral ventricle right	54	inferior frontal gyrus right
9	medial frontal gyrus left	55	precentral gyrus left
10	superior frontal gyrus right	56	temporal lobe WM left
11	globus pallidus right	57	medial front-orbital gyrus left
12	globus pallidus left	58	perirhinal cortex right
13	putamen left	59	superior parietal lobule right
14	inferior frontal gyrus left	60	lateral front-orbital gyrus left
15	putamen right	61	perirhinal cortex left
16	frontal lobe WM right	62	inferior temporal gyrus left
17	parahippocampal gyrus left	63	temporal pole left
18	angular gyrus right	64	entorhinal cortex left
19	temporal pole right	65	inferior occipital gyrus right
20	subthalamic nucleus right	66	superior occipital gyrus left
21	nucleus accumbens right	67	lateral occipitotemporal gyrus right
22	uncus right	68	entorhinal cortex right
23	cingulate region left	69	hippocampal formation left
24	fornix left	70	thalamus left
25	frontal lobe WM left	71	parietal lobe WM right
26	precuneus right	72	insula left
27	subthalamic nucleus left	73	postcentral gyrus right
28	posterior limb of internal capsule left	74	lingual gyrus right
29	posterior limb of internal capsule right	75	medial frontal gyrus right
30	hippocampal formation right	76	amygdala left
31	inferior occipital gyrus left	77	medial occipitotemporal gyrus left
32	superior occipital gyrus right	78	parahippocampal gyrus right
33	caudate nucleus left	79	anterior limb of internal capsule right
34	supramarginal gyrus left	80	middle temporal gyrus right
35	anterior limb of internal capsule left	81	occipital pole right
36	occipital lobe WM right	82	corpus callosum
37	middle frontal gyrus left	83	amygdala right
38	superior parietal lobule left	84	inferior temporal gyrus right
39	caudate nucleus right	85	superior temporal gyrus right
40	cuneus left	86	middle occipital gyrus left
41	precuneus left	87	angular gyrus left
42	parietal lobe WM left	88	medial occipitotemporal gyrus right
43	temporal lobe WM right	89	cuneus right
44	supramarginal gyrus right	90	lateral occipitotemporal gyrus left
45	superior temporal gyrus left	91	thalamus right
46	uncus left	92	occipital pole left
		93	fornix right

Fig. A.1. The names of the selected brain regions in this work.

Table 6

Optimal hyperparameters on four binary datasets.

Datasets	$\beta_1$	$\beta_2$	$\beta_3$	Datasets	$\beta_1$	$\beta_2$	$\beta_3$
AD-NC	$10^{-1}$	$10^{-5}$	$10^{-4}$	AD-MCI	$10^{-1}$	$10^{-5}$	$10^{-2}$
NC-MCI	$10^{-2}$	$10^{-4}$	$10^{-3}$	MCIin-MCIp	$10^{-2}$	$10^{-4}$	$10^{-4}$

## CRedit authorship contribution statement

**Shunqin Jiang:** Writing – review & editing, Writing – original draft, Methodology. **Qiyuan Feng:** Software, Methodology, Funding acquisition. **Hengxin Li:** Writing – review & editing, Software, Investigation. **Zhenyun Deng:** Writing – review & editing, Supervision. **Qinghong Jiang:** Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data that has been used is confidential.

## Appendix

See Table 6 and Figs. 4 and A.1.

## References

- [1] Burcu Akpınar Söylemez, Özlem Küçükçüçlü, Merve Aliye Akyol, Ahmet Turan Işık, Quality of life and factors affecting it in patients with alzheimer's disease: a cross-sectional study, *Health Qual. Life Outcomes* 18 (1) (2020) 1–7.
- [2] Jiangzhang Gan, Ziwen Peng, Xiaofeng Zhu, Rongyao Hu, Junbo Ma, Guorong Wu, Brain functional connectivity analysis based on multi-graph fusion, *Med. Image Anal.* 71 (2021) 102057.
- [3] Gill Livingston, Jonathan Huntley, Andrew Sommerlad, David Ames, Clive Ballard, Sube Banerjee, Carol Brayne, Alistair Burns, Jiska Cohen-Mansfield, Claudia Cooper, et al., Dementia prevention, intervention, and care: 2020 report of the lancet commission, *Lancet* 396 (10248) (2020) 413–446.
- [4] Marco Colizzi, Antonio Lasalvia, Mirella Ruggeri, Prevention and early intervention in youth mental health: is it time for a multidisciplinary and trans-diagnostic model for care? *Int. J. Mental Health Syst.* 14 (1) (2020) 1–14.
- [5] Emma Thomas-Jones, Amy Lloyd, Damian Roland, Gerri Sefton, Lyvonne Tume, Kerry Hood, Chao Huang, Dawn Edwards, Alison Oliver, Richard Skone, et al., A prospective, mixed-methods, before and after study to identify the evidence base for the core components of an effective paediatric early warning system and the development of an implementation package containing those core recommendations for use in the UK: Paediatric early warning system–utilisation and mortality avoidance—the PUMA study protocol, *BMC Pediatr.* 18 (2018) 1–13.
- [6] Ammarah Farooq, SyedMuhammad Anwar, Muhammad Awais, Saad Rehman, A deep CNN based multi-class classification of alzheimer's disease using MRI, in: *IST*, 2020, pp. 1–6.
- [7] Zhao Fan, Fanyu Xu, Xuedan Qi, Cai Li, Lili Yao, Classification of alzheimer's disease based on brain MRI and machine learning, *Neural Comput. Appl.* 32 (2020) 1927–1936.
- [8] Yonghua Zhu, Junbo Ma, Changan Yuan, Xiaofeng Zhu, Interpretable learning based dynamic graph convolutional networks for alzheimer's disease analysis, *Inf. Fusion* 77 (2022) 53–61.
- [9] Dong Jin Park, Min Woo Park, Homin Lee, Young-Jin Kim, Yeongsik Kim, Young Hoon Park, Development of machine learning model for diagnostic disease prediction based on laboratory tests, *Sci. Rep.* 11 (1) (2021) 7567.
- [10] Linda K. McEvoy, Christine Fennema-Notestine, J. Cooper Roddey, Donald J. Hagler Jr., Dominic Holland, David S. Karow, Christopher J. Pung, James B. Brewer, Anders M. Dale, Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment, *Radiology* 251 (1) (2009) 195–205.
- [11] Chong-Yaw Wee, Pew-Thian Yap, Kevin Denny, Jeffrey N. Browndyke, Guy G. Potter, Kathleen A. Welsh-Bohmer, Lihong Wang, Dinggang Shen, Resting-state multi-spectrum functional connectivity networks for identification of MCI patients, *PLoS One* 7 (5) (2020) e37828.

- [12] Stephen Z. Levine, Kazufumi Yoshida, Yair Goldberg, Myrto Samara, Andrea Cipriani, Orestis Efthimiou, Takeshi Iwatsubo, Stefan Leucht, Toshi A. Furukawa, Linking the mini-mental state examination, the alzheimer's disease assessment scale–cognitive subscale and the severe impairment battery: evidence from individual participant data from five randomised clinical trials of donepezil, *BMJ Ment. Health* 24 (2) (2021) 56–61.
- [13] Sofia Lahrichi, Maryem Rhanoui, Mounia Mikram, Bouchra El Asri, Toward a multimodal multitask model for neurodegenerative diseases diagnosis and progression prediction, in: *Proceedings of the 10th International Conference on Data Science, Technology and Applications*, 2021, pp. 322–328.
- [14] Kwansook Oh, Jee Seok Yoon, Heung-II Suk, Learn-explain-reinforce: counterfactual reasoning and its guidance to reinforce an alzheimer's disease diagnosis model, *IEEE Trans. Pattern Anal. Mach. Intell.* 45 (4) (2022) 4843–4857.
- [15] Louise Bloch, Christoph M Friedrich, Alzheimer's Disease Neuroimaging Initiative, Machine learning workflow to explain black-box models for early alzheimer's disease classification evaluated for multiple datasets, *SN Comput. Sci.* 3 (6) (2022) 509.
- [16] Rongyao Hu, Jiangzhang Gan, Xiaofeng Zhu, Tong Liu, Xiaoshuang Shi, Multi-task multi-modality SVM for early COVID-19 diagnosis using chest CT data, *Inf. Process. Manage.* 59 (1) (2022) 102782.
- [17] Mingxia Liu, Jun Zhang, Ehsan Adeli, Dinggang Shen, Joint classification and regression via deep multi-task multi-channel learning for alzheimer's disease diagnosis, *IEEE Trans. Biomed. Eng.* 66 (5) (2018) 1195–1206.
- [18] Wei Liang, Kai Zhang, Peng Cao, Xiaoli Liu, Jinzhu Yang, Osmar Zaiane, Rethinking modeling alzheimer's disease progression from a multi-task learning perspective with deep recurrent neural network, *Comput. Biol. Med.* 138 (2021) 104935.
- [19] Rongyao Hu, Debo Cheng, Wei He, Guoqiu Wen, Yonghua Zhu, Jilian Zhang, Shichao Zhang, Low-rank feature selection for multi-view regression, *Multimedia Tools Appl.* 76 (2017) 17479–17495.
- [20] Rongyao Hu, Ziwen Peng, Xiaofeng Zhu, Jiangzhang Gan, Yonghua Zhu, Junbo Ma, Guorong Wu, Multi-band brain network analysis for functional neuroimaging biomarker identification, *IEEE Trans. Med. Imaging* 40 (12) (2021) 3843–3855.
- [21] Jiangzhang Gan, Rongyao Hu, Yujie Mo, Zhao Kang, Liang Peng, Yonghua Zhu, Xiaofeng Zhu, Multigraph fusion for dynamic graph convolutional network, *IEEE Trans. Neural Netw. Learn. Syst.* (2022).
- [22] Liang Peng, Yujie Mo, Jie Xu, Jialie Shen, Xiaoshuang Shi, Xiaoxiao Li, Heng Tao Shen, Xiaofeng Zhu, GRLC: Graph representation learning with constraints, *IEEE Trans. Neural Netw. Learn. Syst.* (2023).
- [23] Alex Kendall, Yarin Gal, Roberto Cipolla, Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022, pp. 7482–7491.
- [24] Heng Tao Shen, Xiaofeng Zhu, Zheng Zhang, Shui-Hua Wang, Yi Chen, Xing Xu, Jie Shao, Heterogeneous data fusion for predicting mild cognitive impairment conversion, *Inf. Fusion* 66 (2021) 54–63.
- [25] N.J. Kabani, David MacDonald, Colin J. Holmes, Alan C. Evans, 3D anatomical atlas of the human brain, *NeuroImage* 7 (2017).
- [26] Vladimir Svetnik, Andy Liaw, Christopher Tong, J. Christopher Culberson, Robert P. Sheridan, Bradley P. Feuston, Random forest: a classification and regression tool for compound classification and QSAR modeling, *J. Chem. Inf. Comput. Sci.* 43 (6) (2003) 1947–1958.
- [27] Tianqi Chen, Carlos Guestrin, Xgboost: A scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, ACM, New York, NY, USA, 2016, pp. 785–794, <http://dx.doi.org/10.1145/2939672.2939785>, URL <http://doi.acm.org/10.1145/2939672.2939785>.
- [28] Thomas N. Kipf, Max Welling, Semi-supervised classification with graph convolutional networks, in: *J. International Conference on Learning Representations*, ICLR 2017, 2020.
- [29] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, Graph attention networks, in: *Computational Social Networks - 4th International Conference, CSoNet 2015, Beijing, China, August 4-6, 2015, Proceedings*, Vol. 9197, 2022, pp. 296–306.
- [30] Lijun Zhang, Qizheng Yang, Xiao Liu, Hui Guan, Rethinking hard-parameter sharing in multi-domain learning, in: *2022 IEEE International Conference on Multimedia and Expo, ICME, IEEE, 2022*, pp. 01–06.
- [31] Ishan Misra, Abhinav Shrivastava, Abhinav Gupta, Martial Hebert, Cross-stitch networks for multi-task learning, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022, pp. 3994–4003.
- [32] M. Flint Beal, Michael F. Mazurek, Vinh T. Tran, Geetinder Chattha, Edward D. Bird, Joseph B. Martin, Reduced numbers of somatostatin receptors in the cerebral cortex in alzheimer's disease, *Science* 229 (4710) (2017) 289–291.
- [33] Alejandro Bellogín, Pablo Castells, Iván Cantador, Statistical biases in information retrieval metrics for recommender systems, *Inf. Retrieval* 20 (6) (2017) 606–634.