Brain Structure-Function Fusing Representation Learning Using Adversarial Decomposed-VAE for Analyzing MCI

Qiankun Zuo, Ning Zhong, Yi Pan, Huisi Wu, Baiying Lei, *Senior Member, IEEE*, and Shuqiang Wang, *Senior Member, IEEE*

***Abstract*— Integrating the brain structural and functional connectivity features is of great significance in both exploring brain science and analyzing cognitive impairment clinically. However, it remains a challenge to effectively fuse structural and functional features in exploring the complex brain network. In this paper, a novel brain structure-function fusing-representation learning (BSFL) model is proposed to effectively learn fused representation from diffusion tensor imaging (DTI) and resting-state functional magnetic resonance imaging (fMRI) for mild cognitive impairment (MCI) analysis. Specifically, the decomposition-fusion framework is developed to first decompose the feature space into the union of the uniform and unique spaces for each modality, and then adaptively fuse the decomposed features to learn MCI-related representation. Moreover, a knowledge-aware transformer module is designed to automatically capture local and global connectivity features throughout the brain. Also, a uniform-unique contrastive loss is further devised to make the decomposition more effective and enhance the complementarity of structural and functional features. The extensive experiments demonstrate that the proposed model achieves better performance than other competitive methods in predicting and analyzing MCI. More importantly, the proposed model could be a potential tool for reconstructing unified brain networks and predicting abnormal connections during the degenerative processes in MCI.**

Manuscript received 1 June 2023; revised 19 August 2023 and 9 September 2023; accepted 4 October 2023. Date of publication 10 October 2023; date of current version 18 October 2023. This work was supported in part by the National Natural Science Foundations of China under Grant 62172403 and Grant U22A2041, in part by the Distinguished Young Scholars Fund of Guangdong under Grant 2021B1515020019, in part by the Excellent Young Scholars of Shenzhen under Grant RCYX20200714114641211, and in part by the Shenzhen Key Basic Research Projects under Grant JCYJ20200109115641762. *(Corresponding authors: Baiying Lei; Shuqiang Wang.)*

Qiankun Zuo is with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with the Hubei Internet Finance Information Engineering Technology Research Center and the School of Information Engineering, Hubei University of Economics, Wuhan, Hubei 430205, China.

Ning Zhong is with the Faculty of Engineering, Maebashi Institute of Technology, Maebashi 371-0816, Japan.

Yi Pan and Shuqiang Wang are with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China (e-mail: sq.wang@siat.ac.cn).

Huisi Wu is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China.

Baiying Lei is with the School of Biomedical Engineering, Shenzhen University, Shenzhen, 518060, China (e-mail: leiby@szu.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3323432

***Index Terms*— Structural-functional fusion, decomposed representation learning, knowledge-aware transformer, graph convolutional network, mild cognitive impairment.**

I. Introduction

MILD cognitive impairment (MCI) is considered an early stage of Alzheimer's Disease (AD) among older people [1]. It is characterized by memory loss, aphasia, and other brain function decline. Although not all older adults with MCI will develop AD, the annual conversion rate is 10%-15%. As the stage of AD is irreversible and incurable, early cognitive training and rehabilitation treatment are the keys to delaying or preventing the onset of dementia. Therefore, it is essential to develop effective methods for the diagnosis of MCI [2], [3], [4].

The brain network is suitable for characterizing the structural or functional relationships between brain regions by diffusion tensor imaging (DTI) or resting-state functional magnetic resonance imaging (fMRI) [6]. As parts of the brain's structural or functional connections may alter in people with MCI [7], it is common to extract connectivity-based features for early cognitive disease detection. Different from extracting features in Euclidean space [8], [9], [10], the general way of describing these features is to first split the whole brain into several spatially distributed regions of interest (ROIs) and then compute the connection strength between them from imaging data. Previous studies extracted the connectivity-based features from unimodal data and then built a classifier for cognitive disease detection. Since neuroimages from different modalities carry complementary information, current works [11], [12], [13] mainly focus on multimodal fusion by graph convolutional networks (GCN) and have achieved superior performance in disease diagnosis. However, these works heavily depend on the structural connectivity by empirical methods, which may lead to a large error in connection strength calculation because of the manually different parameter settings in certain software toolboxes. It may lose valuable information for disease prediction. Besides, the high noise and changing connectives derived from fMRI make it difficult to fuse with DTI, which cannot fully capture the complex brain network features in cognitive disease analysis.

The generative artificial intelligence has attracted widespread attention in medical image computing [5], [14],

[15], [16]. We combine the generative adversarial network (GAN) and variational autoencoder (VAE) for constructing brain networks from imaging data. As the transformer network [17] has an excellent ability to capture global information and model longer-distance dependencies for image recognition [18], [19], it is more suitable to automatically extract structural features of the spatially distributed brain ROIs and determine the connection strength among them. Since the common and complementary information from unimodal data is often mixed together, the multimodal fusion effect has been dramatically improved by decomposition-fusion representation learning via VAE in disease analysis [20], [21]. However, these works focus on image feature extraction in Euclidean space and ignore representation learning in topological space. It cannot analyze the connectivity features among spatially distributed ROIs in brain disease prediction. Moreover, the decomposed representations need to be adaptively integrated to learn effective connectivity features for disease analysis.

Inspired by the above observations, in this paper, a novel model termed brain structure-function fusing-representation learning (BSFL) is proposed to generate unified brain networks for predicting abnormal brain connections based on fMRI and DTI. Specifically, the knowledge-aware transformer network is designed to extract structural features for each ROI from DTI. Then the structural and functional features extracted from fMRI are sent to the decomposed variational graph autoencoders to decompose the feature space into uniform and unique spaces representing the common and complementary information for each modality. After that, the decomposed representations are utilized to reconstruct the input features to retain the unimodal information. Meanwhile, the representation-fusing generator combines these representations and generates unified brain networks, which are sent to the dual discriminator to make them class-discriminative and distribution-consistent. To ensure the effectiveness of decomposition, a uniform-unique contrastive loss function is utilized to constrain the distance in the decomposed representations within each modality and between modalities. As a result, the unified connectivity-based features are obtained to fully capture MCI-related information and provide reliable analysis of brain network abnormalities. The main contributions of this framework are as follows:

- The novel BSFL model is proposed to first learn the uniform and unique representations of each modality in topological spaces, and then adaptively integrate them to generate unified brain networks. It can greatly enhance the structural-functional feature fusion and effectively recognize the connectivity features that are highly related to MCI.
- The uniform-unique contrastive loss is devised to maximize the distance of the uniform and unique representations within each modality and minimize the distance of uniform representations between modalities, which makes the decomposition more effective and enhances the complementarity of structural and functional features.
- The knowledge-aware transformer (KAT) is designed to extract brain region features from DTI by introducing knowledge of the brain parcellation atlas. The proposed KAT can automatically learn the local and global connectivity features and capture the MCI-related structural information.

The rest of this paper is organized as follows: The related works are briefly described in Section II. The details of the proposed model are presented in Section III. In Section IV, the proposed BSFL and other competing methods are compared, and experimental results are presented on the public database. The reliability of the experimental results and the limitations of the proposed model are discussed in Section V. Finally, Section VI concludes the remarks of this study.

II. Related Work

The current brain network analysis methods for cognitive disease can be summarized in three categories: structural connectivity-based, functional connectivity-based, and multimodal connectivity-based approaches. The first approach focuses on morphology or water diffusion information to extract interrelated features of predefined ROIs for AD analysis. Pereira et al. [22] utilized cortical thickness information from T1-MRI to construct brain networks to analyze the abnormal topology properties between patient groups and healthy controls. Similarly, Wang et al. [23] constructed structural brain networks from DTI data to evaluate graph topological coefficients and demonstrated that the AD group had decreased global efficiency and local efficiency compared with normal controls. The brain network can also be characterized by the neural activity measured in each brain region. The second approach constructed functional connectivity from fMRI or Electroencephalogram (EEG) data and built classifiers to diagnose early AD. The work in [24] investigated subgroups of functional connectivities using EEG data and found abnormal changes in hub regions in AD patients. By defining spatially distributed ROIs, Jie et al. [25] utilized fMRI to extract connectivity-based features with multiple ROIs, which improved MCI diagnosis accuracy and provided valuable biomarkers for treatment. Considering multimodal data provides complementary information, the third approach used structural and functional connectivity to construct unified brain networks for AD diagnosis and treatment. To discover interpretable connections, Lei et al. [26] presented an auto-weighted centralized multi-task model to combine the two kinds of connectivities for the MCI study, which has achieved excellent diagnosis performance and estimated essential brain connections for further treatment. Nevertheless, the previous models adopted structural connectivity directly from the empirical methods, which may be inaccurate and ineffective for downstream feature extraction because of different manual parameter settings. Moreover, the functional connections from fMRI may be influenced by the selection of sliding time windows and the high noise.

There are two strategies to learn representations from multimodal images for brain disease analysis: mixed representation learning and decomposed representation learning. The former strategy extracts latent features from unimodal data separately

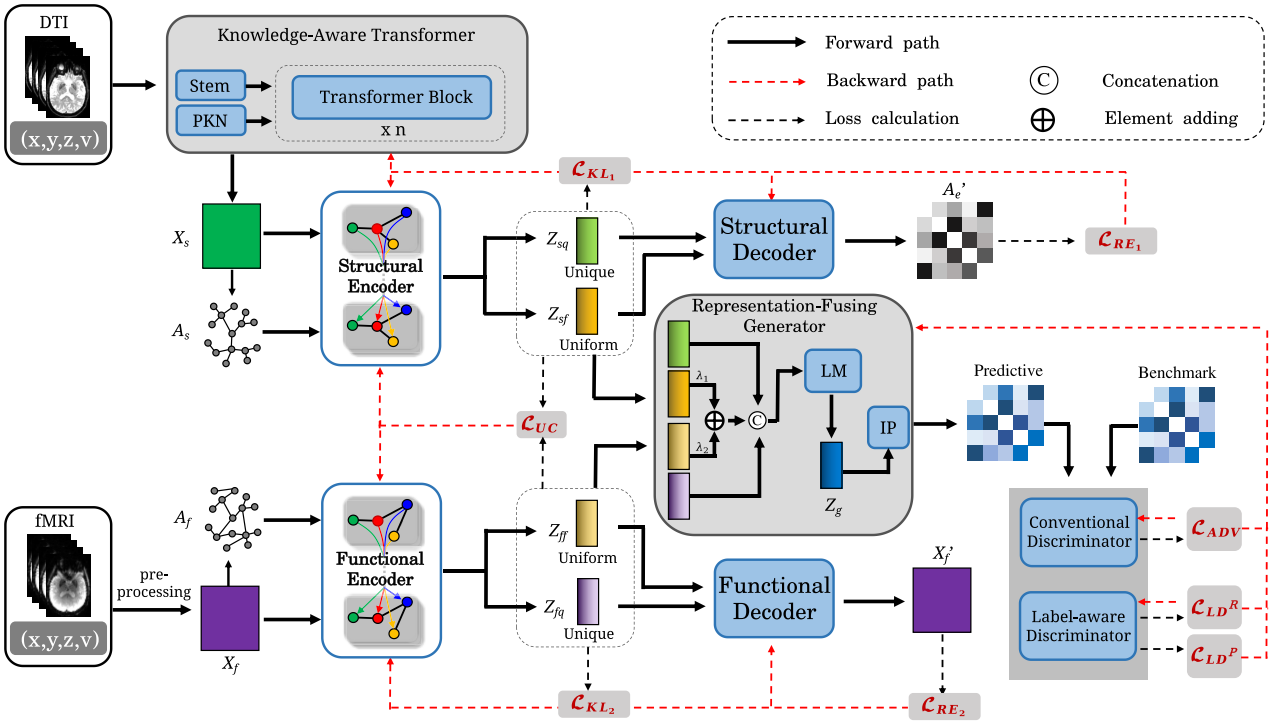

Fig. 1. The overall framework of the proposed BSFL for MCI diagnosis using fMRI, DTI, and the knowledge of template-segmented brain regions. It consists of eight components: a transformer, two encoders, two decoders, a generator, and two discriminators. The output of the representation-fusing generator is the unified brain network, which is used for MCI analysis.

and fuses these features by concatenation or other specific mechanisms (i.e., averaging or weighting). Zhou et al. [27] used the combination of volumetric measures calculated from T1-weighted MRI, metabolic measures generated from positron emission tomography (PET), and genetic measurements extracted from single nucleotide polymorphism (SNP) as input features to diagnose AD. Also, the work in [26] merged structural connectivity features with functional connectivity features using a weighting scheme and then adopted an SVM classifier for early AD diagnosis. Decomposed representation learning via VAE has shown great potential and has become mainstream in medical image analysis [28], [29]. It jointly encodes each modal image into latent representations with separate meanings and combines these multimodal representations for downstream tasks. Zhang et al. [30] proposed a VAE-based model to decompose multi-view brain networks from DTI and learn a unified representation, which improves MCI diagnosis performance. Similarly, Cheng et al. [31] applied multimodal VAE to learn common and distinctive representations from preoperative multimodal images for glioma grading. In general, mixed representation learning may lead to common information redundancy and the degradation of the fusion effect. And because decomposed representation learning concentrates on the euclidian space for disease diagnosis, it is not suitable for brain disease analysis in terms of brain topological characteristics.

III. Method

A. Overview

The flowchart of BSFL is shown in Fig. 1. After some preprocessing steps, given the fMRI and DTI, the proposed model learns a complicated non-linear mapping network to transform the bimodal images into brain networks for detecting abnormal brain connections at different stages of MCI. The proposed model consists of four parts: 1) the knowledge-aware transformer; 2) the decomposed variational graph autoencoders; 3) the representation-fusing generator; and 4) the dual discriminator. The last three parts are defined as the decomposition-fusion framework. First, the transformer-based network extracts structural features from DTI by incorporating location and volume information for predefined ROIs. Then, the feature space is decomposed into unique and uniform spaces for each modality by the decomposed variational graph autoencoders. Finally, the decomposed representations are fused to generate unified brain networks by the representation-fusing generator and the dual discriminator. The proposed model is featured by incorporating the following objective functions: the Kullback-Leibler (KL) loss, the reconstruction loss, the adversarial loss, the classification loss, and the uniform-unique contrastive loss. These loss functions aim to ensure decomposition thoroughness and enhance structural-functional fusion.

B. Architectures

1) Knowledge-Aware Transformer: In this section, the transformer-based network is adopted to extract structural features for each predefined ROI from DTI. The knowledge-aware transformer consists of a stem module, a prior knowledge normalization (PKN) module, and a transformer block. A stem is a common form of convolution with defined kernels and strides that is applied to extract low-level features from the fraction anisotropy (FA) image (as described in Section IV-A).

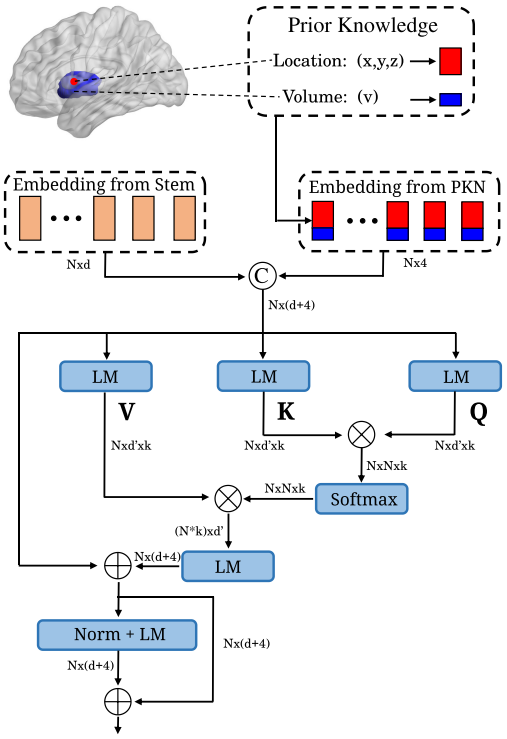

Fig. 2. The detailed structure of the transformer block in KAT. The stem embedding does not correspond to the ROIs. The output of this module is the ROI's structural features.

Suppose the brain is divided into N ROIs. In the proposed model, the $3 \times 3 \times 3$ convolutions [32], [33] are stacked with 2-stride and 1-stride at intervals, followed by a single $1 \times 1 \times 1$ convolution in the penultimate layer to match the N-channel feature for ROIs. The filter numbers are 8, 8, 16, 16, 32, 64, and N for the seven convolutional layers. In the last layer of the stem module, a one-layer linear mapping (LM) network transforms the flattened ROI feature into d-dimensional embedding and inputs it to the transformer block.

Prior knowledge refers to the spatial location and morphology information of the predefined ROIs. According to the standard anatomical template, it can provide the central location (i.e., $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$) and volume (i.e., $\boldsymbol{v}$) information for each ROI. The PKN module normalizes the location and volume information into the range $-1 \sim 1$. For example, $PKN(\boldsymbol{x}_i) = 2(\boldsymbol{x}_i - min(\boldsymbol{x}))/(max(\boldsymbol{x}) - min(\boldsymbol{x})) - 1$, where $\boldsymbol{x} \in \mathbb{R}^N$, N is the number of ROIs. This formula can be applied to other prior knowledge (i.e., $\boldsymbol{y}$, $\boldsymbol{z}$, and $\boldsymbol{v}$).

As illustrated in Fig. 2, both outputs of the stem and PKN are sent to the transformer block to learn spatial and morphological features for each ROI. One stage of the transformer block consists of prior multi-head self-attention (PMS) and feed-forward networks (FFN). Here, the $F_e \in \mathbb{R}^{N \times d}$ and $F_p \in \mathbb{R}^{N \times 4}$ are denoted as the embedding of the Stem and PKN modules, respectively, where d indicates the output dimension of the stem module. The output of each transformer block is:

$$E_{out} = E_{hidden} + FFN(LM(E_{hidden})) \quad (1)$$

$$E_{hidden} = F_e + PMS(LM(F_e, F_p)) \quad (2)$$

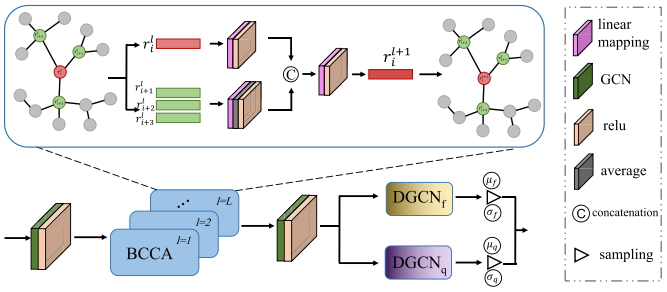

Fig. 3. The network structure of the structural encoder. It accepts both structural connectivity and structural features and outputs two pairs of variables: (μ_f, σ_f), and (μ_q, σ_q). Each pair is used to obtain the decomposed representations (i.e., uniform and unique).

where, $E_{hidden} \in \mathbb{R}^{N \times d}$ and $E_{out} \in \mathbb{R}^{N \times d}$. In particular, the embedding F_e is projected to get value $\mathbf{V}$ by applying k parallel linear mapping layers (i.e., heads). Also, the query $\mathbf{Q}$ and key $\mathbf{K}$ are obtained in the same way by concatenating the embeddings F_e and F_p as the input of the linear mapping layers. The dimension of the three tokens is d', where $d' = d/k$. For example, $k = 1$, and the PMS can be simplified to prior single-head self-attention (PSS). It can be defined as:

$$PSS(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = Softmax(\mathbf{Q}\mathbf{K}^T/\sqrt{d'})\mathbf{V} \quad (3)$$

The output values of each PSS are concatenated and linearly transformed to generate the hidden result E_{hidden}. Then, it is sent to FFN with one linear mapping layer and a $tanh$ activation function. Finally, the output E_{out} is combined with the normalized prior knowledge F_p to input the next transformer block. The output E_{out} of the last transformer block is linearly mapped into X_s with the dimension d.

2) Decomposed Variational Graph Autoencoders: After the structural features $X_s \in \mathbb{R}^{N \times d}$ and the functional features $X_f \in \mathbb{R}^{N \times d}$ extracted from the empirical method have been mined separately, decomposing the bimodal features can significantly improve the common and complementary information fusion for representation learning. Given features extracted from two modalities, this section can learn about decomposed representations among modalities. There are two parts: two encoders and two decoders.

The two encoders share the same structure in Fig. 1. Firstly, the structural connectivity A_s is constructed by the matrix inner product: $A_s = \sigma(X_s X_s^T)$. The functional connectivity A_f is constructed by the matrix inner product: $A_f = \sigma(X_f X_f^T)$. Then, each modal graph feature (i.e., X_s, A_s) is sent to the encoder to get a pair of variables (i.e., σ, μ). After that, the pair of variables is inferred to learn decomposed representations. Finally, the decomposed representations are utilized to reconstruct the input features. The detailed information on the structural encoder is shown in Fig. 3. The GCN is a two-layer network with 128 and 64 neurons. Each layer is followed by the rectified linear unit ($ReLU$) activation function. The brain connectivity and central attention (BCCA) block is added to capture the global correlation between two ROIs iteratively. Every ROI's feature is updated by combing its feature with other ROIs' features. The linear mapping is one dense layer with 64 neurons. The dual graph convolutional

network (DGCN) block is added to generate a pair of latent variables. It consists of two separate GCNs with one h-neuron layer and can output one pair of latent variables representing common and complementary information. The outputs of the structural encoder are μ_{sf}, σ_{sf}, μ_{sq}, and σ_{sq}, and the outputs of the functional encoder are μ_{ff}, σ_{ff}, μ_{fq}, and σ_{fq}.

To infer representations in latent space, a standard normal distribution constraint at the end of the encoder is added to get latent representations. The formula can be expressed as:

$$Z_{sf} = \mu_{sf} + \sigma_{sf} \odot \varepsilon_1, and\ Z_{sq} = \mu_{sq} + \sigma_{sq} \odot \varepsilon_2 \quad (4)$$

$$Z_{ff} = \mu_{ff} + \sigma_{ff} \odot \varepsilon_3, and\ Z_{fq} = \mu_{fq} + \sigma_{fq} \odot \varepsilon_4 \quad (5)$$

where, μ_{sq}, σ_{sq} are the mean and standard deviation matrix of the structure-specific component, while μ_{sf} and σ_{sf} are the mean and standard deviation matrix of the uniform component in the structural encoder. The symbols in the functional encoder also have the same meaning. $\odot$ denotes an element-wise product. ε_i $(i \in 1, 2, 3, 4)$ means a matrix sampled from a Gaussian distribution. Z_{sf}, Z_{sq}, Z_{ff}, and Z_{fq} share the same size $N \times h$.

The reconstruction module can retain unimodal information and enhance the stability of the model. For structural decoder, it accepts both structure-specific representation Z_{sq} and uniform representation Z_{sf} and outputs structural adjacent matrix A'_e with the dimension size $N \times N$. The network is the reverse operation of the structural encoder, followed by an inner product operation and a $sigmoid$ activation function. Similarly, the functional decoder transforms the function-specific representation Z_{fq} and uniform representation Z_{ff} into the original function time series $X'_f \in \mathbb{R}^{N \times d}$ by using the inverse network structure of the functional encoder.

3) Representation-Fusing Generator: Since the latent representations have been decomposed into unique and uniform components, it is easy to find the best weighting parameters between different components in the fusion process. The Multi-Layer Perceptron (MLP)-based generator is designed to fuse the decomposed representations to generate unified brain networks A_p. The generator can adaptively adjust the weight between decomposed representations, which fully reflects the common and complementary information among modalities.

In the representation-fusing generator, the uniform representations from structural-functional data are first added with certain weight values, then concatenated with the unique representations. The formula can be expressed as:

$$Z_c = (Z_{sq} \,||\, Z_m \,||\, Z_{fq}) \quad (6)$$

$$Z_m = \lambda_1 Z_{sf} + \lambda_2 Z_{ff} \quad (7)$$

where λ_1 and λ_2 determine the relative importance of the uniform representations from the bimodal data. In the experiment, both of them are set to 0.5. $||$ means concatenation. $Z_c \in \mathbb{R}^{N \times 3h}$ is the concatenated representation. After that, a two-layer linear mapping network with $2h$ and h neurons is designed to adaptively fuse the learned representations and obtain the fused representation Z_g:

$$Z_g = LM(Z_c) \quad (8)$$

Finally, the fused representation $Z_g \in \mathbb{R}^{N \times h}$ is transformed into the unified brain network through inner product (IP) operation. The predictive unified brain network $A_p \in \mathbb{R}^{N \times N}$ is expressed as:

$$A_p = \sigma(Z_g Z_g^T) \quad (9)$$

where, the σ is a $sigmoid$ function.

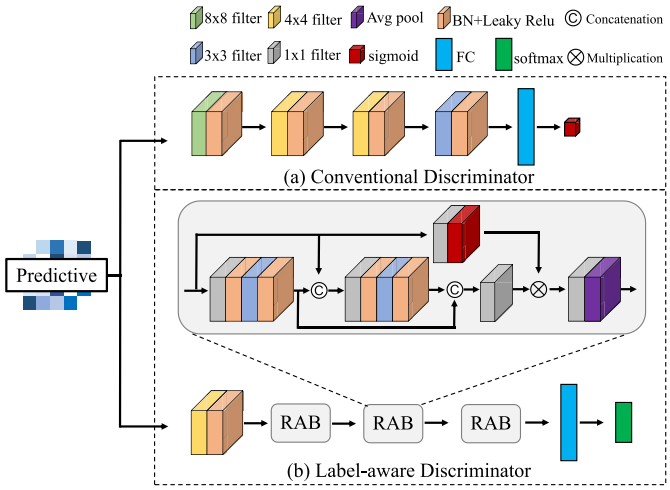

Fig. 4. The illustration of the dual discriminator in the overall framework. The label-aware discriminator is used for disease prediction.

4) Dual Discriminator: As shown in Fig. 4, the conventional discriminator is used to keep the output (i.e., A_p) consistent with the real sample (i.e., A_b) distribution, where the real sample is computed using the graph-based deep model (GBDM) [34]. The filter numbers are 8, 16, 32, and 64, and the fully connected (FC) layer has 1024 neurons. The label-aware discriminator can classify if the input matrix is normal control or patients. It consists of one convolution layer, three residual attention blocks (RAB), and a fully connected layer. The kernel numbers of 4×4, 3×3 and 1×1 in this discriminator are 8, 8, and 16, respectively. After three RABs, a 576-neuron FC layer and a softmax layer are added to make the feature A_p discriminative.

C. Hybrid Loss Function

In this study, the KAT, encoders, decoders, generator, and discriminators are combined into the BSFL model to learn MCI-related representations and jointly trained with the following losses.

1) KL Loss: Assuming the latent representations obey the normal Gaussian distribution $\mathbb{N}(0, 1)$, the output of encoders is defined by $E_s(X_s)$ and $E_f(X_f)$. KL divergence is adapted to constrain the output to match a Gaussian distribution by introducing the reparameterization technique [35]. The expression is defined below:

$$\begin{aligned}\mathcal{L}_{KL} &= \mathcal{L}_{KL_1} + \mathcal{L}_{KL_2} \\ &= \mathbb{E}_{X_s \sim P_{DTI(X_s)}}[\emptyset_{KL}(\mathbb{E}(E_s(X_s))|\mathbb{N}(0,1))] \\ &\quad + \mathbb{E}_{X_f \sim P_{fMRI(X_f)}}[\emptyset_{KL}(\mathbb{E}(E_f(X_f))|\mathbb{N}(0,1))]\end{aligned} \quad (10)$$

where, $\mathbb{E}$ indicates expected value, $\emptyset$ indicates KL divergence.

2) Reconstruction Loss: The reconstruction can stabilize the representation learning. The uniform and unique representations are combined to reconstruct structural or functional features. Here, the D_s and D_f are denoted as the structural and functional decoders, respectively. A_e is the empirical structural connectivity computed from the PANDA toolbox. The loss is defined as:

$$\begin{aligned}\mathcal{L}_{RE} &= \mathcal{L}_{RE_1} + \mathcal{L}_{RE_2} \\ &= [\mathbb{E}_{Z_{sf}\sim P_{(Z_{sf})}, Z_{sq}\sim P_{(Z_{sq})}} ||A_e - D_s(Z_{sf}, Z_{sq})||_2 \\ &\quad + \mathbb{E}_{X_s\sim P_{(X_s)}} ||A_e - \sigma(X_s X_s^T))||_2] \\ &\quad + \mathbb{E}_{Z_{ff}\sim P_{(Z_{ff})}, Z_{fq}\sim P_{(Z_{fq})}} ||X_f - D_f(Z_{ff}, Z_{fq})||_2\end{aligned} \tag{11}$$

3) Adversarial Loss: The output of the representation-fusing generator is the unified brain network, which is defined as A_p. Here, $A_p = G(Z_{sf}, Z_{sq}, Z_{ff}, Z_{fq})$. The benchmark brain network is denoted as A_b, computed from the GBDM method by deeply fusing fMRI and DTI. The conventional discriminator is represented as D_c. The adversarial loss can be written as:

$$\mathcal{L}_D = \mathbb{E}_{A_p\sim P_{A_p}}[(D_c(A_p))^2] + \mathbb{E}_{A_b\sim P_{A_b}}[(D_c(A_b) - 1)^2] \tag{12}$$

$$\mathcal{L}_G = \mathbb{E}_{A_p\sim P_{A_p}}[(D_c(A_p) - 1)^2] \tag{13}$$

$$\mathcal{L}_{ADV} = \mathcal{L}_G + 0.1\mathcal{L}_D \tag{14}$$

4) Classification Loss: To discriminate the unified brain network, a label-aware discriminator is defined as D_l to classify if A_p and A_r are normal controls or patients. The formula is defined as:

$$\begin{aligned}\mathcal{L}_{CL} &= \mathcal{L}_{LD}^R + \mathcal{L}_{LD}^P \\ &= \mathbb{E}_{A_b\sim P_{(A_b)}}[-I \cdot log(D_l(A_b))] \\ &\quad + \mathbb{E}_{A_p\sim P_{(A_p)}}[-I \cdot log(D_l(A_p))]\end{aligned} \tag{15}$$

where, I is the truth label.

5) Uniform-Unique Contrastive Loss: To constrain the learned decomposed representations, a uniform-unique contrastive (UC) loss function is applied to constrain the distance between them. The expression is defined as:

$$\begin{aligned}\mathcal{L}_{UC} &= \frac{1}{4}(\mathcal{L}_{UC_1} + \mathcal{L}_{UC_2}) + \frac{1}{2}\mathcal{L}_{UC_3} \\ &= \frac{1}{4}\mathbb{E}_{Z_{sq}, Z_{sf}}[max(margin - ||(Z_{sq} - Z_{sf})||_2, 0)] \\ &\quad + \frac{1}{4}\mathbb{E}_{Z_{fq}, Z_{ff}}[max(margin - ||(Z_{fq} - Z_{ff})||_2, 0)] \\ &\quad + \frac{1}{2}\mathbb{E}_{Z_{sf}, Z_{ff}}(||Z_{sf} - Z_{ff}||_2)\end{aligned} \tag{16}$$

here, $margin$ indicates the threshold with a default value of 1. The detailed training steps of the proposed BSFL are described in Algorithm 1 for reference.

Algorithm 1 The optimization procedure for the BSFL

Input: each mode's features DTI and X_f, empirical structural connectivity A, real unified brain network A_r, maximal iterative number $maxIter$, training step parameter t, model parameters Θ, hyper-parameters λ_1, λ_2 (set both as 0.5), prior knowledge x, y, z, v.

Output: predictive unified brain network A_p, model parameters Θ;

initialization: Θ, t, $maxIter$.
repeat
  $t \leftarrow t + 1$
  compute structural features based on the KAT module: $X_s = KAT(DTI, x, y, z, v)$;
  compute the structural uniform and unique variables based on the encoder E_s: (μ_{sf}, σ_{sf})=$Uniform(E_s(X_s))$, (μ_{sq}, σ_{sq})=$Unique(E_s(X_s))$;
  compute the functional uniform and unique variables: based on the encoder E_f: (μ_{ff}, σ_{ff})=$Uniform(E_f(X_f))$, (μ_{fq}, σ_{fq})=$Unique(E_f(X_f))$;
  compute the uniform and unique representations based on the above variables and the random noise ε sampled from a standard normal distribution $\mathbb{N}(0, 1)$: $Z_{sf} = \mu_{sf} + \sigma_{sf} \odot \varepsilon_1$, $Z_{sq} = \mu_{sq} + \sigma_{sq} \odot \varepsilon_2$, $Z_{ff} = \mu_{ff} + \sigma_{ff} \odot \varepsilon_3$, $Z_{fq} = \mu_{fq} + \sigma_{fq} \odot \varepsilon_3$;
  compute the unified brain network A_p based on the representation-fusing generator G: $A_p = G(Z_{sf}, Z_{sq}, Z_{ff}, Z_{fq})$;
  reconstruct the input features A' and $X_f^{'}$ with the decoders D_s and D_f: $A_e' = D_s(Z_{sf}, Z_{sq})$, $X_f^{'} = D_f(Z_{ff}, Z_{fq})$;
  Update the Θ in conventional discriminator D_c by back propagating the gradient $\nabla_\Theta L_D^t$;
  add the KL loss Eq. (10), the reconstruct loss Eq. (11), the generator loss Eq. (13), the classification loss Eq. (15) and the uniform-unique contrastive loss Eq. (16) to the loss L_{merge}^t;
  calculate the gradient loss $\nabla_\Theta L_{merge}^t$;
  replace A_p with A_p^t and update the Θ in encoders, decoders, generators and label-aware discriminator by taking adaptive gradient steps.
until $t > maxIter$

IV. Experiments

A. Data Description and Preprocessing

In this experiment, there are four stages associated with cognitive disease degeneration: normal control (NC), significant memory concern (SMC), early mild cognitive impairment (EMCI), and late mild cognitive impairment (LMCI). Based on the Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu/) dataset, we downloaded about 324 subjects by setting the following criteria: (1) The subject should have all the fMRI, DTI, and T1-weighted MRI; (2) the gradient directions of the DTI are set from 6 to 126. (3) The acquisition type is 3D, and the field strength is 3.0 Tesla. In the preprocessing procedures, either DTI or fMRI are combined with T1-weighted MRI to register in native space by PANDA [36] or GRETNA [37] software. Because there are two subjects suffered from preprocessing errors, we excluded these two subjects and obtained a total

TABLE I
SUMMARY OF THE SUBJECT INFORMATION IN THE EXPERIMENT

Group	NC(82)	SMC(82)	EMCI(82)	LMCI(76)
Male/Female	39M/43F	35M/47F	40M/42F	43M/33F
Age(mean±SD)	74.2±8.1	76.1±5.4	75.9±7.5	75.8±6.4

of 322 subjects, as shown in Table I. Note that SMC is the transitional stage from NC to EMCI. EMCI and LMCI are subtypes of MCI.

The DTI gradient directions are in the range $6 \sim 126$. The parameters TR (Time of Repetition) and TE (Time of Echo) are in the range of $3.4s \sim 17.5s$ and $56ms \sim 105ms$, respectively. PANDA toolbox [36] is used to perform sample preprocessing on DTI to get a fraction anisotropy (FA) image and empirical structural connectivity (A_e). To normalize individual FA images for downstream analysis, we non-linearly registered individual FA images of native space to the FA template in the MNI space by resampling into a customized spatial resolution ($2\times2\times2\ mm^3$). The FA template in the MNI space has the dimension size $91 \times 109 \times 91$. As a result, the final output FA image has a dimensional size of $91\times109\times91$, which is sent into the proposed model for structural feature extraction.

The fMRI scanned by a 3T MRI equipment has different slice thicknesses in the range of $2.5mm \sim 3.4mm$. The image resolution ranges from $2.5mm$ to $3.75mm$ in both plane dimensions. The TR value is between $0.607s$ to $3.0s$, and the TE value ranges from $30ms$ to $32ms$. The duration of scanning data is 10 minutes. In the preprocessing stage, the GRETNA toolbox [37] is utilized to acquire functional time series based on the Automated Anatomical Labelling (AAL) atlas [38]. The preprocessing steps include correcting head-motion artifacts, spatial normalization, smoothing, removing linearized drift, and band-pass filtering. Finally, the 90 non-overlapping ROIs time series are obtained by normalizing them into the same TR value. The output of this procedure is the input feature X_f of the proposed model with the dimension size 90×187.

B. Experimental Settings

In this experiment, six binary classification tasks are performed: (1) SMC vs. NC, (2) EMCI vs. NC, (3) LMCI vs. NC, (4) EMCI vs. SMC, (5) LMCI vs. SMC, and (6) LMCI vs. EMCI. Besides, we combined the EMCI and LMCI to evaluate the multi-label classification of NC vs. SMC vs. MCI. The experiments are conducted using 10-fold cross-validation to ensure the results are stable. Also, the proposed model is compared with other related methods to demonstrate its superiority. There are three methods in the comparison: (1) the empirical method that derives the structural connectivity (SC) and static functional connectivity (FC) from the commonly used software toolboxes (i.e., PANDA and GRETNA) and then averages them to obtain empirical brain networks; (2) the GBDM method that deeply fuses the SC and functional time series and then generates benchmark brain networks; (3) our method that transforms the DTI and functional time series to unified brain networks. The results of each method are sent to the same classifier (i.e., SVM [39], DNN [40], and GCN [41]) to compare their prediction performance.

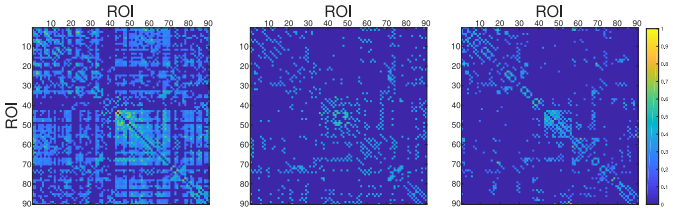

Fig. 5. An example of the generated brain networks using three methods: the empirical method (left), the GBDM method (middle), and the proposed method (right).

The model parameter settings in the experiments are: $N = 90, d = 187, q = 64, h = 32, n = 3$. TensorFlow1[1] is utilized to implement a convergent model on an NVIDIA TITAN RTX2080 GPU device for 10 hours with about 600 epochs. The initial learning rate of the transformer, encoders, and decoders is 10^{-3} and will decrease to 10^{-4} after 200 epochs. The learning rates of the generator and the conventional discriminator are set to 0.0001 and 0.0004, respectively. For the label-aware discriminator, the learning rate is set to 0.0001. The dropout ratio in both the generator and the dual discriminator is set at 0.5. The Adam is adopted to optimize the training process with batch size 16. The prediction performance is calculated by meaning the values of accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-score, and the area under the receiver operating characteristic curve (AUC). The AUC is used to evaluate the classifier's overall performance ($0 \leq AUC \leq 1$).

C. Unified Brain Network Analysis

The proposed model aims to generate unified brain networks for disease analysis. This section analyzes the generated brain networks in terms of prediction tasks. To compare the performance of our method with other related methods, six binary and one multi-label prediction task are conducted to calculate the mean values of evaluation indicators (i.e., ACC, SEN, SPE, and AUC) using a cross-validation strategy. Three different classifiers are adopted to evaluate the prediction performance of the generated brain networks. Fig. 5 shows a qualitative example of three brain networks using different methods. The quantitative prediction results are displayed in Table II and Table III. In Table IV, the four metrics (ACC, SEN, SPE, and AUC) of NC vs. SMC vs. MCI are computed by the macro-averaging method. Specifically, macro averaging is to calculate the results of each category independently and averages the results of each category, which is suitable for situations where the performance of each category is equally important to the overall performance. The best performance is marked in bold black. The results show that our method achieves the best prediction performance among the three methods in terms of different classifiers. It indicates that the proposed model can make full use of the common and complementary information from the structural-functional data and thus make the fusion more effective.

[1] http://www.tensorflow.org/

TABLE II
MEAN PREDICTION RESULTS OF THE GENERATED BRAIN NETWORKS BY DIFFERENT CLASSIFIERS FOR SMC VS. NC, EMCI VS. NC, AND LMCI VS. NC.(%)

Classifier	Methods	SMC vs. NC				EMCI vs. NC				LMCI vs. NC			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
SVM	Empirical	76.82	85.36	68.29	77.84	78.05	73.17	82.93	75.04	82.27	82.89	81.70	83.27
	GBDM	77.43	85.36	69.51	78.84	79.88	73.17	86.59	79.95	85.44	84.21	86.59	84.53
	Ours	**79.88**	**89.02**	**70.73**	**87.09**	**84.16**	**74.39**	**93.90**	**95.05**	**87.97**	**88.16**	**87.80**	**95.31**
DNN [40]	Empirical	76.83	81.71	71.95	77.54	81.70	85.36	78.04	86.85	84.17	78.94	89.02	82.50
	GBDM	78.66	81.71	75.61	78.42	83.53	86.58	80.48	88.17	86.07	82.89	89.02	86.26
	Ours	**82.92**	**84.15**	**81.71**	**86.92**	**89.02**	**87.80**	**90.24**	**96.06**	**91.77**	**88.16**	**95.12**	**94.74**
GCN [41]	Empirical	79.26	87.80	70.73	78.77	84.75	89.02	80.48	86.91	88.60	85.52	91.46	88.17
	GBDM	81.70	89.02	74.39	81.81	86.58	89.02	84.14	89.75	91.13	86.84	95.12	89.68
	Ours	**85.97**	**96.34**	**75.61**	**93.88**	**90.85**	**93.90**	**87.80**	**97.85**	**94.30**	**93.42**	95.12	**98.12**

TABLE III
MEAN PREDICTION RESULTS OF THE GENERATED BRAIN NETWORKS BY DIFFERENT CLASSIFIERS FOR EMCI VS. SMC, LMCI VS. SMC, AND LMCI VS. EMCI.(%)

Classifier	Methods	EMCI vs. SMC				LMCI vs. SMC				LMCI vs. EMCI			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
SVM	Empirical	77.44	80.49	74.39	78.48	83.54	85.53	81.71	84.74	80.38	82.89	78.05	81.98
	GBDM	78.66	79.27	78.05	79.16	86.71	85.53	**87.80**	85.93	83.54	81.58	**85.37**	85.40
	Ours	**82.93**	**87.80**	78.05	**87.46**	**88.61**	**90.79**	86.59	**91.70**	**85.44**	**88.16**	82.93	**90.47**
DNN [40]	Empirical	79.88	82.93	76.83	80.73	85.44	81.58	89.02	82.99	82.91	84.21	81.71	85.00
	GBDM	82.32	85.37	79.27	80.52	89.24	86.84	91.46	90.60	86.08	88.16	84.15	87.66
	Ours	**85.37**	**87.80**	**82.93**	**84.98**	**92.41**	**93.42**	91.46	**93.50**	**90.51**	**89.47**	**91.46**	**92.04**
GCN [41]	Empirical	81.71	84.15	79.27	79.85	89.24	90.79	87.80	91.00	85.44	82.89	87.80	86.39
	GBDM	85.98	87.80	84.15	88.22	93.04	93.42	92.68	92.78	88.61	89.47	87.80	88.29
	Ours	**87.80**	**91.46**	84.15	**91.52**	**95.57**	**94.74**	**96.34**	**95.72**	**91.14**	**93.42**	**89.02**	**93.24**

TABLE IV
MEAN PREDICTION RESULTS OF THE GENERATED BRAIN NETWORKS BY DIFFERENT CLASSIFIERS FOR A MULTI-LABEL CLASSIFICATION TASK.(%)

Classifier	Methods	(EMCI+LMCI) vs. SMC vs. NC			
		ACC	SEN	SPE	AUC
SVM	Empirical	70.81	56.37	77.52	70.32
	GBDM	74.74	62.53	80.55	73.55
	Ours	**76.19**	**64.98**	**81.71**	**76.04**
DNN [40]	Empirical	72.05	58.61	78.80	74.62
	GBDM	75.78	64.37	81.43	78.17
	Ours	**78.05**	**67.08**	**83.09**	**79.08**
GCN [41]	Empirical	73.71	60.69	79.72	74.78
	GBDM	78.05	67.66	83.22	**79.17**
	Ours	**79.30**	**70.10**	**84.37**	77.71

To analyze the effect of different brain regions on the prediction task, we first shield one brain region of the unified brain network and compute the corresponding mean accuracy, and then obtain this brain region's importance score by subtracting the computed accuracy from value 1. By repeating this procedure for each ROI, we get the importance score for all 90 ROIs. After sorting the importance scores in descending order, the top 10 corresponding ROIs are the important brain regions for the classification task. The top 10 disease-related ROIs are displayed in Fig. 6. Specifically, the top 10 related ROIs are PHG.L, CAL.R, DCG.L, PCUN.R, THA.L, ORBinf.R, AMYG.L, OLF.R, SOG.R, and FFG.L in the SMC vs. NC prediction task. For EMCI vs. NC, the ten important ROIs are in the frontal lobe (SFGdor.L, ORBinf.R, OLF.L, SFGmed.R), temporal lobe (AMYG.L, TPOsup.L), parietal lobe (SPG.L, PCL.R), and subcortical area (PAL.R). The relevant brain ROIs for LMCI vs. NC are in the frontal lobe (ORBsupmed.L, SFGmed.R, ORBinf.R, OLF.R), parietal lobe (SPG.R, PCG.L), temporal lobe (TPOmid.R, PHG.R, TPOsup.L), and subcortical area (PUT.R). From SMC to EMCI, the ten important ROIs are PHG.R, ACG.R, PCG.R, PoCG.R, PreCG.R,ORBmid.L, OLF.L, ORBinf.L, AMYG.L, and PoCG.L. For LMCI vs. SMC, the SMA.L, PoCG.R, PUT.R, PreCG.L, ORBinf.R, AMYG.L, HIP.L, PoCG.L, TPOsup.R, and ACG.R are the important brain regions. From EMCI to LMCI, the top 10 ROIs are CAL.R, DCG.R, ANG.R, ITG.R, ORBinf.L, SOG.R, PHG.L, HIP.L, PHG.R, and SMA.L.

D. Prediction of Abnormal Brain Connections

In this section, predicting abnormal brain connections is studied during the cognitive disease's progression. After the model has been trained to converge, it can output the unified brain network (UBN) for each subject with bimodal images (i.e., fMRI and DTI). Based on the standard two-sample t-test method, we can evaluate the significant brain connections between two groups (e.g., LMCI vs. NC) by setting the p-value threshold. In Fig. 7, the first and second columns show three examples of p-values on brain connections between patients (i.e., LMCI, EMCI, and SMC) and NC groups by setting the threshold lower than 0.05. For ease of visualization, the significant brain connections are denoted by a blue color.

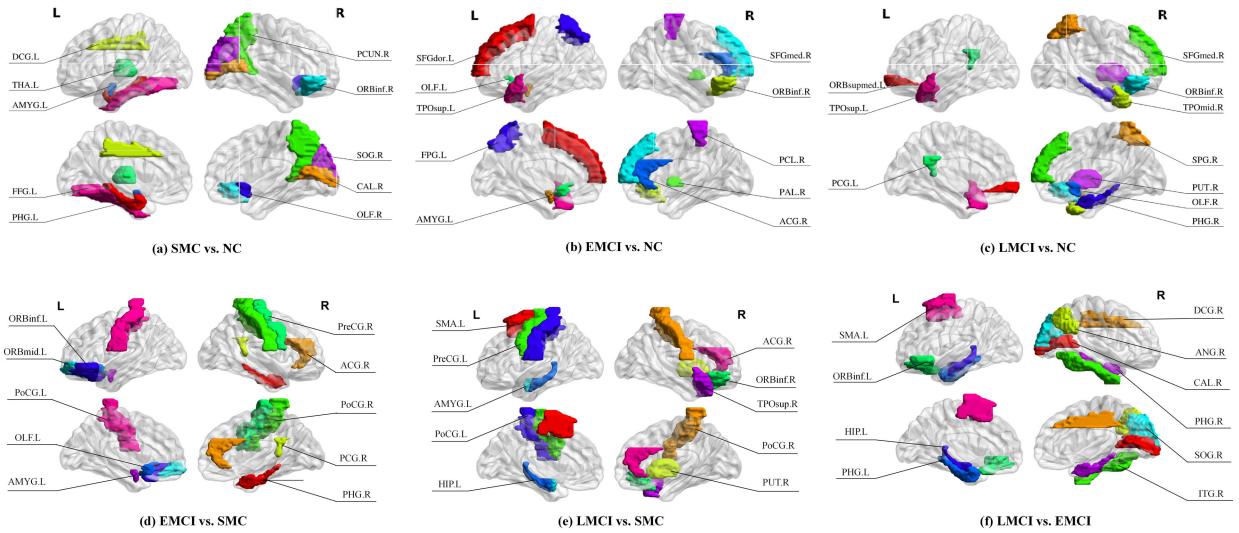

Fig. 6. The spatial distribution of the top 10 related ROIs for (a) SMC vs. NC, (b) EMCI vs. NC, (c) LMCI vs. NC, (d) EMCI vs. SMC, (e) LMCI vs. SMC, and (f) LMCI vs. EMCI.

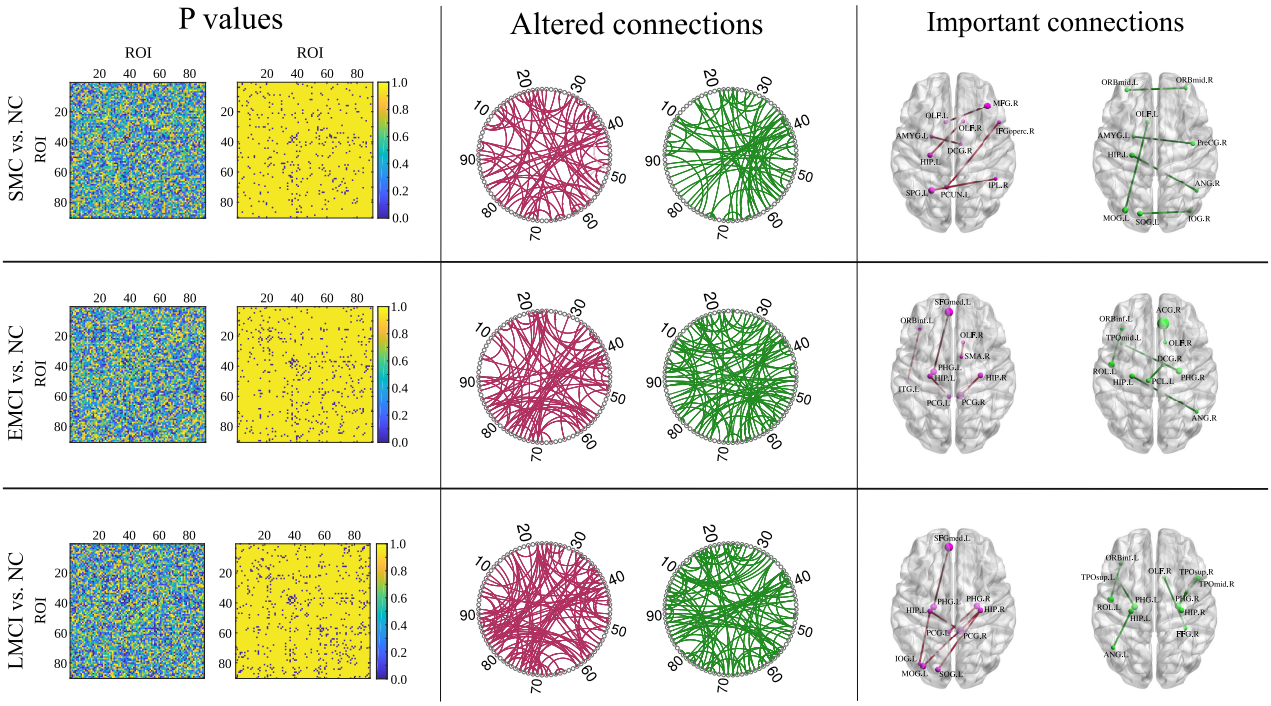

Fig. 7. The altered brain connections for SMC vs. NC, EMCI vs. NC, and LMCI vs. NC, respectively. The first column shows the p-values of the connections between all paired ROIs, the second column displays the significant connections with p-value < 0.05, the third and fourth columns represent reduced and increased connections, the fifth and sixth columns are the spatial visualization of important connections.

The most densely connected ROIs are mostly overlapped with the results of the above section. These significant brain connections are defined as the altered connections or the abnormal connections. To investigate the property of these altered connections, we subtract the mean connection strength of each patient group (SMC, EMCI, LMCI) from that of the NC group. The positive values mean increased connections, and the negative values mean reduced connections. The third and fourth columns display the circular graph of altered connections at different disease stages. Compared with the normal controls, the number of reduced connections is 154, 162, and 218, while the number of increased connections is 138, 192, and 233 for SMC, EMCI, and LMCI, respectively.

To analyze the important abnormal connections, significant connections with a p-value lower than 0.001 are selected. The spatial location of the important abnormal connections is shown in the fifth and sixth columns of Fig. 7 using the BrainNet Viewer toolbox [43]. The red color means reduced connections, and the green color indicates increased connections. The ROI size defined in the AAL template is the relative volume.

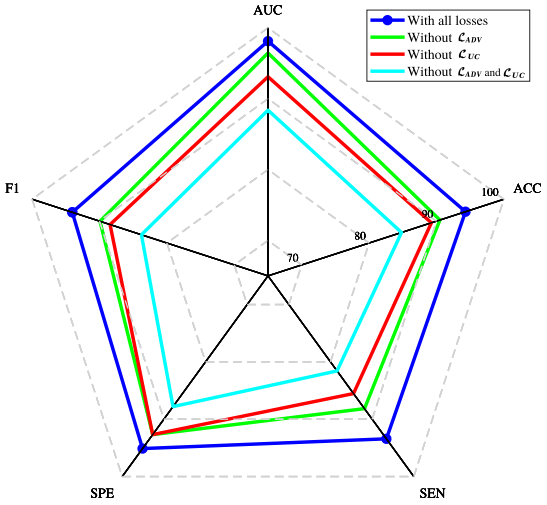

Fig. 8. Comparison of different loss functions on the prediction performance of LMCI vs. NC.

E. Ablation Study

The decomposition-fusion framework in the proposed model is essential for constructing unified brain networks. To investigate the effectiveness of the decomposed and fused modules, the uniform-unique contrastive and adversarial losses are explored in the prediction performance. Three prediction experiments are conducted in this section: 1) remove the uniform-unique contrastive loss $\mathcal{L}_{UC}$, which means the unique and uniform representations are mixed; 2) remove the adversarial loss $\mathcal{L}_{ADV}$; 3) remove both losses in the proposed model. The effects of removing different loss functions on prediction performance are shown in Fig. 8. The results demonstrate that either uniform-unique contrastive loss or adversarial loss can affect the prediction performance to some extent. Both loss functions can effectively improve the proposed model's performance in terms of ACC, SEN, SPE, F1, and AUC.

V. Discussion

A. Effectiveness of the Generator and the Decoders

In this study, the representation-fusing generator is important for unified brain network construction and analysis. To evaluate if the predictive UBN obtained by the generator is disease-related, the t-distributed stochastic neighbor embedding (t-SNE) tool [44] is used to display how the learned representations are arranged and if they are well separated. Fig. 9 shows the two-dimensional projection of the learned representations without and with the representation-fusing generator for NC vs. LMCI. The representations obtained by BSFL with the generator are well arranged by class information, while the ones obtained by BSFL without the generator are scattered and not badly separated. Thus, the BSFL can extract MCI-related features and capture complementary information between modalities.

The structural and functional decoders are used to reconstruct the empirical SC and functional time series from the decomposed representations. And these representations are fused to generate unified brain networks for disease analysis.

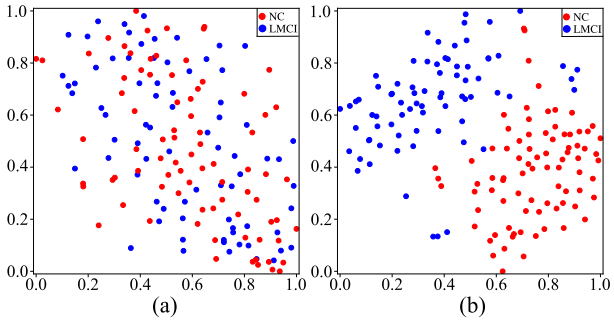

Fig. 9. The t-SNE visualization of the representations obtained by BSFL (a) without and (b) with the representation-fusing generator.

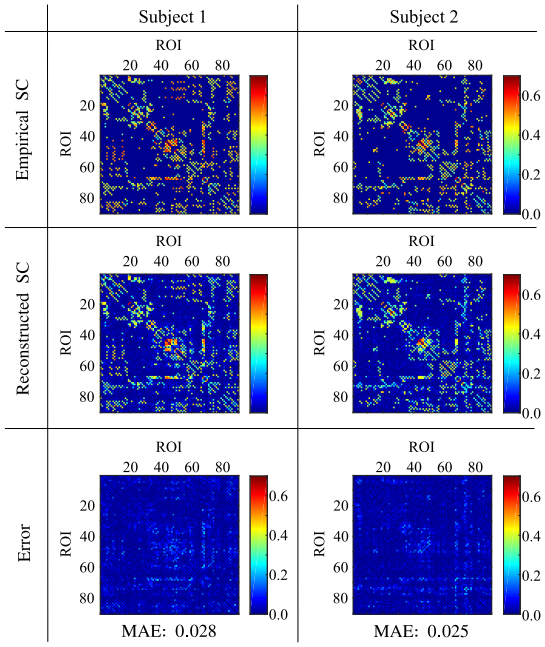

Fig. 10. Visualization of the empirical SC, the reconstructed SC, and the reconstructed error from two representative subjects.

Therefore, the reconstruction process greatly influences the fusion effect and downstream tasks. To demonstrate the reconstruction performance, two representative subjects are selected from LMCI and NC, respectively, and the reconstructed SC is visualized in Fig. 10. The structural decoder can well rebuild the empirical SC. The mean absolute error (MAE) is used to evaluate the reconstruction quality.

B. Comparison With Clinical Studies

Effective disease-related biomarkers are essential for clinicians to early diagnose neurodegenerative disease and develop treatments to delay disease progression. The proposed model can output important ROIs and potential biomarkers in MCI analysis. In this section, comparing our results with clinical studies will be investigated. Considering the prediction tasks and the two-sample t-test, 10 top overlapping ROIs (ORBinf.R, OLF.L, OLF.R, SFGmed.R, PHG.R, AMYG.L, SOG.R, FFG.L, PCL.R, and PUT.R) can be obtained in the identified brain regions in Table V. These ROIs are highly correlated with MCI and can be found in previous studies [45], [46]. For example, the olfactory cortex plays an important

TABLE V
THE TOP TEN MCI-RELATED ROIS WERE VERIFIED BY THE CLINICAL STUDIES

ROI index	ROI name	Location	Verification
16	ORBinf.R	Frontal lobe	Hu *et al.* [47]
21	OLF.L	Frontal lobe	Li *et al.* [48]
22	OLF.R	Frontal lobe	Li *et al.* [48]
24	SFGmed.R	Frontal lobe	Yu *et al.* [49]
40	PHG.R	Temporal lobe	Duan *et al.* [50]
41	AMYG.L	Temporal lobe	Mihaescu *et al.* [51]
50	SOG.R	Occipital lobe	Wang *et al.* [52]
55	FFG.L	Temporal lobe	Hu *et al.* [47]
70	PCL.R	Parietal lobe	Tang *et al.* [53]
74	PUT.R	Subcortical area	Wang *et al.* [54]

role in translating everyday experiences into lasting episodic and working memory, which is preferentially attacked at the early stage of Alzheimer's disease. The parahippocampal gyrus also contributes to memory storage, and its structural damage can cause abnormal emotional and cognitive behavior. The paracentral lobule can recognize spatial relationships, and patients with MCI showed atrophy in the parietal lobe.

In the prediction of abnormal connections, the number of reduced connections increases from SMC to LMCI when compared with NC, while the number of increased connections also shows the same pattern. It indicates that these disease stages behave as a compensatory mechanism [55] to make sure the brain functions normally. Besides, the important abnormal connections obtained partly agree with clinical findings. For example, some increased brain connections have been identified by clinical discoveries [56], including HIP.L-ANG.L, HIP.L-ANG.R, PHG.R-TPOsup.R, and PHG.R-TPOmid.R. And some reduced connections have been verified between the left posterior cingulate gyrus and the left hippocampus, and between the right hippocampus and the left middle occipital gyrus. Furthermore, the changes in brain connection strength increase or decrease dramatically as the disease progresses. The increased connections likely appear on the same brain hemisphere, while the reduced connections are found across the brain hemispheres. It may be illustrated by the previous work that the increased connections have shorter distances compared with the reduced connections [57].

VI. CONCLUSION

In this paper, the proposed BSFL model is proposed to predict brain network abnormalities during MCI progression by combining fMRI and DTI. With the guidance of prior knowledge, the designed transformer can automatically extract the local and global connectivity features throughout the brain. The decomposed variational graph autoencoders decompose the feature space into unique and uniform spaces, and the uniform-unique contrastive loss is utilized to further improve the decomposition's effectiveness. The representation-fusing generator is utilized to fuse the decomposed representations to generate MCI-related connectivity features. The extensive experiments on the ADNI database demonstrate the proposed model's effectiveness compared with other competitive methods. Furthermore, some MCI-related brain regions and abnormal connections identified in our results also show the proposed model's reliability. Altogether, the proposed model is promising for reconstructing unified brain networks for brain disease analysis and providing potential connection-based biomarkers during the degenerative process of MCI.

REFERENCES

[1] Alzheimer's Association, "2019 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 15, no. 3, pp. 321–387, Mar. 2019.

[2] M. N. Sabbagh et al., "Rationale for early diagnosis of mild cognitive impairment (MCI) supported by emerging digital technologies," *J. Prevention Alzheimer's Disease*, vol. 7, no. 3, pp. 158–164, 2020.

[3] E. C. Edmonds et al., "Data-driven vs consensus diagnosis of MCI: Enhanced sensitivity for detection of clinical, biomarker, and neuropathologic outcomes," *Neurology*, vol. 97, no. 13, pp. e1288–e1299, 2021.

[4] B. Lei et al., "Predicting clinical scores for Alzheimer's disease based on joint and deep learning," *Expert Syst. Appl.*, vol. 187, Jan. 2022, Art. no. 115966.

[5] M. Li et al., "A retinal vessel segmentation method based on super-pixel and generative adversarial networks," *J. Integr. Technol.*, vol. 9, no. 6, pp. 21–28, 2020.

[6] C. Gong et al., "Generative AI for brain image computing and brain network computing: A review," *Frontiers Neurosci.*, vol. 17, Jun. 2023, Art. no. 1203104.

[7] J. B. Pereira et al., "Abnormal structural brain connectome in individuals with preclinical Alzheimer's disease," *Cerebral Cortex*, vol. 28, no. 10, pp. 3638–3649, Oct. 2018.

[8] S. Hu, W. Yu, Z. Chen, and S. Wang, "Medical image reconstruction using generative adversarial network for Alzheimer disease assessment with class-imbalance problem," in *Proc. IEEE 6th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2020, pp. 1323–1327.

[9] W. Yu, B. Lei, M. K. Ng, A. C. Cheung, Y. Shen, and S. Wang, "Tensorizing GAN with high-order pooling for Alzheimer's disease assessment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4945–4959, Sep. 2022.

[10] W. Yu et al., "Morphological feature visualization of Alzheimer's disease via multidirectional perception GAN," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 4401–4415, Aug. 2023.

[11] S. Yu et al., "Multi-scale enhanced graph convolutional network for early mild cognitive impairment detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2020, pp. 228–237.

[12] J. Pan, B. Lei, Y. Shen, Y. Liu, Z. Feng, and S. Wang, "Characterization multimodal connectivity of brain network by hypergraph GAN for Alzheimer's disease analysis," in *Proc. 4th Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*, Beijing, China. Springer, Oct./Nov. 2021, pp. 467–478.

[13] Q. Zuo, B. Lei, Y. Shen, Y. Liu, Z. Feng, and S. Wang, "Multimodal representations learning and adversarial hypergraph fusion for early Alzheimer's disease prediction," in *Proc. 4th Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*, Beijing, China. Springer, Oct./Nov. 2021, pp. 479–490.

[14] S. Hu, B. Lei, S. Wang, Y. Wang, Z. Feng, and Y. Shen, "Bidirectional mapping generative adversarial networks for brain MR to PET synthesis," *IEEE Trans. Med. Imag.*, vol. 41, no. 1, pp. 145–157, Jan. 2022.

[15] S. You et al., "Fine perceptive GANs for brain MR image super-resolution in wavelet domain," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 7, 2022, doi: 10.1109/TNNLS.2022.3153088.

[16] B. Hu, C. Zhan, B. Tang, B. Wang, B. Lei, and S.-Q. Wang, "3-D brain reconstruction by hierarchical shape-perception network from a single incomplete image," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 11, 2023, doi: 10.1109/TNNLS.2023.3266819.

[17] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.

[18] Q. Zhang, and Y. B. Yang, "Rest: An efficient transformer for visual recognition," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021.

[19] T. Xiao, P. Dollar, M. Singh, E. Mintun, T. Darrell, and R. Girshick, "Early convolutions help transformers see better," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021.

[20] D. Hu et al., "Disentangled-multimodal adversarial autoencoder: Application to infant age prediction with incomplete multimodal neuroimages," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 4137–4149, Dec. 2020.

[21] H. Yue et al., "MLDRL: Multi-loss disentangled representation learning for predicting esophageal cancer response to neoadjuvant chemoradiotherapy using longitudinal CT images," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102423.

[22] J. B. Pereira et al., "Disrupted network topology in patients with stable and progressive mild cognitive impairment and Alzheimer's disease," *Cerebral Cortex*, vol. 26, no. 8, pp. 3476–3493, Aug. 2016.

[23] T. Wang et al., "Multilevel deficiency of white matter connectivity networks in Alzheimer's disease: A diffusion MRI study with DTI and HARDI models," *Neural Plasticity*, vol. 2016, pp. 1–14, 2016.

[24] M. M. Engels, C. J. Stam, W. M. van der Flier, P. Scheltens, H. de Waal, and E. C. van Straaten, "Declining functional connectivity and changing hub locations in Alzheimer's disease: An EEG study," *BMC Neurol.*, vol. 15, no. 1, pp. 1–8, Dec. 2015.

[25] B. Jie, C.-Y. Wee, D. Shen, and D. Zhang, "Hyper-connectivity of functional networks for brain disease diagnosis," *Med. Image Anal.*, vol. 32, pp. 84–100, Aug. 2016.

[26] B. Lei et al., "Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease," *Med. Image Anal.*, vol. 61, Apr. 2020, Art. no. 101652.

[27] T. Zhou, K. Thung, X. Zhu, and D. Shen, "Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis," *Hum. Brain Mapping*, vol. 40, no. 3, pp. 1001–1016, Feb. 2019.

[28] P. K. Gyawali, Z. Li, S. Ghimire, and L. Wang, "Semi-supervised learning by disentangling and self-ensembling over stochastic latent space," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2019, pp. 766–774.

[29] J. Jiang, and H. Veeraraghavan, "Unified cross-modality feature disentangler for unsupervised multi-domain MRI abdomen organs segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2020, pp. 347–358.

[30] Y. Zhang, L. Zhan, S. Wu, P. Thompson, and H. Huang, "Disentangled and proportional representation learning for multi-view brain connectomes," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2021, pp. 508–518.

[31] J. Cheng et al., "Multimodal disentangled variational autoencoder with game theoretic interpretability for glioma grading," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 2, pp. 673–684, Feb. 2022.

[32] K. Hara, H. Kataoka, and Y. Satoh, "Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and ImageNet?" in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6546–6555.

[33] S. Wang, H. Wang, A. C. Cheung, Y. Shen, and M. Gan, "Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease," *Deep Learn. Appl.*, pp. 53–73, 2020.

[34] L. Zhang et al., "Deep fusion of brain structure-function in mild cognitive impairment," *Med. Image Anal.*, vol. 72, Aug. 2021, Art. no. 102082.

[35] Z. Ding et al., "Guided variational autoencoder for disentanglement learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7917–7926.

[36] Z. Cui, S. Zhong, P. Xu, Y. He, and G. Gong, "PANDA: A pipeline toolbox for analyzing brain diffusion images," *Frontiers Hum. Neurosci.*, vol. 7, p. 42, 2013.

[37] J. Wang, X. Wang, M. Xia, X. Liao, A. Evans, and Y. He, "GRETNA: A graph theoretical network analysis toolbox for imaging connectomics," *Frontiers Hum. Neurosci.*, vol. 9, p. 386, Jun. 2015.

[38] N. Tzourio-Mazoyer et al., "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, Jan. 2002.

[39] C. Cortes and V. Vapnik, "Support vector machine," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[40] Y. Kong, J. Gao, Y. Xu, Y. Pan, J. Wang, and J. Liu, "Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier," *Neurocomputing*, vol. 324, pp. 63–68, Jan. 2019.

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

[42] M. Krzywinski et al., "Circos: An information aesthetic for comparative genomics," *Genome Res.*, vol. 19, no. 9, pp. 1639–1645, Sep. 2009.

[43] M. Xia, J. Wang, and Y. He, "BrainNet viewer: A network visualization tool for human brain connectomics," *PLoS ONE*, vol. 8, no. 7, Jul. 2013, Art. no. e68910.

[44] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, 2008.

[45] L. Xu et al., "Prediction of progressive mild cognitive impairment by multi-modal neuroimaging biomarkers," *J. Alzheimer's Disease*, vol. 51, no. 4, pp. 1045–1056, Apr. 2016.

[46] T. Kawagoe, K. Onoda, and S. Yamaguchi, "Subjective memory complaints are associated with altered resting-state functional connectivity but not structural atrophy," *NeuroImage, Clin.*, vol. 21, 2019, Art. no. 101675.

[47] Q. Hu et al., "Intrinsic brain activity alterations in patients with mild cognitive impairment-to-normal reversion: A resting-state functional magnetic resonance imaging study from voxel to whole-brain level," *Frontiers Aging Neurosci.*, vol. 13, Jan. 2022.

[48] Y. Li et al., "Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features," *Neurobiol. Aging*, vol. 33, no. 2, pp. 427.e15–427.e30, Feb. 2012.

[49] M. Yu et al., "Selective impairment of hippocampus and posterior hub areas in Alzheimer's disease: An MEG-based multiplex network study," *Brain*, vol. 140, no. 5, pp. 1466–1485, May 2017.

[50] H. Duan et al., "Differences in Aβ brain networks in Alzheimer's disease and healthy controls," *Brain Res.*, vol. 1655, pp. 77–89, Jan. 2017.

[51] A. S. Mihaescu et al., "Graph theory analysis of the dopamine D2 receptor network in Parkinson's disease patients with cognitive decline," *J. Neurosci. Res.*, vol. 99, no. 3, pp. 947–965, Mar. 2021.

[52] B. Wang et al., "Decreased complexity in Alzheimer's disease: Resting-state fMRI evidence of brain entropy mapping," *Frontiers Aging Neurosci.*, vol. 9, p. 378, Nov. 2017.

[53] J. Tang et al., "Aberrant white matter networks mediate cognitive impairment in patients with silent lacunar infarcts in basal ganglia territory," *J. Cerebral Blood Flow Metabolism*, vol. 35, no. 9, pp. 1426–1434, Sep. 2015.

[54] D. Wang et al., "Topological disruption of structural brain networks in patients with cognitive impairment following cerebellar infarction," *Frontiers Neurol.*, vol. 10, p. 759, Jul. 2019.

[55] M. Montembeault, I. Rouleau, J.-S. Provost, and S. M. Brambati, "Altered gray matter structural covariance networks in early stages of Alzheimer's disease," *Cerebral Cortex*, vol. 26, no. 6, pp. 2650–2662, Jun. 2016.

[56] D. Berron, D. van Westen, R. Ossenkoppele, O. Strandberg, and O. Hansson, "Medial temporal lobe connectivity and its associations with cognition in early Alzheimer's disease," *Brain*, vol. 143, no. 4, pp. 1233–1248, Apr. 2020.

[57] Y. Liu et al., "Impaired long distance functional connectivity and weighted network architecture in Alzheimer's disease," *Cerebral Cortex*, vol. 24, no. 6, pp. 1422–1435, Jun. 2014.