Graph-Based Fusion of Imaging, Genetic and Clinical Data for Degenerative Disease Diagnosis

Guo, Rui ; Tian, Xu; Lin, Hanhe; McKenna, Stephen; Li, Hong-Dong; Guo, Fei

DOI:
10.1109/TCBB.2023.3335369

Publication date:
2023

Document Version
Peer reviewed version

Author Accepted Manuscript: Guo, R, Tian, X, Lin, H, McKenna, S, Li, H-D, Guo, F & Liu, J 2023, 'Graph-Based Fusion of Imaging, Genetic and Clinical Data for Degenerative Disease Diagnosis', IEEE/ACM Transactions on Computational Biology and Bioinformatics . https://doi.org/10.1109/TCBB.2023.3335369

Graph-based fusion of imaging, genetic and clinical data for degenerative disease diagnosis

Rui Guo[1], Xu Tian[1], Hanhe Lin, Stephen McKenna, Hong-Dong Li, Fei Guo, Jin Liu*, *Member, IEEE*

***Abstract*—Graph learning methods have achieved noteworthy performance in disease diagnosis due to their ability to represent unstructured information such as inter-subject relationships. While it has been shown that imaging, genetic and clinical data are crucial for degenerative disease diagnosis, existing methods rarely consider how best to use their relationships. How best to utilize information from imaging, genetic and clinical data remains a challenging problem. This study proposes a novel graph-based fusion (GBF) approach to meet this challenge. To extract effective imaging-genetic features, we propose an imaging-genetic fusion module which uses an attention mechanism to obtain modality-specific and joint representations within and between imaging and genetic data. Then, considering the effectiveness of clinical information for diagnosing degenerative diseases, we propose a multi-graph fusion module to further fuse imaging-genetic and clinical features, which adopts a learnable graph construction strategy and a graph ensemble method. Experimental results on two benchmarks for degenerative disease diagnosis (Alzheimer's Disease Neuroimaging Initiative and Parkinson's Progression Markers Initiative) demonstrate its effectiveness compared to state-of-the-art graph-based methods. Our findings should help guide further development of graph-based models for dealing with imaging, genetic and clinical data.**

***Index Terms*—Graph learning, Imaging and genetic data, Clinical data, Degenerative disease**

I. Introduction

ADVANCES in technologies have facilitated the acquisition of medical data in different modalities. Medical data acquired from multiple modalities can provide more useful information than those in a single modality. Moreover, they present an opportunity to extract more robust representations by modelling relationships between different modalities [1]–[3]. Although recent studies found that imaging, genetic and clinical data are crucial for degenerative disease diagnosis [4], [5], how to make use of these data for effective diagnosis is still a long-standing research topic.

A number of methods have been proposed to fuse imaging and genetic data in recent years. According to the fusion strategy, these methods can be categorized as data fusion methods [6]–[8], decision fusion methods [9], [10], and representation fusion methods [11], [12]. Data fusion methods generate a new form of data from the imaging and genetic data. Due to data heterogeneity, it might be difficult to achieve promising performance with such methods. Decision fusion methods make a final diagnosis by fusing diagnostic results from two separate models, where one model is trained on imaging data, and the other is trained on genetic data. Such methods ignore complementary information between imaging and genetic data when making their modality-specific diagnostic predictions. In contrast, representation fusion methods are capable of exploiting the relationships between modalities; these relationships have been modelled using generative adversarial networks [13], [14], knowledge distillation [15], hypergraphs [4], [16], and attention mechanisims [1]. Despite the advantages of representation fusion methods, existing methods do not optimally take into account both joint imaging-genetic information and modality-specific information.

With the ability to extract information from a wide range of unstructured data [17]–[19], graph learning has been successfully applied for imaging-based disease diagnosis. Parisot et al. [20] are the first to propose a framework that represented the population of subjects as a graph. This framework exploited graph convolutional networks (GCN) and involved representing populations as a sparse graph in which nodes were associated with imaging-based features, and clinical information was integrated as edge weights. With the help of a graph made up of subjects, the features of subjects similar to a given subject could be easily aggregated. This allows for a smoother high-dimensional manifold over subjects and enables the model to make consistent judgements across similar subjects. To integrate clinical information with imaging data, a series of methods have been proposed for constructing the subject graph. Huang et al. [21] distinguished image features and non-image features and used them as node features and edge features of the graph for data fusion. Tong et al. [22] proposed to construct a single graph for each modality based on a handcrafted kernel and then combined them into a unified graph for diagnosis. These methods simply combine graphs that have been constructed from different modalities without taking into account relationships between those different graphs.

To address the above shortcomings, we propose a novel graph-based fusion (GBF) method for degenerative disease diagnosis. It consists of an imaging-genetic fusion module and a multi-graph fusion module. The former uses an attention mechanism to extract both specific and joint information from imaging and genetic data. The latter applies multi-graph learning to fuse extracted image-genetic features with clinical data, taking account of both inter- and intra-graph relationships.

Our main contributions are summarized as follows.

1) Taking account of the characteristics of imaging, genetic and clinical data, we propose a novel graph-based fusion

Rui Guo, Xu Tian, Hong-Dong Li, Fei Guo and Jin Liu are with School of Computer Science and Engineering, Central South University, Changsha, 410083, China.

Hanhe Lin and Stephen McKenna are with the School of Science and Engineering, University of Dundee, DD1 4HN Dundee, United Kingdom .

*Corresponding author: Jin Liu (liujin06@csu.edu.cn)

[1]These authors contributed equally to this work.

approach for diagnosing degenerative disease.
2) We develop an imaging-genetic fusion module to exploit the relationship between imaging and genetic data, where their specific and joint information are extracted by an attention mechanism.
3) We develop a multi-graph fusion model to further fuse imaging, genetic and clinical features. Specifically, each type of data constructs a separate graph, whose inter-graph and intra-graph information are extracted by joint graph convolution and specific graph convolution, respectively.
4) We apply the proposed GBF method to benchmark datasets for two degenerative disease diagnosis tasks: mild cognitive impairment (MCI) and Parkinson's disease (PD). We present an ablation study as well as direct comparisons with state-of-the-art graph-based methods.

II. Related work

A. Multi-modal fusion

In recent years, an increasing number of methods have been proposed to improve the understanding of brain diseases utilizing imaging and genetic data [23], [24]. Cheng et al. [25] proposed a multiparametric magnetic resonance imaging (MRI) based approach for glioma grade prediction. In this study, multi-modal feature selection was used to exploit the complementarity of data in different modalities. Pahuja et al. [26] fused manual features acquired from MRI with features from cerebrospinal fluid (CSF) for PD diagnosis. However, since it was not an end-to-end approach, the representation of different modalities could not be optimized during training. To address this issue, Lu et al. [27] proposed a multi-modal joint learning approach for the diagnosis of Alzheimer's disease (AD), where the representations of different modalities were optimized. Considering that genetics is one of the important causes of degenerative diseases, Ying et al. [28] introduced a multi-modal AD diagnosis neural network that uses both imaging and genetic data. To further exploit the consistency between imaging and genetic data, Ko et al. [13] proposed an MCI diagnosis model. By using generative adversarial structures, this approach was more concerned with inter-modal consistency, and this consistent performance makes the multi-modal representations more robust. To detect potential biomarkers associated with AD progression, Huang et al. [29] proposed a new time-group sparsity regression and additive model using longitudinal imaging data and genetic data. Kim et al. [30] proposed a new connectivity-based penalty and combined it with lasso to investigate the association between multi-modal imaging data and genetic data.

Apart from imaging and genetic data, clinical variables such as age and sex play an important part in degenerative disease diagnosis. Zhang et al. [4] used a hypergraph to model the relationship between different modalities including clinical data. However, using a hypergraph is difficult to finely describe the specific associations between genes and brain regions. Shi et al. [31] proposed a new feature selection method which learns an adaptive similarity matrix between modalities and performs feature selection by group sparsity-induced $l_{2,1}$ parameterization. Zheng et al. [1] proposed a multi-modal fusion method to fuse imaging data, demographic data, and cognitive test data. This method constructs pairwise associations between multi-modal features through a self-attention mechanism, and feature interactions between the modalities are performed through such associations. Although these studies have initially attempted multi-modal fusion, these approaches paid little attention to the complex relationships between features in different modalities.

B. Graph learning

To exploit the relationship between subjects, graph learning has been widely used with biomedical data. Parisot et al. [20] proposed a graph-based framework for the diagnosis of autism spectrum disorder (ASD) and AD. They associated nodes in the graph with image-based features, while integrating demographic information such as sex and age as the weights of the edges. Lei et al. [32] designed a multi-scale enhancement-based graph convolution network for MCI diagnosis. It constructs a single graph to combine clinical data with structural and functional features of the brain obtained from multi-modal imaging data. However, it remains a challenge to balance the information from both imaging and non-imaging data within one graph.

To address this limitation, multi-graph learning [10], [11], [33] has been proposed, where each graph corresponds to one specific type of data. Liu et al. [10] proposed to construct separate graphs for different features from MRI and single nucleotide polymorphisms (SNP) by clinical data and to fuse the predictions from these graphs. Since graph convolution is sensitive to constructed graphs, there is increasing attention on graph structure learning. An edge-variable method was proposed in [21]. In this method, pairwise association encoders based on learnable functions were used for edge construction by non-imaged features. Zhang et al. [33] proposed a local-to-global graph convolution network for brain disease diagnosis. Their proposed method combines imaging and clinical data through a brain-based network and a subject-based network of graph convolutions. To balance imaging and non-imaging data, Song et al. [11] proposed a multi-centre attention graph consisting of subjects. Considering information related to subjects' diagnosis such as site, gender, and collection device, the method uses an attention mechanism to capture their effects on edge weights and performs graph fusion based on attention weights. In addition, Zheng et al. [1] proposed further smoothing and sparse terms to regularize the graph structure. The above methods have attempted to extract multi-graph information from biomedical data. However, they do not take into account joint information across multi-graphs and specific information within each graph, which has a negative impact on model performance due to potential information loss.

III. Method

The framework for our proposed graph-based fusion (GBF) approach is shown in Fig. 1. It consists of two modules, i.e., imaging-genetic fusion (IGF) and multi-graph fusion (MGF). For imaging and genetic features, IGF uses multi-subspace

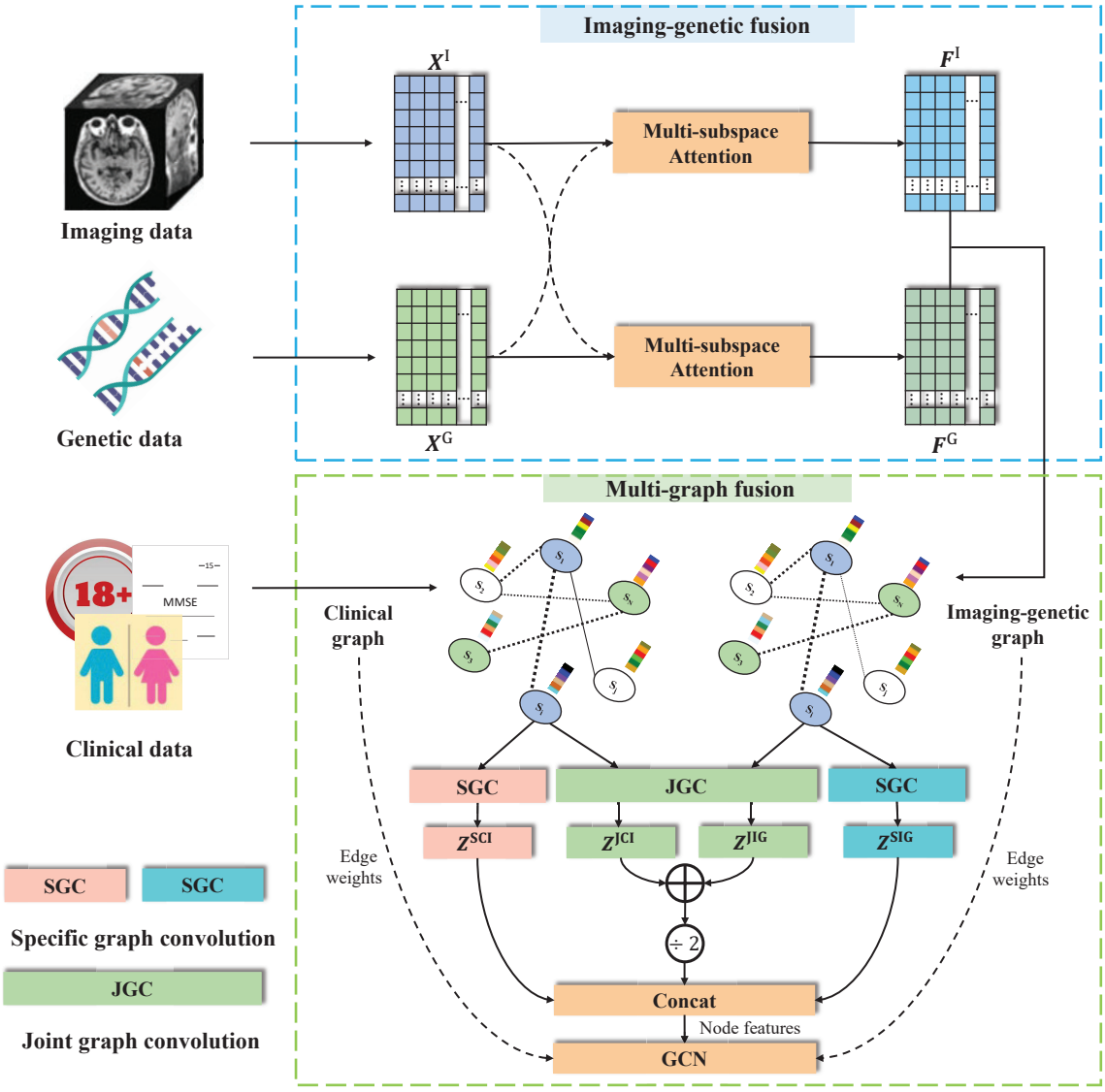


Fig. 1. Our proposed graph-based fusion (GBF) framework. Three different types of features (imaging, genetic, and clinical) are extracted before being used for fusion. GBF consists of two fusion modules that perform imaging-genetic fusion (IGF) and multi-graph fusion (MGF). IGF (top) uses a multi-subspace attention mechanism to fuse imaging features and genetic features, yielding imaging-genetic features. MGF (bottom) uses joint graph convolution and specific graph convolution to fuse imaging-genetic features and clinical features, and a graph convolution network (GCN) to diagnose degenerative disease.

attention (MSA) to exploit their relationships and yield fused imaging-genetic features. Then, MGF constructs two separate graphs using imaging-genetic features and clinical features, where joint graph convolution and specific graph convolution are used for graph fusion and degenerative disease diagnosis.

A. Imaging-genetic fusion

We propose an imaging-genetic fusion model to effectively extract joint and specific information from imaging and genetic data. Let $X^{\mathrm{I}} \in \mathbb{R}^{n\times d}$ and $X^{\mathrm{G}} \in \mathbb{R}^{n\times d}$ be imaging and genetic features obtained after preprocessing and feature extraction, where n is the number of subjects and d is the number of extracted features. To exploit the rich associations in imaging and genetic data, we project X^{I} and X^{G} into s dimensional subspaces to obtain their multi-subspace embeddings. Inspired by the original definition of attention mechanism for natural language processing [34], [35], we regard the embeddings of different subspaces as words and obtain attention scores by matching these embeddings with each other.

The schematic diagram of MSA for extracting imaging features that contain joint and specific information is shown in Fig. 2. Given X^{I} and X^{G}, we first use projection matrices $\mathcal{W}^{\mathrm{I}}$ and $\mathcal{W}^{\mathrm{G}}$ to obtain their multi-subspace embeddings $M^{\mathrm{G}} \in \mathbb{R}^{n\times s\times d_e}$ and $M^{\mathrm{I}} \in \mathbb{R}^{n\times s\times d_e}$, from which the attention matrix are generated. To avoid information bias, the number of features in each subspace embedding is d_e. For both embeddings, we use the dot product function to obtain their matching degree and use it to further calculate the attention score.

There are two attention matrices related to imaging features, one corresponds to joint information ($P^{\mathrm{JI}} \in \mathbb{R}^{n\times s\times s}$), and the other corresponds to specific information ($P^{\mathrm{SI}} \in \mathbb{R}^{n\times s\times s}$). For

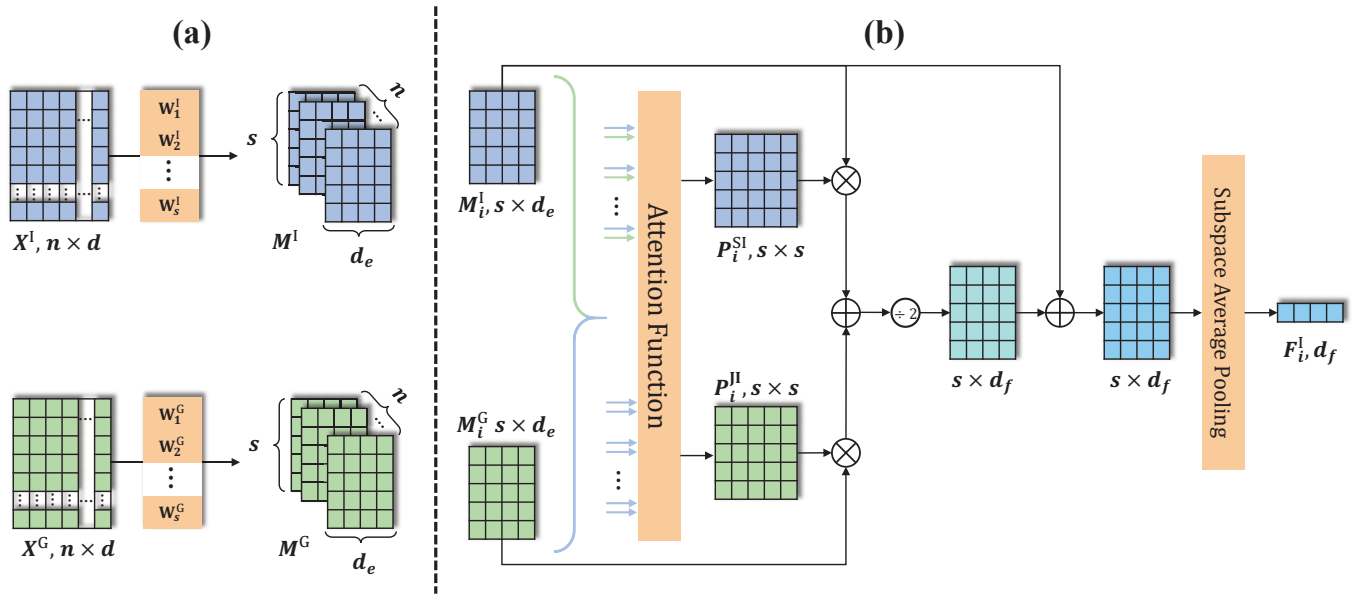


Fig. 2. Schematic diagram of multi-subspace attention (MSA). (a) Multi-subspace representations are obtained from extracted imaging and genetic features, respectively. (b) Imaging features of the i-th subject that contain joint and specific information are obtained by attention score within imaging data and genetic data.

the i-th subject, the attention vector between j-th subspace imaging features and genetic subspaces features is computed as:

$$P_{ij}^{\mathrm{JI}} = \frac{\exp\left(M_{ij}^{\mathrm{I}}\left(M_{i}^{\mathrm{G}}\right)^{\mathrm{T}}\right)}{\sum_{k=1}^{s}\exp\left(M_{ij}^{\mathrm{I}}\left(M_{ik}^{\mathrm{G}}\right)^{\mathrm{T}}\right)}. \tag{1}$$

where $M_{ij}^{\mathrm{I}} \in \mathbb{R}^{1\times d_e}$ denotes the feature vector of the j-th subspace of imaging data, and $M_i^{\mathrm{G}} \in \mathbb{R}^{s\times d_e}$ denotes genetic data in all subspace.

Correspondingly, the attention matrix w.r.t. joint information (i.e., imaging and genetic features) of the i-th subject is denoted as:

$$P_i^{\mathrm{JI}} = \left[\left(P_{i1}^{\mathrm{JI}}\right)^{\mathrm{T}}, \left(P_{i2}^{\mathrm{JI}}\right)^{\mathrm{T}}, \cdots, \left(P_{is}^{\mathrm{JI}}\right)^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{2}$$

To obtain the feature associations of different subspaces within imaging data, the attention matrix w.r.t. specific information (i.e., imaging features only) of the i-th subject is computed as:

$$P_{ij}^{\mathrm{SI}} = \frac{\exp\left(M_{ij}^{\mathrm{I}}\left(M_{i}^{\mathrm{I}}\right)^{\mathrm{T}}\right)}{\sum_{k=1}^{s}\exp\left(M_{ij}^{\mathrm{I}}\left(M_{ik}^{\mathrm{I}}\right)^{\mathrm{T}}\right)}, \tag{3}$$

$$P_i^{\mathrm{SI}} = \left[\left(P_{i1}^{\mathrm{SI}}\right)^{\mathrm{T}}, \left(P_{i2}^{\mathrm{SI}}\right)^{\mathrm{T}}, \cdots, \left(P_{is}^{\mathrm{SI}}\right)^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{4}$$

Given the attention matrices P_i^{JI} and P_i^{SI} of the i-th subject, the imaging features F_i^{I} that contain both joint and specific information are defined as:

$$F_i^{\mathrm{I}} = \frac{1}{s}\sum_{j=1}^{s}\left(M_{ij}^{\mathrm{I}} + \frac{1}{2}\left(P_{ij}^{\mathrm{SI}}M_{ij}^{\mathrm{I}} + P_{ij}^{\mathrm{JI}}M_{ij}^{\mathrm{G}}\right)\right). \tag{5}$$

Similarly, the attention matrices regarding genetic features of the i-th subject (P_i^{JG} and P_i^{SG}) can be derived in the same way, where the genetic features F_i^{G} are defined as:

$$F_i^{\mathrm{G}} = \frac{1}{s}\sum_{j=1}^{s}\left(M_{ij}^{\mathrm{G}} + \frac{1}{2}\left(P_{ij}^{\mathrm{SG}}M_{ij}^{\mathrm{G}} + P_{ij}^{\mathrm{JG}}M_{ij}^{\mathrm{I}}\right)\right). \tag{6}$$

Finally, the imaging-genetic features F^{IG} of the i-th subject are obtained by aggregating imaging features and genetic features:

$$F_i^{\mathrm{IG}} = \frac{1}{2}\left(F_i^{\mathrm{I}} + F_i^{\mathrm{G}}\right). \tag{7}$$

By means of the MSA mechanism, the IGF captures the joint information between imaging data and genetic data, as well as their individual specific information. The fused features are obtained by aggregating this information. Making use of the correlation between genetic data at the micro-level and imaging data at the macro-level, IGF provides more comprehensive and robust information for further fusion.

B. Multi-graph fusion

A graph is a suitable data structure for describing the relationships between subjects. Existing methods that construct a population graph for disease diagnosis usually use only clinical data. More importantly, the methods for graph construction are pre-defined, static, and thus not learnable from data. To obtain more accurate and adaptable graph structures for degenerative disease diagnosis, we used both imaging-genetic and clinical features for learnable multi-graph construction. Then, taking into account intra- and inter-graph relations, we propose a graph ensemble learning strategy to fuse them.

1) Graph construction: Due to the differences between imaging, genetic and clinical features, we construct an imaging-genetic graph (denoted as IG graph) and a clinical graph (denoted as CI graph), respectively.

Imaging-genetic graph We use a weighted graph $G^{\mathrm{IG}} = (F^{\mathrm{IG}}, E^{\mathrm{IG}})$ to define the imaging-genetic graph.

Here F^{IG} corresponds to the imaging-genetic features of all subjects, and $E^{\mathrm{IG}} = \{E_{ij}^{\mathrm{IG}} \mid i, j = 1, 2, ..., N\}$ corresponds to the similarities between each pair of subjects, defined as:

$$C_{ij} = 1 - \frac{\left(F_i - \overline{F_i}\right)\left(F_j - \overline{F_j}\right)}{\left|F_i - \overline{F_i}\right|_2 \left|F_j - \overline{F_j}\right|_2}, \tag{8}$$

$$E_{ij}^{\mathrm{IG}} = \exp\left(-\frac{{C_{ij}}^2}{2\overline{C_i}^2}\right), \tag{9}$$

where C_{ij} is the normalized cosine distance between i-th subject and j-th subject, $\overline{C_i}$ is the average value of C_{ij} in the range of $j \in [1, N]$ which can constrain the range of edge weights, and E_{ij}^{IG} corresponds to the similarity between i-th subject and j-th subject.

Clinical graph Compared to imaging-genetic data, clinical data are more likely to be corrupted by subjective factors, such as clinical scale data. Therefore, we propose a learnable graph construction method for clinical data.

The clinical graph is defined as $G^{\mathrm{CI}} = (F^{\mathrm{CI}}, E^{\mathrm{CI}})$ with vertices denoting subjects. F^{CI} is defined as the extracted clinical features in Sec. IV-B. We present $E^{\mathrm{CI}} = \{E_{ij}^{\mathrm{CI}} \mid i, j = 1, 2, ..., N\}$ as a weight matrix to indicate the similarity between subjects, where E_{ij}^{CI} is the connection weight between i-th subject and j-th subject.

In general, the weight matrix of the clinical graph is pre-defined. The pre-defined graph structure is static and not learnable, and therefore difficult to adapt to different diagnosis tasks. A better way is to construct an adaptable graph, and the edges of the graph should be generated dynamically. Inspired by the work of EV-GCN [21], we define the learnable function $\mathbb{F} : (F_i^{\mathrm{CI}}, F_j^{\mathrm{CI}})$ with parameters $\Omega = \{W^{\mathrm{E}}, b\}$ to estimate E_{ij}^{CI}. The pairwise encoder first normalizes the paired inputs F_i^{CI} and F_j^{CI}, respectively. After normalization, a projection layer is used to map each normalized input to a features space $\tilde{F} \in \mathbb{R}^D$. The representation of $\tilde{F}_i^{\mathrm{CI}}$ is calculated as:

$$\tilde{F}_i^{\mathrm{CI}} = \sigma\left(F_i^{\mathrm{CI}} W^{\mathrm{E}} + b\right), \tag{10}$$

where σ denotes the activation function of ReLU, W_F denotes the parameters of the pairwise encoder, and $\tilde{F}_i^{\mathrm{CI}}$ is the normalized F_i^{CI}. The weights between i-th subject and j-th subject are defined as the cosine similarity of their feature representations, defined as:

$$E_{ij}^{\mathrm{CI}} = \left(\cos\left(\tilde{F}_i^{\mathrm{CI}}, \tilde{F}_i^{\mathrm{CI}}\right) + 1\right)/2. \tag{11}$$

It is worth noting that the constructed imaging-genetic and clinical graphs model associations between subjects from different perspectives, and these graphs still need to be fused for disease diagnosis.

2) Graph Ensemble Learning: We propose a graph ensemble learning method to fuse the imaging-genetic graph and clinical graph. As illustrated in Fig. 1, it consists of four components. Specifically, the two graphs are fed into one joint graph convolution (JGC) layer with shared weights to extract joint information across graphs, and two individual specific graph convolution (SGC) layers to extract specific information within each graph. Their outputs are further fed into a graph convolution network (GCN), consisting of three Chebyshev graph convolutional layers, a fusion block, and an MLP for degenerative disease diagnosis.

Joint graph convolution (JGC) Given the imaging-genetic graph $G^{\mathrm{IG}} = (F^{\mathrm{IG}}, E^{\mathrm{IG}})$ and clinical graph $G^{\mathrm{CI}} = (F^{\mathrm{CI}}, E^{\mathrm{CI}})$, we use JGC with a shared weight W^{J} to yield a joint imaging-genetic embedding Z^{JIG} and a clinical embedding Z^{JCI}, defined as:

$$Z^{\mathrm{JIG}} = \left((D^{\mathrm{IG}})^{-\frac{1}{2}} E^{\mathrm{IG}} (D^{\mathrm{IG}})^{-\frac{1}{2}}\right) F^{\mathrm{IG}} W^{\mathrm{J}}, \tag{12}$$

$$Z^{\mathrm{JCI}} = \left((D^{\mathrm{CI}})^{-\frac{1}{2}} E^{\mathrm{CI}} (D^{\mathrm{CI}})^{-\frac{1}{2}}\right) F^{\mathrm{CI}} W^{\mathrm{J}}, \tag{13}$$

where D^{IG} and D^{CI} are the degree matrices of the similarity matrix E^{IG} and E^{CI}, respectively. The degree matrix D is a diagonal matrix whose elements in row i are the sum of the elements of row i in E, which performs the Laplacian normalization of the adjacency matrix to keep the scale of features unchanged in the graph convolution. The JGC [36], [37] allows the shared weight W^{J} to extract the joint information from the potential feature patterns that exist in both graphs.

The joint embedding Z^{J} is defined as:

$$Z^{\mathrm{J}} = \frac{1}{2}\left(Z^{\mathrm{JIG}} + Z^{\mathrm{JCI}}\right). \tag{14}$$

Specific graph convolution (SGC) We use two SGCs' to obtain specific embeddings separately. Given the imaging-genetic graph $G^{\mathrm{IG}} = (F^{\mathrm{IG}}, E^{\mathrm{IG}})$, the learned output embedding Z^{SIG}, parameterized by W^{IG}, is calculated as:

$$Z^{\mathrm{SIG}} = \left((D^{\mathrm{IG}})^{-\frac{1}{2}} E^{\mathrm{IG}} (D^{\mathrm{IG}})^{-\frac{1}{2}}\right) F^{\mathrm{IG}} W^{\mathrm{IG}}. \tag{15}$$

Likewise, given the clinical graph $G^{\mathrm{CI}} = (F^{\mathrm{CI}}, E^{\mathrm{CI}})$, the learned output embedding Z^{SCI} is calculated as

$$Z^{\mathrm{SCI}} = \left((D^{\mathrm{CI}})^{-\frac{1}{2}} S^{\mathrm{CI}} (D^{\mathrm{CI}})^{-\frac{1}{2}}\right) F^{\mathrm{CI}} W^{\mathrm{CI}}, \tag{16}$$

where W^{CI} is the learned weights.

Finally, the joint and specific embeddings are concatenated, denoted as $Z = \mathrm{Concat}(Z^{\mathrm{SCI}}, Z^{\mathrm{SIG}}, Z^{\mathrm{J}})$.

Graph convolution network (GCN) Given the fused adjacency matrix $E = E^{\mathrm{IG}} + E^{\mathrm{CI}}$, we obtain a new graph $G = (Z, E)$ integrating the imaging-genetic and clinical data. It is further input into three Chebyshev graph convolutional layers with dense connections and an MLP predictor [21] for degenerative disease diagnosis.

IV. Experiments

A. Datasets

1) ADNI: The Alzheimer's Disease Neuroimaging Initiative (ADNI: http://adni.loni.usc.edu/) is a multi-centre study of

TABLE I
STATISTICS OF SUBJECTS IN THIS STUDY ('STD' DENOTES STANDARD DEVIATION.

Datasets	Group	Female / Male	Age (mean±std)
ADNI	HC	52 / 61	73.60±6.03
	MCI	48 / 43	72.32±7.37
PPMI	HC	41 / 46	60.94±11.01
	PD	42 / 24	65.09±8.09

diseases such as AD and MCI. So far, there are three studies (ADNI-1, -GO and -2) in the dataset. It has recruited over 1700 subjects, aged between 55 and 90 years, from over 50 sites in the U.S. and Canada. It contains multi-modal data including neuroimaging, clinical, biological, and genetic biomarkers.

After data pre-processing, 204 ADNI subjects who have complete imaging, genetic and clinical data were retained, where 113 were diagnosed as healthy controls (HC) and 91 were diagnosed as MCI. Clinical information in ADNI includes site, age, sex, years of education, APOE4, mini-mental state examination (MMSE), and Rey auditory verbal learning test (RAVLT). Some statistical information about these subjects is shown in Table I.

2) PPMI: The Parkinson Progression Marker Initiative (PPMI: https://www.ppmi-info.org/) is a multi-center study designed to explore biomarkers associated with Parkinson's disease (PD) to improve understanding of the etiology and course of the disease. The PPMI cohort contains imaging and clinical data from approximately 400 patients with PD and 200 healthy subjects.

After data pre-processing, 87 controls and 66 patients with Parkinson's disease were used to validate our proposed method. Clinical information in PPMI contains site, age, sex, and the Geriatric depression scale (GDS). Some statistical information about these subjects is shown in Table I.

B. Data pre-processing

We perform data pre-processing and feature extraction on imaging, genetic, and clinical data before they are fused.

1) Imaging and genetic data: **Imaging data** FreeSurfer [38] was used for preprocessing of T1-weighted MRI (https://www.freesurfer.net/). The main steps include:

1) Head motion correction is carried out and the brain tissue is removed by using the surface deformation framework to achieve skull dissection.
2) MRI is converted to the Montreal Neurosciences Institute (MNI) standard space to complete the registration operation.
3) The brain's gray and white matter are separated.
4) 3D reconstruction of the brain is performed through MRI slices.
5) The cortical tissue is marked according to the Desikan-Killiany atlas [39].

Finally, the quantitative measurement values of surface area, gray matter volume, cortical thickness, mean curvature, and curvature index of each cerebral cortex tissue area in MRI were extracted by the automatic measurement framework in the FreeSurfer.

Genetic data Plink was used for pre-processing of the SNP data (http://www.cog-genomics.org/Plink/2.0). The main steps include:

1) Subjects and SNPs with deletions of more than 0.05 are removed.
2) Sex discrepancy of subjects is checked.
3) Minor allele frequency of all SNPs is calculated.
4) The Hardy-Weinberg equilibrium test is performed.
5) Subjects with high or low heterozygosity rates are excluded.
6) Relative subjects are deleted.

We selected disease-related top-10 genes and encoded the SNP sequence information of these genes as 0, 1 and 2 by additive coding [4].

For both imaging and genetic data, there is a large amount of task-irrelevant noise. Therefore, we used the recursive feature elimination (RFE) algorithm as an estimator to select d features that are most effective for degenerative disease diagnosis.

2) Clinical data: Clinical data often contains both continuous and discrete data. For continuous and ordinal variables such as age or clinical scores, we used them directly. For categorical variables, we encoded them beforehand. For example, sex is encoded as 0/1, where 0 corresponds to male, and 1 corresponds to female.

C. Experimental Settings

In this study, we followed Huang et al. [21] to set the hyper-parameters of our proposed approach:

- The order of Chebyshev graph convolution was 3.
- d, s, and d_e were set to 100, 3, and 100, respectively.
- The learning rate was set to 0.01, weight attenuation to 5×10^{-5}, and the number of iterations to 300.
- We employed cross-entropy loss on the labeled nodes to train the overall model.

Python 3.7, Pytorch 1.4, and scikit-learn package were used to implement our proposed GBF approach. For ADNI, we selected the top ten AD-related genes (APOE, BIN1, ABCA7, CLU, CR1, PICALM, MS4A6A, CD33, MS4A4E, and CD2AP) reported in the AlzGene (http://www.alzgene.org/) database. For PPMI, we selected the top ten PD-related genes (SNCA, ATP13A2, DDC, DRD2, MAOB, PARK7, PINK1, PRKN, SLC18A2, and TH) reported in the DisGeNET database (https://www.disgenet.org/).

The binary classification performance of our proposed approach was evaluated using four widely used metrics, i.e., accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC).

The range of the four metrics is between 0 and 1, where 1 corresponds to the best, and 0 corresponds to the worst. We used k-fold cross-validation (k=10) with 5 repetitions and report the mean and standard deviation of 5 experiments.

TABLE II
THE CLASSIFICATION RESULTS OF OUR PROPOSED GBF WITH DIFFERENT TYPES OF DATA. I: IMAGING DATA. G: GENETIC DATA. CL: CLINICAL DATA.

Task	Data	ACC (%)	SEN (%)	SPE (%)	AUC
HC vs MCI	I	73.33±1.54	66.15±4.08	79.12±3.05	0.677±0.014
	G	72.94±2.13	65.71±6.80	78.94±3.73	0.680±0.022
	CI	79.71±2.01	71.21±3.27	81.42±1.13	0.775±0.021
	IG	80.59±2.18	74.29±3.57	88.14±2.83	0.792±0.026
	IG + Cl	**84.80±1.92**	**78.46±4.26**	**91.86±3.09**	**0.847±0.023**
HC vs PD	I	71.63±2.68	74.85±1.89	78.62±6.71	0.709±0.032
	G	69.02±2.40	67.88±7.13	76.09±9.20	0.673±0.026
	CI	73.20±2.36	**82.42±4.27**	72.41±4.12	0.724±0.024
	IG	75.16±2.36	80.61±5.54	79.77±3.07	0.743±0.022
	IG + Cl	**78.30±3.19**	76.67±4.53	**80.69±1.22**	**0.779±0.041**

TABLE III
THE CLASSIFICATION RESULTS WITH AND WITHOUT RFE FOR FEATURE SELECTION.

Task	RFE	ACC (%)	SEN (%)	SPE (%)	AUC
HC vs MCI	-	82.76±1.68	74.29±2.87	89.38±2.33	0.824±0.020
	✓	**84.80±1.92**	**78.35±4.26**	**91.86±3.09**	**0.847±0.023**
HC vs PD	-	74.12±4.71	74.85±8.85	76.09±2.33	0.742±0.078
	✓	**78.30±3.19**	**76.67±4.53**	**80.69±1.22**	**0.779±0.041**

D. Ablation study

1) Effectiveness of data in different modalities: To investigate the effectiveness of imaging, genetic, and clinical data, we first trained the GBF on imaging data (I), genetic data (G), and clinical (CI) data, respectively. Then we compared the performance of the GBF trained on imaging-genetic data (IG) as well as imaging-genetic and clinical (IG+CI) data. When training on I and G data only, MSA, JGC, and SGC were removed. GCN was directly applied to the graph constructed by I data or G data. Similarly, when training on IG or CI data, JGC and SGC were removed.

The classification results are reported in Table II. It can be observed that data in the three different modalities are effective for degenerative disease diagnosis, where CI performs best, reaching 79.71% ACC in MCI diagnosis and 73.20% ACC in PD diagnosis. The model trained on IG data outperforms that trained on CI data, which implies that fused IG data contains more useful information than CI data. Our GBF model (IG+CI) performs better than that trained on CI or IG data, which demonstrates its effectiveness in IG and CI fusion.

2) Effectiveness of feature selection: In our study, we used the recursive feature elimination (RFE) algorithm to remove task-irrelevant features. To verify the validity of the feature selection, we compared it with a method that does not use RFE.

The performance comparison with different estimators for feature selection is reported in Table III. As can be seen, the performance of methods using RFE is better than that without using RFE. This demonstrates that RFE is capable of removing task-irrelevant features and reducing the noise of the model input.

3) Effectiveness of IGF and MGF: In this study, we propose an IGF module to fuse imaging and genetic features, and an MGF to fuse imaging-genetic and clinical features. To investigate their effectiveness, we compare the results of the GBF approach with and without the IGF and/or MGF modules. Specifically, imaging features and genetic features are simply concatenated while the IGF module is removed. Similarly, the IG graph and CI graph are simply concatenated (without SGC and JGC) while the MGF module is removed.

The classification results are reported in Table IV. For both diseases, methods with only one module (IGF or MGF) outperform those without any module, which demonstrates the effectiveness of IGF and MGF alone. The IGF module allows the model to achieve an ACC improvement of 1.76% and 1.96% on the baseline and MGF-inclusive approaches respectively. The MGF module gives the model an ACC gain of 1.66% and 1.86% for the baseline and IGF-inclusive approaches respectively. In other words, it indicates that IGF is effective in enhancing the representation of both imaging data and genetic data fusion, and MGF effectively exploits the relationships between different graphs. Moreover, the performance of our GBF approach, namely IGF + MGF, is better than those with one single module (IGF or MGF), which indicates the advantage of the proposed framework for feature fusion.

4) Effectiveness of components in IGF and MGF: In our proposed IGF and MGF, different components are used to capture the rich information in multi-modal data from different perspectives. In order to further validate the effectiveness of the different components in the IGF and MGF modules, we fix one module and investigate the contribution of different components in the other module.

TABLE IV
THE CLASSIFICATION RESULTS OF ABLATION STUDY ON IGF AND MGF MODULES.

Task	Module		ACC (%)	SEN (%)	SPE (%)	AUC
	IGF	MGF				
HC vs MCI	-	-	81.18±1.01	71.87±1.98	88.67±2.45	0.790±0.014
	✓	-	82.94±1.43	77.36±2.83	85.66±2.31	0.811±0.018
	-	✓	82.84±1.55	75.38±3.58	89.03±2.14	0.804±0.020
	✓	✓	**84.80±1.92**	**78.35±4.26**	**91.86±3.09**	**0.847±0.023**
HC vs PD	-	-	74.12±0.71	74.24±4.39	72.18±2.81	0.717±0.027
	✓	-	75.16±2.36	80.61±5.54	79.77±3.07	0.743±0.022
	-	✓	76.21±0.61	74.24±1.24	79.08±0.89	0.765±0.001
	✓	✓	**78.30±3.19**	**76.67±4.53**	**80.69±1.22**	**0.779±0.041**

TABLE V
THE CLASSIFICATION RESULTS OF ABLATION STUDY ON THE COMPONENTS OF THE IGF MODULE.

Task	Components		ACC (%)	SEN (%)	SPE (%)	AUC
	$\mathbf{MSA}_{\mathrm{I}}$	$\mathbf{MSA}_{\mathrm{G}}$				
HC vs MCI	-	-	78.14±1.76	74.29±3.62	83.01±3.42	0.770±0.020
	✓	-	79.22±1.43	76.04±3.48	85.13±3.23	0.779±0.011
	-	✓	79.80±0.76	**76.92±1.49**	82.12±1.64	0.788±0.017
	✓	✓	**80.59±2.18**	74.29±3.57	**88.14±2.83**	**0.792±0.026**
HC vs PD	-	-	73.99±2.90	77.27±1.24	76.55±6.13	0.713±0.008
	✓	-	74.64±1.55	76.97±3.26	72.41±4.05	0.717±0.019
	-	✓	74.90±1.94	79.70±5.34	70.34±2.23	0.728±0.026
	✓	✓	**75.16±2.36**	**80.61±5.54**	**79.77±3.07**	**0.743±0.022**

There are two key components in the IGF module, MSA for imaging features ($\mathbf{MSA}_{\mathrm{I}}$) and MSA for genetic features ($\mathbf{MSA}_{\mathrm{G}}$). $\mathbf{MSA}_{\mathrm{I}}$ captures specific information within imaging features and joint information between imaging and genetic features. $\mathbf{MSA}_{\mathrm{G}}$ captures specific information within genetic features and joint information between genetic and imaging features. We compare the results of models with or without $\mathbf{MSA}_{\mathrm{I}}$ and/or $\mathbf{MSA}_{\mathrm{G}}$ components while fixing the MGF module. The results are reported in Table V. The performance of the model with only one MSA component ($\mathbf{MSA}_{\mathrm{I}}$ or $\mathbf{MSA}_{\mathrm{G}}$) is better than that without an MSA component. Furthermore, the performance of the model with both components is better than that with only one MSA component. It demonstrates that our proposed IGF module effectively fuses imaging and genetic features.

Likewise, there are two types of important component in the MGF module, JGC and SGC. SGC captures the specific information within the IG graph or CI graph. JGC captures the joint information between the IG graph and the CI graph. In Table VI, we compare the results of the MGF module with or without JGC and/or SGC components while fixing the IGF module. It can be observed that the performance of the model with SGC or JGC outperforms that without any component. Meanwhile, the model with both JGC and SGC (i.e., the GBF approach) achieves the best performance on ACC, SPE, and AUC, which indicates the advantage of the MGF module.

E. Comparison with state-of-the-art methods

We compared our proposed GBF for degenerative diseases diagnosis with state-of-the-art (SOTA) methods, including PopulationGCN [20], InceptionGCN [40], EV-GCN [21], MOGONET [21], MMGL [1], Triplet Attention Net [12], and LG-GNN [33]. To fairly compare the performance of the different methods, we used the same data splitting and experimental setup.

Table VII shows the performance of these methods and our proposed GBF for MCI and PD diagnosis. Compared to PopulationGCN [20] that first applied GCN for degenerative disease diagnosis, InceptionGCN [40] obtained 4.51% accuracy increase in MCI diagnosis but 0.70% accuracy decrease in PD diagnosis by optimizing graph convolution, whereas EV-GCN [21] improved accuracy by 7.35% in MCI diagnosis and by 12.55% in PD diagnosis by optimizing graph construction. By using the attention mechanism, MMGL [1] and Triplet Attention Net [12] significantly improve the performance in comparison to PopulationGCN. MOGONET [41] and LG-GNN [33] enhance the performance through a simple multi-graph learning approach, achieving 76.18% and 79.12% accuracy in MCI diagnosis and 72.81% and 73.59% accuracy in PD diagnosis. Our proposed GBF approach provides a more efficient representation of the GCN through MSA between imaging and genetic data, adaptively constructs graphs more suitable for degenerative disease diagnosis, and provides rich information for graph ensemble learning through SGC and CGC. Our GBF approach performed the best, obtaining

TABLE VI
THE CLASSIFICATION RESULTS OF ABLATION STUDY ON THE COMPONENTS OF THE MGF MODULE.

Task	Components		ACC (%)	SEN (%)	SPE (%)	AUC
	SGC	JGC				
HC vs MCI	-	-	82.94±1.43	77.36±2.83	85.66±2.31	0.811±0.018
	✓	-	83.33±0.50	78.02±2.20	88.14±2.06	0.819±0.007
	-	✓	83.73±0.62	**79.78±2.26**	87.96±2.06	0.828±0.010
	✓	✓	**84.80±1.92**	78.35±4.26	**91.86±3.09**	**0.847±0.023**
HC vs PD			75.16±2.36	80.61±5.54	79.77±3.07	0.743±0.022
	✓	-	76.99±2.50	**83.33±4.60**	70.80±5.86	0.731±0.045
	-	✓	77.25±1.39	80.91±2.97	73.33±5.29	0.754±0.021
	✓	✓	**78.30±3.19**	76.67±4.53	**80.69±1.22**	**0.779±0.041**

TABLE VII
THE CLASSIFICATION RESULTS OF OUR PROPOSED GBF AND SOTA METHODS.

Task	Method	ACC (%)	SEN (%)	SPE (%)	AUC (%)
HC vs MCI	PopulationGCN [20]	69.51±4.38	68.79±7.98	64.78±5.11	0.671±0.059
	InceptionGCN [40]	74.02±3.18	70.33±5.48	72.39±6.87	0.728±0.053
	EV-GCN [21]	76.86±1.22	68.57±4.62	83.54±2.60	0.741±0.031
	MOGONET [41]	76.18±5.96	76.26±8.72	87.61±5.71	0.744±0.067
	MMGL [1]	76.86±3.26	59.34±10.06	91.15±3.13	0.752±0.038
	Triplet Attention Net [12]	82.84±3.94	75.82±5.28	86.90±3.61	0.826±0.029
	LG-GNN [33]	79.12±5.55	74.51±5.39	84.25±6.23	0.804±0.059
	GBF (Ours)	**84.80±1.92**	**78.35±4.26**	**91.86±3.09**	**0.847±0.023**
HC vs PD	PopulationGCN [20]	62.22±6.20	65.76±8.13	68.05±7.35	0.629±0.072
	InceptionGCN [40]	61.44±7.13	73.94±8.92	73.26±4.31	0.604±0.065
	EV-GCN [21]	74.77±4.35	73.33±5.38	75.63±4.70	0.734±0.057
	MOGONET [41]	72.81±5.38	68.18±6.63	75.63±6.81	0.712±0.061
	MMGL [1]	75.55±3.62	70.30±6.17	78.39±2.12	0.752±0.038
	Triplet Attention Net [12]	74.64±5.79	72.73±6.06	74.48±3.87	0.739±0.049
	LG-GNN [33]	73.59±7.96	71.52±4.76	73.79±5.14	0.720±0.054
	GBF (Ours)	**78.30±3.19**	**76.67±4.53**	**80.69±1.22**	**0.779±0.041**

84.80% accuracy for MCI and 78.30% accuracy for PD.

F. Important brain regions

To explore important brain regions associated with the early stages of MCI and PD, we mapped the attention scores of MRI back to the brain regions and sorted them according to attention scores in descending order. The five brain regions with the highest attention scores were considered to be important brain regions.

The locations of important brain regions associated with MCI and PD in the Desikan-Killiany atlas [39] are shown in Fig. 3. For MCI diagnosis, the regions of interest (ROIs) are rostral middle frontal, inferior parietal, superior frontal, and fusiform gyrus. Our finding is consistent with previous studies on Alzheimer's disease [42]–[46]. Boots et al. [47] found that atrophy in brain regions such as the frontal lobes was associated with cognitive performance such as language learning, memory, and visuospatial ability. Lindemer et al. [48] suggested that white matter abnormalities in brain regions such as the superior and inferior parietal lobes were associated with cerebrovascular dysfunction in the elderly and could further contribute to brain aging and dementia. Richter et al. [49] found that pyknotic gyrus volume was associated with facial matching function. For PD diagnosis, the ROIs are rostral anterior cingulate, superior temporal, superior frontal, and inferior parietal. These brain regions have also been frequently mentioned in previous PD-related studies [50]–[53]. Kikuchi et al. [54] suggested that brain regions such as the anterior cingulate and frontal lobe might contribute to dysphagia, one of the typical symptoms of PD patients. Blanchet et al. [55] suggested that abnormalities in the anterior cingulate and prefrontal lobe contribute to neuropathic pain in PD patients. New et al. [56] showed through analysis that the left and right superior temporal gyrus might be associated with speech motor disorders in PD, which were characterized by reduced pitch and loudness. Tahmasian et al. [57] suggested that the inferior parietal might contribute to abnormalities associated with behavioral perception and execution, as evidenced by the study of Herz et al. [58]. It is worth noting that the inferior parietal lobe is included in the ROIs of both MCI and PD. We

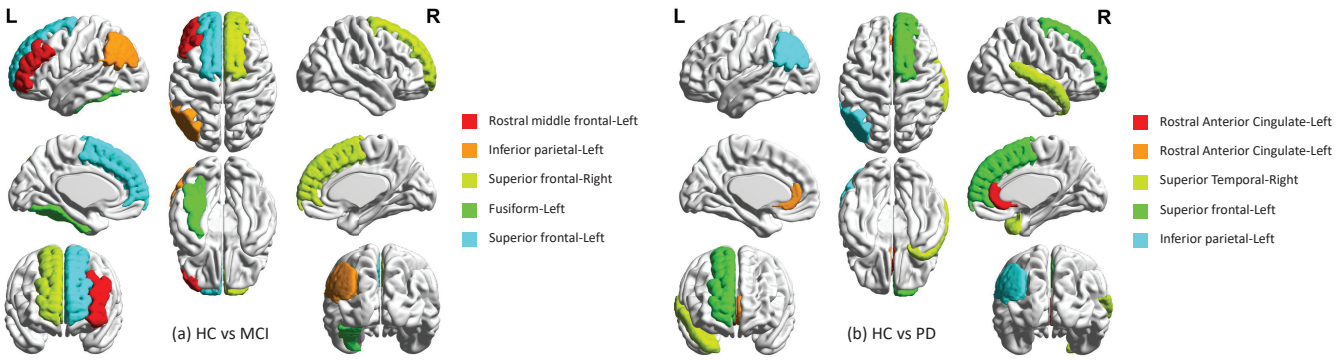


Fig. 3. Illustration of important brain regions associated with (a) MCI and (b) PD diagnosis.

believe it may be associated with the possible concomitant appearance of MCI and PD [59].

G. Limitations and future work

This study proposes a graph-based fusion approach for imaging, genetic and clinical data, including an image-genetic fusion module and a multi-graph fusion module. Although the expected results were achieved, there is still much room for improvement in future research.

From a data perspective, this study makes use of imaging, genetic, and clinical data for degenerative disease diagnosis. However, the etiology and symptoms of degenerative diseases are very complex. Multiple types of data such as gene sequences, gene expression, brain structure, function and metabolism as well as higher level brain network information can provide valuable information for the diagnosis of degenerative diseases. It is therefore important to consider the diagnosis of degenerative diseases in terms of more types of data such as imaging, genetic and clinical.

From a model perspective, this study proposes multi-subspace attention and graph ensemble learning, which to some extent address the problems in existing multi-modal fusion methods, such as how to consider the relationships between different data. However, the existing methods still face some problems, such as the lack of validation of large-scale cohorts, which may reduce the generalisability of the methods. In addition, there are persistent problems in the field of degenerative disease diagnosis such as strong-weak modal fusion, which also severely limit the performance of existing methods. Therefore, addressing the problems in the field of degenerative disease diagnosis from a model design perspective is also an important way to further improve model performance.

V. Conclusion

We proposed a novel graph-based fusion approach (GBF) to effectively exploit the relationships between imaging, genetic and clinical data for degenerative disease diagnosis. A comprehensive comparison with state-of-the-art methods validates the superiority of our proposed method. Using this approach, we obtain regions of interest in the diagnosis of degenerative diseases and validate their associations with the diseases in biomedical studies, which demonstrates the interpretability of our proposed approach. Notably, our proposed method not only accurately identifies MCI and PD, but also provides a guide in exploring the diagnosis of diseases from the perspective of multi-modal graph learning.

Acknowledge

This work was supported in part by the National Natural Science Foundation of China under Grant 62172444, in part by the Natural Science Foundation of Hunan Province under Grant 2022JJ30753, in part by the Central South University Innovation-Driven Research Programme under Grant 2023CXQD018, and in part by the High Performance Computing Center of Central South University.

References

[1] S. Zheng, Z. Zhu, Z. Liu, Z. Guo, Y. Liu, Y. Yang, and Y. Zhao, "Multi-modal graph learning for disease prediction," *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2207–2216, 2022.

[2] J. Liu, X. Tian, J. Wang, R. Guo, and H. Kuang, "Mtfil-net: automated alzheimer's disease detection and mmse score prediction based on feature interactive learning," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1002–1007, IEEE, 2021.

[3] L. Du, F. Liu, K. Liu, X. Yao, S. L. Risacher, J. Han, L. Guo, A. J. Saykin, L. Shen, and A. D. N. Initiative, "Identifying diagnosis-specific genotype–phenotype associations via joint multitask sparse canonical correlation analysis and classification," *Bioinformatics*, vol. 36, no. Supplement_1, pp. i371–i379, 2020.

[4] Y. Zhang, H. Zhang, L. Xiao, Y. Bai, V. D. Calhoun, and Y.-P. Wang, "Multi-modal imaging genetics data fusion via a hypergraph-based manifold regularization: Application to schizophrenia study," *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2263–2272, 2022.

[5] L. Du, J. Zhang, F. Liu, H. Wang, L. Guo, J. Han, A. D. N. Initiative, *et al.*, "Identifying associations among genomic, proteomic and imaging biomarkers via adaptive sparse multi-view canonical correlation analysis," *Medical Image Analysis*, vol. 70, p. 102003, 2021.

[6] J. Fu, W. Li, J. Du, and B. Xiao, "Multimodal medical image fusion via laplacian pyramid and convolutional neural network reconstruction with local gradient energy strategy," *Computers in Biology and Medicine*, vol. 126, p. 104048, 2020.

[7] H. Xu and J. Ma, "Emfusion: An unsupervised enhanced medical image fusion network," *Information Fusion*, vol. 76, pp. 177–186, 2021.

[8] L. Du, J. Zhang, Y. Zhao, M. Shang, L. Guo, J. Han, A. D. N. Initiative, *et al.*, "inmtscca: An integrated multi-task sparse canonical correlation analysis for multi-omic brain imaging genetics," *Genomics, Proteomics & Bioinformatics*, 2023, doi: https://doi.org/10.1016/j.gpb.2023.03.005.

[9] J. Liu, J. Wang, Z. Tang, B. Hu, F.-X. Wu, and Y. Pan, "Improving alzheimer's disease classification by combining multiple measures," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 15, no. 5, pp. 1649–1659, 2018.

[10] J. Liu, H. Du, R. Guo, H. X. Bai, H. Kuang, and J. Wang, "Mmgk: Multimodality multiview graph representations and knowledge embedding for mild cognitive impairment diagnosis," *IEEE Transactions on Computational Social Systems*, 2022, doi: https://doi.org/10.1109/TCSS.2022.3216483.

[11] X. Song, F. Zhou, A. F. Frangi, J. Cao, X. Xiao, Y. Lei, T. Wang, and B. Lei, "Multicenter and multichannel pooling gcn for early ad diagnosis based on dual-modality fused brain network," *IEEE Transactions on Medical Imaging*, vol. 42, no. 2, pp. 354–367, 2022.

[12] Q. Zhu, H. Wang, B. Xu, Z. Zhang, W. Shao, and D. Zhang, "Multimodal triplet attention network for brain disease diagnosis," *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3884–3894, 2022.

[13] W. Ko, W. Jung, E. Jeon, and H.-I. Suk, "A deep generative–discriminative learning for multimodal representation in imaging genetics," *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2348–2359, 2022.

[14] Y. Wang, Y. Feng, L. Zhang, J. T. Zhou, Y. Liu, R. S. M. Goh, and L. Zhen, "Adversarial multimodal fusion with attention mechanism for skin lesion classification using clinical and dermoscopic images," *Medical Image Analysis*, vol. 81, p. 102535, 2022.

[15] Q. Wang, L. Zhan, P. Thompson, and J. Zhou, "Multimodal learning with incomplete modalities by knowledge distillation," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1828–1838, 2020.

[16] Y. Zhu, X. Zhu, M. Kim, J. Yan, D. Kaufer, and G. Wu, "Dynamic hyper-graph inference framework for computer-assisted diagnosis of neurodegenerative diseases," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 608–616, 2018.

[17] J. Liu, M. Li, W. Lan, F.-X. Wu, Y. Pan, and J. Wang, "Classification of alzheimer's disease using whole brain hierarchical network," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 15, no. 2, pp. 624–632, 2016.

[18] J. Li, J. Liu, H. Yue, J. Cheng, H. Kuang, H. Bai, Y. Wang, and J. Wang, "Darc: Deep adaptive regularized clustering for histopathological image classification," *Medical image analysis*, vol. 80, p. 102521, 2022.

[19] P. Yi, L. Jin, T. Xu, L. Wei, and G. Rui, "Hippocampal segmentation in brain mri images using machine learning methods: A survey," *Chinese Journal of Electronics*, vol. 30, no. 5, pp. 793–814, 2021.

[20] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, and D. Rueckert, "Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease," *Medical image analysis*, vol. 48, pp. 117–130, 2018.

[21] Y. Huang and A. Chung, "Edge-variational graph convolutional networks for uncertainty-aware disease prediction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 562–572, Springer, 2020.

[22] T. Tong, K. Gray, Q. Gao, L. Chen, D. Rueckert, A. D. N. Initiative, *et al.*, "Multi-modal classification of alzheimer's disease using nonlinear graph fusion," *Pattern recognition*, vol. 63, pp. 171–181, 2017.

[23] L. Du, F. Liu, K. Liu, X. Yao, S. L. Risacher, J. Han, A. J. Saykin, and L. Shen, "Associating multi-modal brain imaging phenotypes and genetic risk factors via a dirty multi-task learning method," *IEEE transactions on medical imaging*, vol. 39, no. 11, pp. 3416–3428, 2020.

[24] L. Du, H. Wang, J. Zhang, S. Zhang, L. Guo, J. Han, and A. D. N. Initiative, "Adaptive structured sparse multiview canonical correlation analysis for multimodal brain imaging association identification," *Science China Information Sciences*, vol. 66, no. 4, p. 142106, 2023.

[25] J. Cheng, J. Liu, H. Yue, H. Bai, Y. Pan, and J. Wang, "Prediction of glioma grade using intratumoral and peritumoral radiomic features from multiparametric mri images," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 2, pp. 1084–1095, 2020.

[26] G. Pahuja and B. Prasad, "Deep learning architectures for parkinson's disease detection by using multi-modal features," *Computers in Biology and Medicine*, vol. 146, p. 105610, 2022.

[27] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, "Multimodal and multiscale deep neural networks for the early diagnosis of alzheimer's disease using structural mr and fdg-pet images," *Scientific reports*, vol. 8, no. 1, pp. 1–13, 2018.

[28] Q. Ying, X. Xing, L. Liu, A.-L. Lin, N. Jacobs, and G. Liang, "Multi-modal data analysis for alzheimer's disease diagnosis: An ensemble model using imagery and genetic features," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3586–3591, IEEE, 2021.

[29] M. Huang, X. Chen, Y. Yu, H. Lai, and Q. Feng, "Imaging genetics study based on a temporal group sparse regression and additive model for biomarker detection of alzheimer's disease," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1461–1473, 2021.

[30] M. Kim, J. H. Won, J. Youn, and H. Park, "Joint-connectivity-based sparse canonical correlation analysis of imaging genetics for detecting biomarkers of parkinson's disease," *IEEE transactions on medical imaging*, vol. 39, no. 1, pp. 23–34, 2019.

[31] Y. Shi, C. Zu, M. Hong, L. Zhou, L. Wang, X. Wu, J. Zhou, D. Zhang, and Y. Wang, "Asmfs: Adaptive-similarity-based multi-modality feature selection for classification of alzheimer's disease," *Pattern Recognition*, vol. 126, p. 108566, 2022.

[32] B. Lei, Y. Zhu, S. Yu, H. Hu, Y. Xu, G. Yue, T. Wang, C. Zhao, S. Chen, P. Yang, *et al.*, "Multi-scale enhanced graph convolutional network for mild cognitive impairment detection," *Pattern Recognition*, vol. 134, p. 109106, 2023.

[33] H. Zhang, R. Song, L. Wang, L. Zhang, D. Wang, C. Wang, and W. Zhang, "Classification of brain disorders in rs-fmri via local-to-global graph neural networks," *IEEE Transactions on Medical Imaging*, vol. 42, no. 2, pp. 444–455, 2022.

[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[35] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[36] J. Cheng, J. Liu, H. Kuang, and J. Wang, "A fully automated multimodal mri-based multi-task learning for glioma segmentation and idh genotyping," *IEEE Transactions on Medical Imaging*, vol. 41, no. 6, pp. 1520–1532, 2022.

[37] H. Yang, J. Sun, and Z. Xu, "Learning unified hyper-network for multi-modal mr image synthesis and tumor segmentation with missing modalities," *IEEE Transactions on Medical Imaging*, 2023, doi: https://doi.org/10.1109/TMI.2023.3301934.

[38] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

[39] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, *et al.*, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.

[40] A. Kazi, S. Shekarforoush, K. Kortuem, S. Albarqouni, N. Navab, *et al.*, "Self-attention equipped graph convolutions for disease prediction," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1896–1899, IEEE, 2019.

[41] T. Wang, W. Shao, Z. Huang, H. Tang, J. Zhang, Z. Ding, and K. Huang, "Mogonet integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification," *Nature Communications*, vol. 12, no. 1, pp. 1–13, 2021.

[42] D. H. Salat, D. N. Greve, J. L. Pacheco, B. T. Quinn, K. G. Helmer, R. L. Buckner, and B. Fischl, "Regional white matter volume differences in nondemented aging and alzheimer's disease," *Neuroimage*, vol. 44, no. 4, pp. 1247–1258, 2009.

[43] S. J. Greene, R. J. Killiany, A. D. N. Initiative, *et al.*, "Subregions of the inferior parietal lobule are affected in the progression to alzheimer's disease," *Neurobiology of aging*, vol. 31, no. 8, pp. 1304–1311, 2010.

[44] L. A. Rabin, A. J. Saykin, J. D. West, M. J. Borgos, H. A. Wishart, K. E. Nutter-Upham, L. A. Flashman, and R. B. Santulli, "Judgment in older adults with normal cognition, cognitive complaints, mci, and mild ad: Relation to regional frontal gray matter," *Brain imaging and behavior*, vol. 3, no. 2, pp. 212–219, 2009.

[45] D. Ma, I. S. Fetahu, M. Wang, R. Fang, J. Li, H. Liu, T. Gramyk, I. Iwanicki, S. Gu, W. Xu, *et al.*, "The fusiform gyrus exhibits an epigenetic signature for alzheimer's disease," *Clinical epigenetics*, vol. 12, no. 1, pp. 1–16, 2020.

[46] E. Niskanen, M. Könönen, S. Määttä, M. Hallikainen, M. Kivipelto, S. Casarotto, M. Massimini, R. Vanninen, E. Mervaala, J. Karhu, *et al.*, "New insights into alzheimer's disease progression: a combined tms and structural mri study," *PLoS One*, vol. 6, no. 10, p. e26113, 2011.

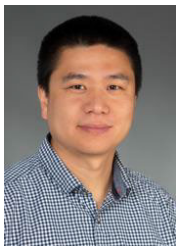
[47] E. A. Boots, S. A. Schultz, J. M. Oh, J. Larson, D. Edwards, D. Cook, R. L. Koscik, M. N. Dowling, C. L. Gallagher, C. M. Carlsson, *et al.*, "Cardiorespiratory fitness is associated with brain structure, cognition, and mood in a middle-aged cohort at risk for alzheimer's disease," *Brain imaging and behavior*, vol. 9, no. 3, pp. 639–649, 2015.

[48] E. R. Lindemer, D. N. Greve, B. R. Fischl, J. C. Augustinack, D. H. Salat, A. D. N. Initiative, *et al.*, "Regional staging of white matter signal

abnormalities in aging and alzheimer's disease," *NeuroImage: Clinical*, vol. 14, pp. 156–165, 2017.
[49] N. Richter, N. Beckers, O. A. Onur, M. Dietlein, M. Tittgemeyer, L. Kracht, B. Neumaier, G. R. Fink, and J. Kukolja, "Effect of cholinergic treatment depends on cholinergic integrity in early alzheimer's disease," *Brain*, vol. 141, no. 3, pp. 903–915, 2018.
[50] K. M. Davies, S. Bohic, A. Carmona, R. Ortega, V. Cottam, D. J. Hare, J. P. Finberg, S. Reyes, G. M. Halliday, J. F. Mercer, *et al.*, "Copper pathology in vulnerable brain regions in parkinson's disease," *Neurobiology of aging*, vol. 35, no. 4, pp. 858–866, 2014.
[51] J.-Y. Wang, Q.-Q. Zhuang, L.-B. Zhu, H. Zhu, T. Li, R. Li, S.-F. Chen, C.-P. Huang, X. Zhang, and J.-H. Zhu, "Meta-analysis of brain iron levels of parkinson's disease patients determined by postmortem and mri measurements," *Scientific reports*, vol. 6, no. 1, pp. 1–13, 2016.
[52] C. H. Adler and T. G. Beach, "Neuropathological basis of nonmotor manifestations of parkinson's disease," *Movement Disorders*, vol. 31, no. 8, pp. 1114–1119, 2016.
[53] D. Long, J. Wang, M. Xuan, Q. Gu, X. Xu, D. Kong, and M. Zhang, "Automatic classification of early parkinson's disease with multi-modal mr imaging," *PloS one*, vol. 7, no. 11, p. e47714, 2012.
[54] A. Kikuchi, T. Baba, T. Hasegawa, M. Kobayashi, N. Sugeno, M. Konno, E. Miura, Y. Hosokai, T. Ishioka, Y. Nishio, *et al.*, "Hypometabolism in the supplementary and anterior cingulate cortices is related to dysphagia in parkinson's disease: a cross-sectional and 3-year longitudinal cohort study," *BMJ open*, vol. 3, no. 3, p. e002249, 2013.
[55] P. J. Blanchet and C. Brefel-Courbon, "Chronic pain and pain processing in parkinson's disease," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 87, pp. 200–206, 2018.
[56] A. B. New, D. A. Robin, A. L. Parkinson, C. R. Eickhoff, K. Reetz, F. Hoffstaedter, C. Mathys, M. Sudmeyer, J. Michely, J. Caspers, *et al.*, "The intrinsic resting state voice network in parkinson's disease," *Human brain mapping*, vol. 36, no. 5, pp. 1951–1962, 2015.
[57] M. Tahmasian, S. B. Eickhoff, K. Giehl, F. Schwartz, D. M. Herz, A. Drzezga, T. van Eimeren, A. R. Laird, P. T. Fox, H. Khazaie, *et al.*, "Resting-state functional reorganization in parkinson's disease: an activation likelihood estimation meta-analysis," *Cortex*, vol. 92, pp. 119–138, 2017.
[58] D. M. Herz, S. B. Eickhoff, A. Løkkegaard, and H. R. Siebner, "Functional neuroimaging of motor control in parkinson's disease: A meta-analysis," *Human brain mapping*, vol. 35, no. 7, pp. 3227–3237, 2014.
[59] G. W. Duncan, M. J. Firbank, J. T. O'Brien, and D. J. Burn, "Magnetic resonance imaging: a biomarker for cognitive impairment in parkinson's disease?," *Movement Disorders*, vol. 28, no. 4, pp. 425–438, 2013.

Rui Guo received the master's degree of computer technology from Central South University, Changsha, Hunan, P.R. China. He is currently a data Analyst in the Second Xiangya Hospital of Central South University. His research interests include disease related multi-modal medical data analysis and machine learning.

Xu Tian received the master's degree of computer technology from Central South University, Changsha, Hunan, P.R. China. He is currently pursuing for the Ph.D degree in computer science at Central South University. His research interests include medical image analysis, multi-modal learning and deep learning.



Hanhe Lin received his Ph.D. at the Department of Information Science, University of Otago, New Zealand in 2016. From 2016 to 2021, he was a postdoc at the Department of Computer and Information Science at the University of Konstanz, Germany, where he was working on project A05 (visual quality assessment) of SFB-TRR 161, funded by the German Research Foundation (DFG). Currently, he is Lecturer in Computing at the University of Dundee, UK. His research interests include image processing, computer vision, machine learning, deep learning, and visual quality assessment.

Stephen McKenna received the B.Sc. degree (Hons.) in computer science from the University of Edinburgh, in 1990, and the Ph.D. degree in medical image analysis from the University of Dundee, in 1994. He was an EU Research Fellow, Italy, from 1994 to 1995, and an EPSRC Postdoctoral Researcher at the Queen Mary University of London, from 1995 to 1998. He has held visiting research positions at George Mason University, BT, USA, and Universidad Iberoamericana. He is currently a Professor (a Personal Chair) and the Computing Research Lead at the University of Dundee. He teaches courses on machine learning, image analysis, and their application to healthcare. His research interests include biomedical image analysis, computer vision, and machine learning. He serves on journal editorial boards which include Machine Vision and Applications, Journal of Imaging, and Scientific Data.

Hongdong Li received the Bachelor (Pharmaceutical Engineering) and PhD (Analytical Chemistry) degree from College of Chemistry and Chemical Engineering, Central South University, Changsha, P.R. China in 2007 and 2012, respectively. He is an Associate Professor in School of Computer Science and Engineering, Central South University, Changsha, Hunan, P.R. China. His research interests include bioinformatics, computational medicine and systems biology.

Fei Guo received the Ph.D. degree in computer science from Shandong University, China. She is currently a professor at the School of Computer Science and Engineering, Central South University, Changsha, Hunan, P.R. China. Her current research interests include bioinformatics and biomedicine. Her research has been cited over 3500 times (Google Scholar) with H-index=32.

Jin Liu received the Ph.D. degree in computer science from Central South University, China, in 2017. He is currently an associate professor at the School of Computer Science and Engineering, Central South University, Changsha, Hunan, P.R. China. His current research interests include medical image analysis and machine learning. His research has been cited over 2500 times (Google Scholar) with H-index=24.