



ARTICLE

## Brain Functional Network Generation Using Distribution-Regularized Adversarial Graph Autoencoder with Transformer for Dementia Diagnosis

Qiankun Zuo<sup>1,4</sup>, Junhua Hu<sup>2</sup>, Yudong Zhang<sup>3,\*</sup>, Junren Pan<sup>4</sup>, Changhong Jing<sup>4</sup>, Xuhang Chen<sup>5</sup>, Xiaobo Meng<sup>6</sup> and Jin Hong<sup>7,8,\*</sup>

<sup>1</sup>School of Information Engineering, Hubei University of Economics, Wuhan, 430205, China

<sup>2</sup>State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing, 100038, China

<sup>3</sup>School of Computing and Mathematic Sciences, University of Leicester, Leicester, LE1 7RH, UK

<sup>4</sup>Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, 518055, China

<sup>5</sup>Faculty of Science and Technology, University of Macau, Macau, 999078, China

<sup>6</sup>School of Geophysics, Chengdu University of Technology, Chengdu, 610059, China

<sup>7</sup>Laboratory of Artificial Intelligence and 3D Technologies for Cardiovascular Diseases, Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 519041, China

<sup>8</sup>Medical Research Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 519041, China

\*Corresponding Authors: Yudong Zhang. Email: yz461@le.ac.uk; Jin Hong. Email: hongj5@mail2.sysu.edu.cn

Received: 04 January 2023 Accepted: 20 March 2023 Published: 03 August 2023

### ABSTRACT

The topological connectivity information derived from the brain functional network can bring new insights for diagnosing and analyzing dementia disorders. The brain functional network is suitable to bridge the correlation between abnormal connectivities and dementia disorders. However, it is challenging to access considerable amounts of brain functional network data, which hinders the widespread application of data-driven models in dementia diagnosis. In this study, a novel distribution-regularized adversarial graph auto-Encoder (DAGAE) with transformer is proposed to generate new fake brain functional networks to augment the brain functional network dataset, improving the dementia diagnosis accuracy of data-driven models. Specifically, the label distribution is estimated to regularize the latent space learned by the graph encoder, which can make the learning process stable and the learned representation robust. Also, the transformer generator is devised to map the node representations into node-to-node connections by exploring the long-term dependence of highly-correlated distant brain regions. The typical topological properties and discriminative features can be preserved entirely. Furthermore, the generated brain functional networks improve the prediction performance using different classifiers, which can be applied to analyze other cognitive diseases. Attempts on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset demonstrate that the proposed model can generate good brain functional networks. The classification results show adding generated data can achieve the best accuracy value of 85.33%, sensitivity value of 84.00%, specificity value of 86.67%. The proposed model also achieves superior performance compared with other related augmented models. Overall, the proposed model effectively improves cognitive disease diagnosis by generating diverse brain functional networks.



**KEYWORDS**

Adversarial graph encoder; label distribution; generative transformer; functional brain connectivity; graph convolutional network; dementia

**1 Introduction**

The brain is an information-processing system with many complicated and precise computations when dealing with various daily activities [1]. Daily physiological activities are always associated with the interaction between multiple neurons, neuronal clusters, or multiple brain regions. This interaction is called brain functional network (BFN), which describes the relationship between temporal blood-oxygen-level-dependent (BOLD) signals from distant brain areas [2]. Dementia (i.e., Alzheimer's Disease, AD) is a typical kind of neuropathic disorder where patients usually show abnormal functional connections between brain regions [3]. It can result in a series of cognitive symptoms: memory impairment, poor language expression, and changes in vision [4,5]. Functional magnetic resonance imaging (fMRI) can easily capture these abnormal features using the non-intrusive scanning technology [6]. The BFN can bring a new way for the diagnosis and analysis of neurodegenerative disorders [7]. Therefore, analysis of BFNs is important to mine complex brain connectivity features and help detect dementia-related biomarkers. It is further important to understand the pathogenic mechanism and the drug discovery for neurodegenerative disorders [8].

The BFN is usually constructed through a software toolbox by splitting the human brain into predefined Region-of-Interests (ROIs) [9]. The element in the BFN matrix indicates the functional correlation between two ROIs. Many approaches based on machine learning were utilized to diagnose neurological disease in an end-to-end scheme [10–16]. For example, Meier et al. [17] applied the support vector machine (SVM) classifier to distinguish older adults from younger adults using functional connectivity data. To boost the Mild Cognitive Impairment (MCI) prediction performance, Yu et al. [18] designed a sparse graph representation learning method with a weighting scheme to generate sparse BFNs. Bi et al. [19] combined convolutional learning and recurrent learning to extract regional connectivity and adjacent positional features, which proves its learning ability in AD diagnosis. More advanced techniques are proposed to explore the complex connectivity-based features [20–22]. Ji et al. [23] devised novel convolutional kernels to capture hierarchical topological characteristics by element-wise weighting brain networks and achieved more accurate classification performance. The work in [24] applied the graph convolutional network (GCN) method to improve the classification accuracy by jointly using the functional time series and connectivity-based matrices. Nevertheless, the limited medical data makes data-driven models challenging to achieve good prediction results.

The straightforward way to improve classification performance is to synthesize more similar medical data and feed it to data-driven models [25,26]. Numerous data-augmented methods have been developed to solve the small data problem in the field of brain imaging analysis. For example, Hong et al. [27] augmented the routine brain magnetic resonance (MR) imaging with scaling, rotation, translation, and gamma correction and achieved accurate results predicting children's brain age through a deep learning model. Hu et al. [28] synthesized the positron emission tomography (PET) from MR using generative models, which can handle the problem of incomplete modalities and is promising for multimodal fusion. However, the above methods cannot be applied to BFN augmentation, because it only considers the local features between adjacent pixels and ignores the

topological information between distant pixels. Many efforts have been tried to generate new graph data in the graph domain. Meszlényi et al. [29] created some simulated connectivity-based datasets by applying noise weights (NW) to improve the classification performance. The study of [30] solved the problem of small-size data by employing the synthetic minority over-sampling technique (SMOTE) algorithm and achieved a good classification accuracy of non-tumorous facial pigmentation disorders. However, these methods do not directly generate new graph data but interpolate existing brain networks to augment the data, which brings some noise and may have some side effects on the model's classification performance.

The generative adversarial networks (GANs) [31] is a two-player game that can produce quite good results by mutual game learning [32]. It has gained broad applications in analyzing medical imaging because of its strong ability in distribution fitting [33]. These applications cover the image-related fields, including cross-modal synthesis [34], point cloud generation [35], image super-resolution [36], disease classification [37–39], regression task [40,41], and organ segmentation [42]. Besides, the prior distribution can guide the model's optimization and thus stabilize the representation learning in the GAN's training [43]. Reference [44] introduced a Gaussian distribution to constrain the graph embedding in adversarial learning and achieved good performance in graph analytics. The GAN-based model has been applied in the BFN augmentation. For example, Tan et al. [45] utilized the Gaussian noise to synthesize BFNs by applying a semi-positive definite manifold constraint. Also, the transformer network [46] can greatly improve image classification performance by combining adversarial strategy, which can model a strong relationship between distant ROIs.

Motivated by these observations, in this study, a novel distribution-regularized adversarial graph autoencoder (DAGAE) model is proposed to generate BFNs for dementia diagnosis. The main works of this paper are as follows: (1) The label distribution is estimated to regularize the latent space learned by the graph encoder, which can make the learning process stable and deduce a robust representation. (2) The transformer-based network in the generator is introduced to map the node representations into node-to-node connections by exploring the global connectivity information between distant ROIs. It preserves the main topological properties and more discriminative features. (3) The generated BFNs enhance disease prediction using different classifiers, which can be applied to analyze other related brain diseases.

## 2 Materials and Methods

### 2.1 Data Preparation

The purpose of the Alzheimer's Disease Neuroimaging Initiative (ADNI) project<sup>1</sup> is to detect the early stage of Alzheimer's disease from clinical, imaging, gene, biomarker, and other aspects. In this study, we mainly focus on the Late Mild Cognitive Impairment (LMCI) stage scanned with functional Magnetic Resonance Imaging (fMRI). To eliminate the influence of category imbalance on model classification performance, we selected the same number of NC subjects as LMCI for the experiment. About 150 subjects with fMRI were selected to test our model's effectiveness, including 75 Normal Controls (NC) and 75 LMCI. The fMRI data were scanned with the filed strength of 3.0 Tesla. The turning angle is 80 degrees, and the time of repetition (TR) is in the range of 0.607~3.0 s. The scanning time for each subject is about 10 min.

The commonly used GRETNA [47] software is adopted to preprocess the fMRI to construct graph data. The detailed procedures include format conversion, first ten volumes removal, slice

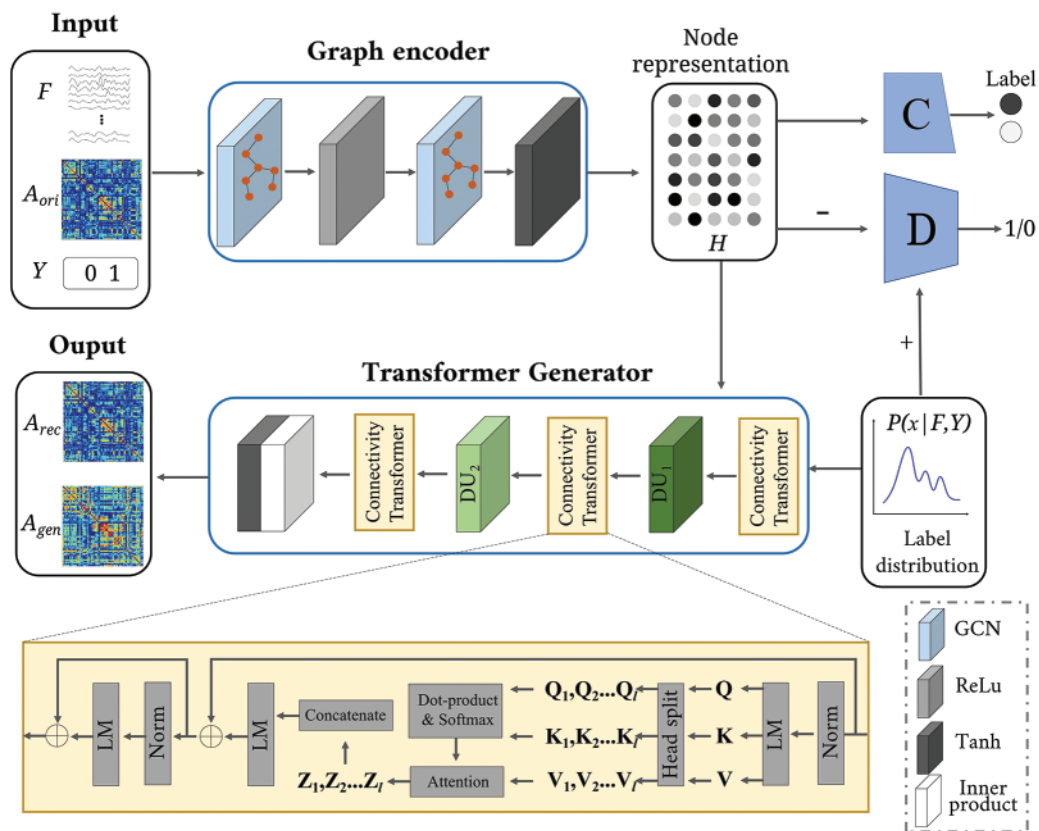
---

<sup>1</sup><http://adni.loni.usc.edu/>

timing, head motion realign, normalizing, spatially smooth, detrend, and temporally filtering (usually 0.01~0.08 Hz). At last, the automated anatomical labeling (AAL) atlas [48] with ninety non-overlapping ROIs is warped to the fMRI volumes for obtaining functional features  $F$ . The  $F$  with the size  $90 \times 187$  is transformed into a BFN  $A_{ori}$  with the dimension size  $90 \times 90$  by the Pearson coefficient algorithm.

## 2.2 Distribution-Regularized Adversarial Graph Autoencoder

The BFN is generated by the designed DAGAE model, which is depicted in Fig. 1. It accepts the graph data (including brain functional feature  $F$  and BFN  $A_{ori}$ ) and corresponding label (i.e.,  $Y = \{0, 1\}$ ), outputs the reconstructed brain network  $A_{rec}$ , and the generated brain network  $A_{gen}$ . The proposed DAGAE contains three parts: the label distribution estimation (LDE), the adversarial graph encoder (AGE), and the transformer generator (TG). The LDE module is devised to compute the probability distribution of latent node features, which can robustly constrain the graph encoder for node representation learning. The TG maps latent node space to graph space, which decodes the node representations to BFNs. Four objective functions are utilized to optimize the model, including adversarial loss, reconstruction loss, node-representation consistent loss, and cross-entropy loss.



**Figure 1:** The architecture of the proposed DAGAE model. It accepts brain functional feature  $F$  and BFN  $A_{ori}$  with a specific label and outputs reconstructed or generated BFN (i.e.,  $A_{rec}$ ,  $A_{gen}$ ). The model has four parts: Graph encoder, classifier, discriminator, and transformer generator

### 2.2.1 Label Distribution Estimation

To improve the performance of node representation in latent space, the label probability distribution estimated by Kernel Density Estimation (KDE) is introduced in the node representation learning. Instead of the traditional normal distribution, it can reflect the accurate distribution of node features and ensure robust representations in the model training. In the feature space, the node feature  $F \in \mathbb{R}^{N \times 187}$  is first passed through a dimension reduction operation to get  $H = \{h_1, h_2, \dots, h_N\} \in \mathbb{R}^{N \times p}$ , and then sent to the KDE for distribution estimation.  $N$  means the number of brain regions,  $p$  is the dimension of  $h_i$ . The estimated label distribution  $P(x|F, Y)$  is defined as:

$$P(x|F, Y) = \frac{1}{Nn_Y b} \sum_{i=1}^{Nn_Y} K\left(\frac{x - h_i}{b}\right) \quad (1)$$

where,  $n_Y$  is the subject number with specific disease (i.e.,  $Y = 0$  means the NC,  $Y = 1$  means the LMCI) subjects.  $K(\cdot)$  is a predefined kernel function (i.e., Gaussian), and  $b$  means the kernel's bandwidth.

### 2.2.2 Adversarial Graph Encoder

The graph encoder accepts both  $F$  and  $A_{ori}$  and outputs the latent node representation  $H$ . To make the node representation learning stable, a prior distribution is introduced to guide the learning process. The graph encoder  $\mathbf{E}$  consists of two GCN layers, where each layer is followed by an activation function. The output dimension of GCN layers is 64 and 32, respectively. The first and second activations are the *ReLU* and *tanh* functions, respectively. The graph encoder can be expressed as:

$$H = \mathbf{E}(A_{ori}, F) = GCN_2(GCN_1(A_{ori}, F)) \quad (2)$$

The node representation  $H$  is treated as a fake sample to send to the discriminator  $\mathbf{D}$ . The loss function of the graph encoder in adversarial training is:

$$\mathcal{L}_{enc} = \mathbb{E}_{F \sim P_{fMRI}, A_{ori} \sim P_{fMRI}} [\log(1 - \mathbf{D}(\mathbf{E}(A_{ori}, F)))] \quad (3)$$

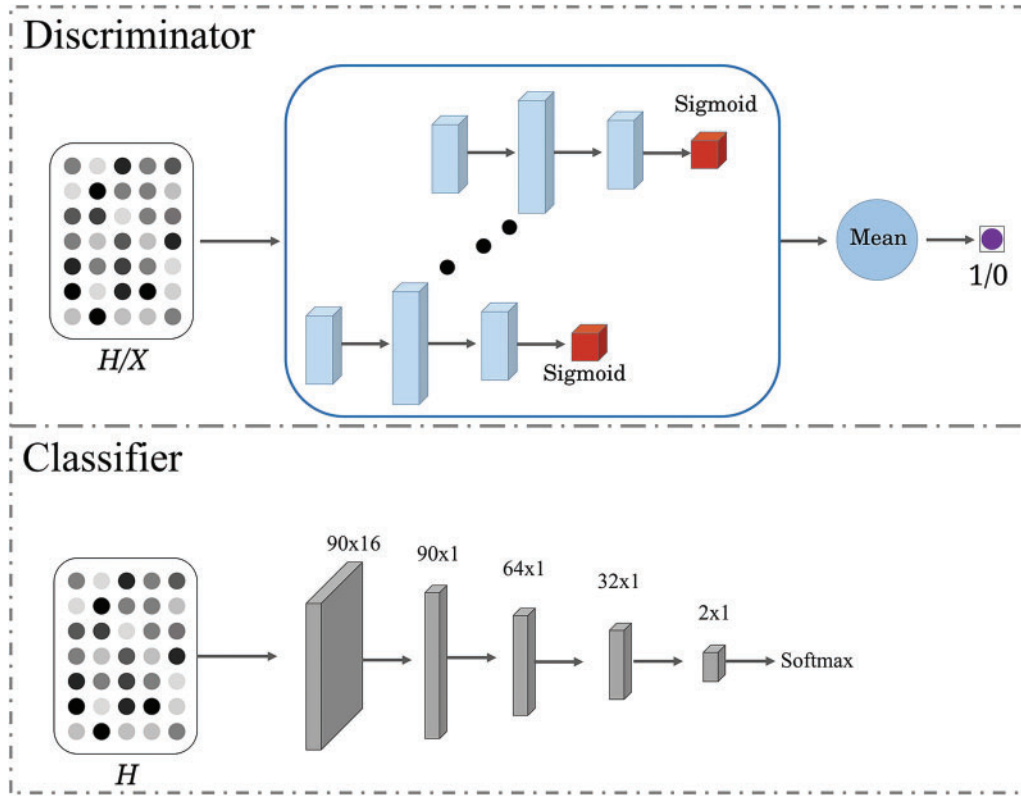
The graph encoder aims to enforce the latent node representation  $H$  to be consistent with the estimated label distribution  $P_x$ . The estimated label distribution  $P_x$  is utilized to guide the graph encoder to learn a robust representation. The discriminator plays as a referee to supervise the graph encoder to learn a distribution-consistent representation. Specifically, we sample a matrix  $X \in \mathbb{R}^{N \times p}$  from the distribution  $P_x$ . The sampled matrix  $X$  is the real sample for adversarial learning, while the fake sample is the output  $H$  of the graph encoder. As shown in Fig. 2, the discriminator comprises  $N$  sub-networks, where each discriminates the true or false of only one ROI representation. Each subnetwork is built on a three-layer perceptron with hidden neurons 32, 64, and 1. Each subnetwork consists of a *sigmoid* activation function to keep the output in the range of 0~1. The output of the discriminator is the mean value of all the subnetwork outputs. The discriminator loss is:

$$\mathcal{L}_{dis} = \mathbb{E}_{F \sim P_{fMRI}, A_{ori} \sim P_{fMRI}} [-\log(1 - \mathbf{D}(\mathbf{E}(A_{ori}, F)))] + \mathbb{E}_{X \sim P_x} [-\log(\mathbf{D}(X))] \quad (4)$$

In addition, to make the node representation discriminative, the cross-entropy loss is introduced to further regularize the learned node representation. The binary classifier  $\mathbf{C}$  is shown in the lower part of Fig. 2. The node representation  $H$  passes five Multi-Layer Perceptron (MLP) and outputs a vector with two elements, followed by a softmax to predict the most likely disease category. The classifier loss can be computed as follows:

$$\mathcal{L}_{cla} = \mathbb{E}_{F \sim P_{fMRI}, A_{ori} \sim P_{fMRI}} [y \cdot \log(\mathbf{C}(\mathbf{E}(A_{ori}, F)))] \quad (5)$$

here, the  $y$  is a one-hot vector (i.e., [0,1] represents the LMCI, and [1,0] represents NC).



**Figure 2:** Illustration of the discriminator and classifier structure. The input of the discriminator is a matrix computed from either latent node representation  $H$  or label distribution  $P_x$ , and the output is true (1) or false (0)

### 2.2.3 Transformer Generator

The transformer generator  $\mathbf{G}$  process maps each node in latent space  $H$  to a reconstructed graph  $A_{rec}$ . Also, the  $\mathbf{G}$  can generate similar brain networks  $A_{gen}$  by inputting matrix  $X$  sampled from the label distribution  $P_x$ . The generator module comprises three connectivity transformer (CT) layers, two dimensions upscaling (DU) layers, and connectivity prediction operation. The connectivity transformer layer contains norm, linear mapping (LM), head split, attention, dot-product, and concatenate. 4, 8, and 11 heads are designed in the three successive CT layers. Note that each CT's input and output dimension is the same. The output dimension of the two DU layers is 64 and 187, respectively. Each DU has only one layer. After latent representation  $H$  passes through the CT and DU layers, the inner product and  $\tanh$  activation function are utilized to predict connectivity with the range  $-1 \sim 1$ . The reconstructed BFN  $A_{rec}$  and generated BFN  $A_{gen}$  are given by:

$$A_{rec} = \mathbf{G}(H) = \tanh(H' \cdot H'^T) \quad (6)$$

$$H' = CT(DU_2(CT(DU_1(CT(H)))))) \quad (7)$$

$$A_{gen} = \mathbf{G}(X) = \tanh(X' \cdot X'^T) \quad (8)$$

$$X' = CT(DU_2(CT(DU_1(CT(X)))))) \quad (9)$$

The reconstruction loss is adopted to preserve the original graph structure while making the autoencoder training stable. We choose the L1 norm to measure the distance between the original  $A_{ori}$  and reconstructed  $A_{rec}$ . It is defined as:

$$\mathcal{L}_{rec} = \mathbb{E}_{A_{ori} \sim P_{fMRI}, F \sim P_{fMRI}} (\|G(E(A_{ori}, F)) - A_{ori}\|_1) \quad (10)$$

Moreover, to make the generator learning more stable, we put the reconstructed brain network  $A_{rec}$  to the graph encoder and obtain consistent node representation  $\hat{H}$ . The node-representation consistent loss is calculated by minimizing the distance generated between  $H$  and  $\hat{H}$ :

$$\mathcal{L}_{nrc} = \mathbb{E}_{H \sim P_H} (\|\hat{H} - H\|_1) \quad (11)$$

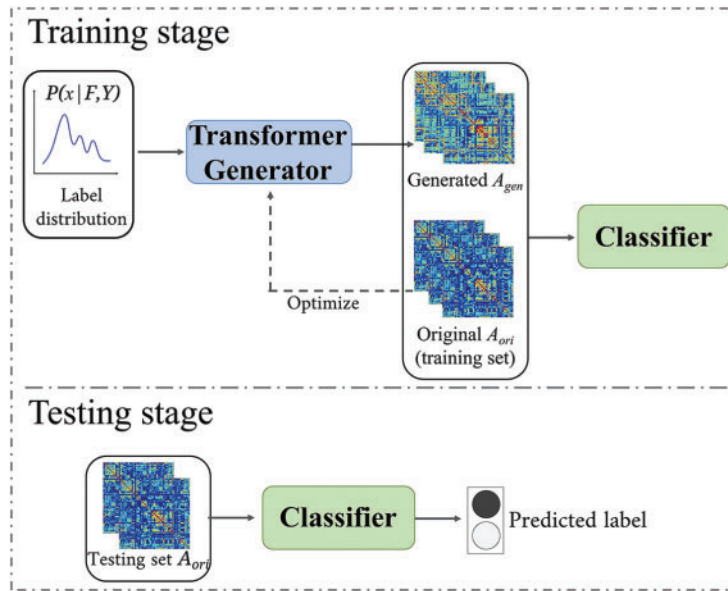
$$\hat{H} = E(G(H)) \quad (12)$$

### 2.3 Classification Training and Evaluation Metrics

In summary, the optimization strategy of the proposed DAGAE updates the weights of the graph encoder, discriminator, classifier, and transformer generator. The hybrid loss is defined by:

$$\mathcal{L}_{all} = \mathcal{L}_{enc} + \mathcal{L}_{dis} + \mathcal{L}_{cla} + \mathcal{L}_{rec} + \mathcal{L}_{nrc} \quad (13)$$

As illustrated in Fig. 3, we send the training set to the DAGAE model for training and augment the BFN with the transformer generator. The detailed training of the DAGAE model is shown in Algorithm 1. Inspired by the method [49,50], the latent feature learning can be stable when the optimization converges. For each label (i.e., NC and LMCI), we sample matrices from the distribution  $P_x$  and generate  $k$  times the number of original BFNs. It should be noted that the generated BFNs are not seen in the test set.



**Figure 3:** The entire workflow of this work. In the training stage, the transformer generator is first extracted from the trained DAGAE and then maps the distribution-sampled representation into generated BFNs ( $A_{gen}$ ). At last, the combination of the generated and original BFNs is utilized for training the classifier. In the testing stage, only the original BFNs in the testing set are used to predict the disease label

In the classification stage, we build a sample graph classifier modified from [24], including two GCN layers with 32 and 16 hidden neurons, one graph pooling, and one MLP layer with two neurons. The original training set is first used to train the graph classifier. Then the generated BFNs are mixed in the training set to finetune the classifier for enhancing classification performance. At last, the trained classifier predicts disease labels of the testing set for performance evaluation. There are four commonly used metrics for the prediction assessment: Accuracy (ACC), Specificity (SPE), Sensitivity (SEN), and the Area Under the receiver operating characteristic Curve (AUC). They are defined as:

$$ACC = \frac{TL + TN}{TL + TN + FL + FN} \quad (14)$$

$$SEN = \frac{TL}{TL + FN} \quad (15)$$

$$SPE = \frac{TN}{TN + FL} \quad (16)$$

where,  $TN$  means that NC is correctly predicted,  $TL$  means that LMCI is correctly predicted.  $FN$  means that NC is incorrectly predicted,  $FL$  means that LMCI is incorrectly predicted.

---

**Algorithm 1:** Optimizing the DAGAE model
 

---

**Input:**  $F$ : brain functional feature;  
 $A_{ori}$ : original BFN;  
 $P_x(F, Y)$ : label distribution;  
 $O$ : the number of iterations;  
 $T$ : the number of steps for updating discriminator

**Output:**  $A_{gen} \in \mathbb{R}^{N \times N}$ : the generated BFN

- 1: **for**  $i = 1, 2, \dots, O$  **do**
- 2:     **for**  $j = 1, 2, \dots, T$  **do**
- 3:         Compute the latent representation matrix  $H$  using Eq. (2)
- 4:         Sample a true matrix  $X = \{x_1, x_2, \dots, x_N\}$  from the label distribution  $p_x\{F, Y\}$
- 5:         Compute the loss function  $\mathcal{L}_{dis}$
- 6:         Update the discriminator  $\mathbf{D}$  by propagating the gradient  $\nabla \mathcal{L}_{dis}^j[\log(\mathbf{D}(X)) + \log(1 - \mathbf{D}(H))]$
- 7:     **end for**
- 8:     Compute the combined loss function  $\mathcal{L} = \mathcal{L}_{enc} + \mathcal{L}_{cla} + \mathcal{L}_{rec} + \mathcal{L}_{mrc}$
- 9:     Update the encoder, classifier, and generator by back-propagating the gradient  $-\nabla \mathcal{L}^i$
- 10:     Compute the generated BFN  $A_{gen}^i$  using Eq. (8)
- 11:     Replace the generated  $A_{gen}$  with  $A_{gen}^i$
- 12: **end for**

---

### 3 Experiment and Results

#### 3.1 Experimental Setup

We adopt the 5-fold cross-validation strategy in the experiment to conduct the training and testing. The preprocessed data is evenly separated into five folds, meaning each fold contains 15 NCs and 15 LMCIs. We first selected one-fold data and sent the rest of the four folds data (60 NCs and 60 LMCIs) into the DAGAE model for training. Next, the trained generator is used to generate  $k$  (default value 1.0) times the training data. Then the generated and original training data are jointly to train the

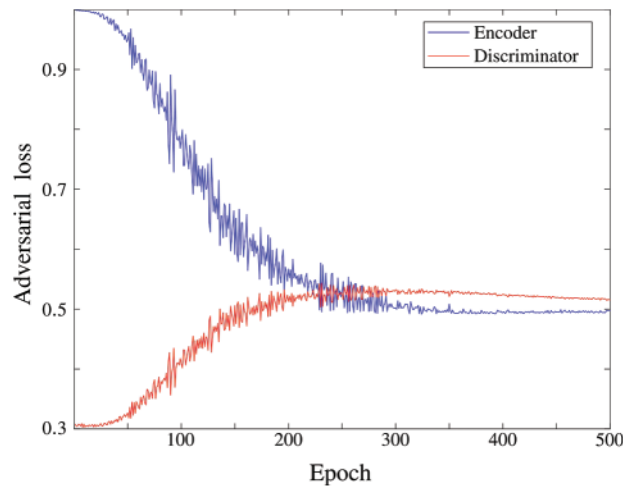


classifier. Finally, the classifier predicts the disease label of the selected one-fold data for performance evaluation. In this experiment, three other augmented methods (i.e., EW [29], SMOTE [30] and ARAE [44]) and three classical classifiers (i.e., SVM [51], DNN [52], and GCN [53]) are introduced to test the effectiveness of the proposed model.

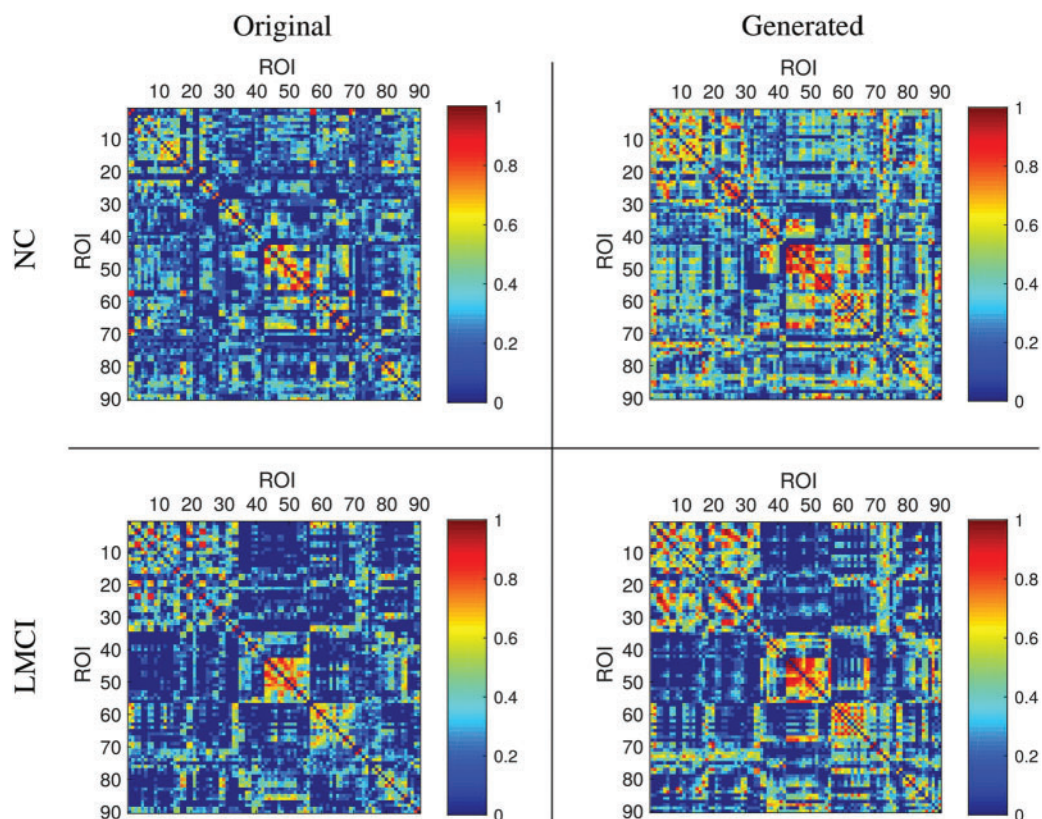
The DAGAE is trained on Ubuntu18.04 using the TensorFlow framework for BFN synthesis. The graphical device is one GPU with NVIDIA Quadro P4000 8.0 GB. We set the model's parameters with the values as follows:  $N = 90, p = 32$ . In the DAGAE training, we first update the weighting parameters of the graph encoder, classifier, and discriminator and then optimize the generator parameters. The learning rate for the encoder and the discriminator is set at 0.001 and 0.0001, respectively. The learning rate of the classifier and the generator are the same as the discriminator. The Adam algorithm is selected for training with batch size 16. The learning process terminates when the discriminator cannot identify the input from the node representation or the prior label distribution and the change of total loss is stable. After the DAGAE has been trained, the generator is used to augment BFNs for training the classifier. The learning rate of the classifier is set as 0.0001. We took about 1000 epochs for training to ensure the classifier's convergence. The ACC value is defined as the classification performance evaluation, which is used to optimize the classifiers.

### 3.2 Prediction Performance

In the experiment, it is essential to constrain the latent node representation to follow the label distribution. This constraint can diminish the model overfitting and stabilize the representation learning. Fig. 4 shows the adversarial loss over epochs in the training process. In the beginning, the encoder loss falls, and the discriminator rises. After 250 epochs, both keep around 0.5 steadily, which means the adversarial training converges. After the training, the transformer generator generates new BFNs by accepting node representation matrices sampled from the label distribution  $P_x$ . As shown in Fig. 5, examples of the original and generated BFNs are compared qualitatively. It can be seen that the generated BFN can preserve the main patterns of the original BFN.



**Figure 4:** The loss curve of the adversarial training. It is utilized to constrain the latent node representations in the label distribution

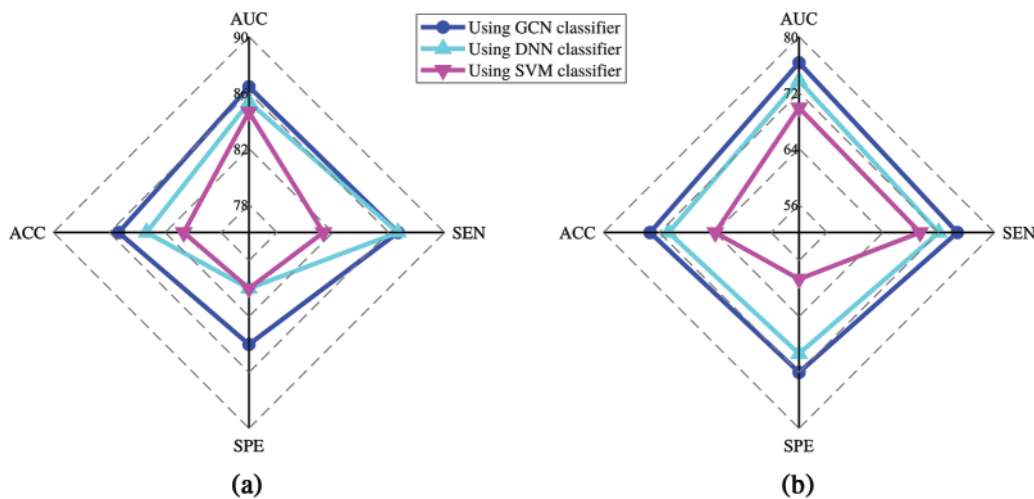


**Figure 5:** Visualization of the original and generated BFNs. The left column shows the original BFNs at NC and LMCI stages, and the right column shows the generated BFNs at NC and LMCI stages

We analyze the classification performance with different classifiers to investigate the proposed model's effectiveness. As shown in [Table 1](#), the generated BFNs can gain better classification performance over original BFNs. Among the three augmented methods, our model achieved superior results in three classifiers with more than 10 percent of ACC value compared to results using original BFNs. Also, our model increases the ACC value by 1.3%, 2.0%, and 3.3% compared with the competing ARAE method for GCN, DNN, and SVM classifiers, respectively. This evidence proves that the proposed model can generate more effective BFNs for classification improvement. To detail the effectiveness of the GCN-based classifier, [Figs. 6a](#) and [6b](#) show that the prediction results using a GCN-based classifier achieves the best performance than other traditional classifiers. Note that both original and generated BFNs using different methods are sent to the same classifier for classification performance evaluation. [Fig. 7](#) also shows better performance of the GCN-based classifier. Our model shows superior prediction performance in terms of ACC, SEN, SPE, and AUC by achieving 85.33%, 84.0%, 86.67%, and 86.42%. It probably indicates that the GCN-based classifier can benefit the topological properties buried in the BFNs and enhance the classification of BFNs.

**Table 1:** Comparison of classification performance using different generated BFNs

BFNs generated from	Classifier	ACC	SEN	SPE	AUC
Original	SVM	64.00%	58.66%	69.33%	69.85%
EW [29]	SVM	68.00%	70.67%	65.33%	74.63%
SMOTE [30]	SVM	75.33%	78.67%	72.00%	79.70%
ARAE [44]	SVM	77.33%	70.67%	74.00%	79.11%
<b>Our model</b>	SVM	<b>80.67%</b>	<b>80.00%</b>	<b>81.33%</b>	<b>84.62%</b>
Original	DNN	70.67%	69.33%	72.00%	73.71%
EW [29]	DNN	75.33%	70.67%	79.99%	77.10%
SMOTE [30]	DNN	80.67%	78.67%	82.67%	82.66%
ARAE [44]	DNN	81.33%	77.33%	85.33%	84.44%
<b>Our model</b>	DNN	<b>83.33%</b>	<b>80.00%</b>	<b>86.67%</b>	<b>85.32%</b>
Original	GCN	73.33%	72.00%	74.67%	76.28%
EW [29]	GCN	80.67%	81.33%	79.99%	82.95%
SMOTE [30]	GCN	82.67%	81.33%	84.00%	83.56%
ARAE [44]	GCN	84.00%	82.67%	85.33%	86.91%
<b>Our model</b>	GCN	<b>85.33%</b>	<b>84.00%</b>	<b>86.67%</b>	<b>86.42%</b>

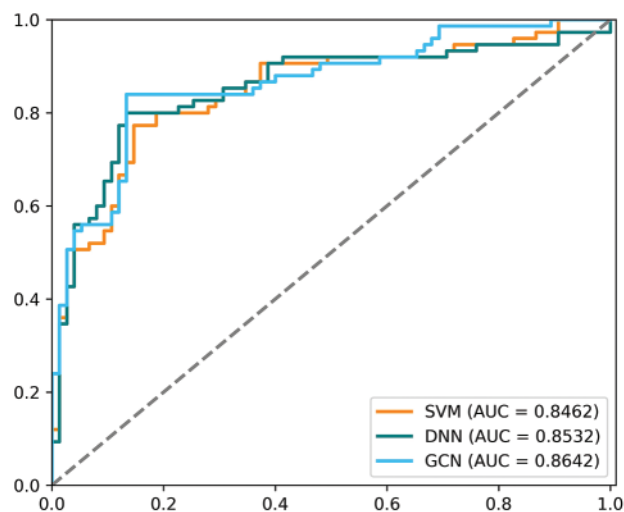


**Figure 6:** The prediction performance comparison using three classifiers by inputting BFNs from (a) our model, (b) original

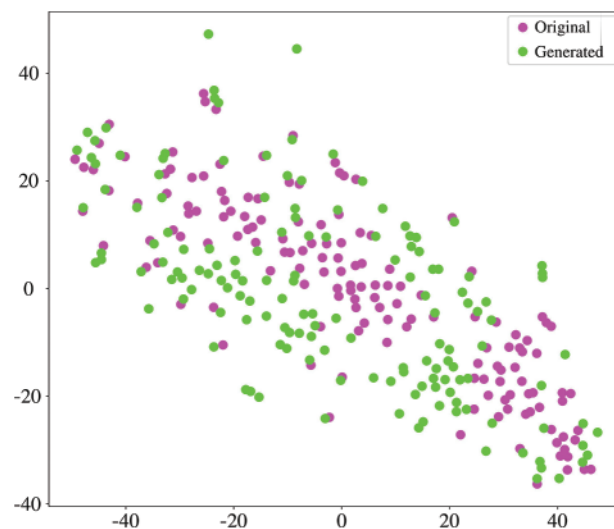
### 3.3 Evaluation of the Generated BFNs

In this section, we evaluate the coherence between original and generated BFNs. We generated the same size as the original data in each one-fold training. We employed the t-distributed Stochastic Neighbour Embedding (t-SNE) tool [54] to analyze the graphical characteristics. Fig. 8 shows the projection of the embedding representation of original and generated BFNs from high-dimensional

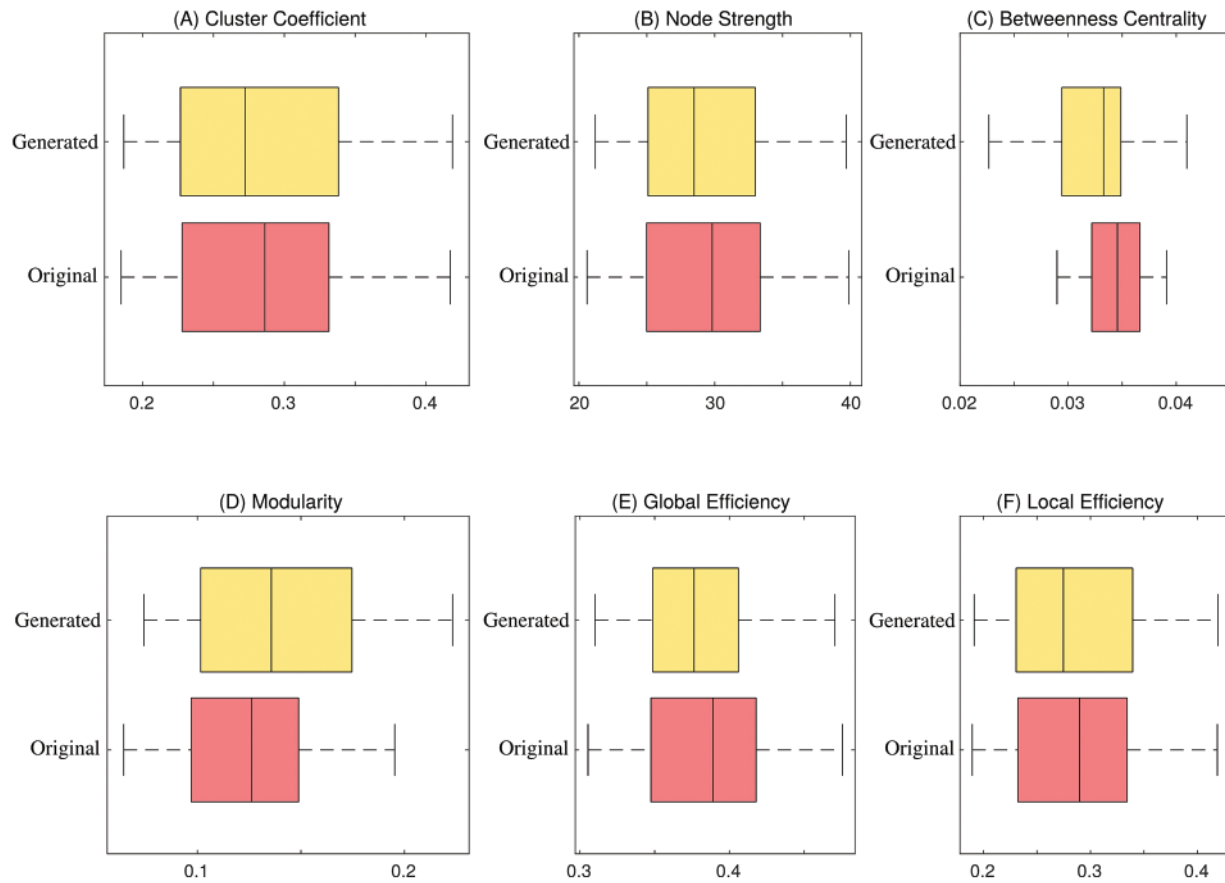
space to two-dimensional space. The generated data is consistent with the original data distribution, which ensures the similarity between the generated and original FBNs. In addition, six common metrics are utilized to quantitatively measure the effectiveness of the generated data. These six metrics can provide a relatively reliable measure of generated BFN's quality, including clustering coefficient, node strength, betweenness centrality, modularity, local efficiency, and global efficiency. As shown in Fig. 9, the boxplot distribution of each metric is compared between the generated and original BFNs. The generated data can mostly cover the range of graph metrics from the original data. Therefore, the generated BFNs by our model contains non-Euclidean characteristics and preserve the overall nature of brain connectivity, which is suitable to augment the BFNs for dementia diagnosis.



**Figure 7:** The comparison of ROC curves using three classifiers. The gray dotted line represents the random classifier



**Figure 8:** The comparison of embedded t-SNE representation between the original and generated BFNs

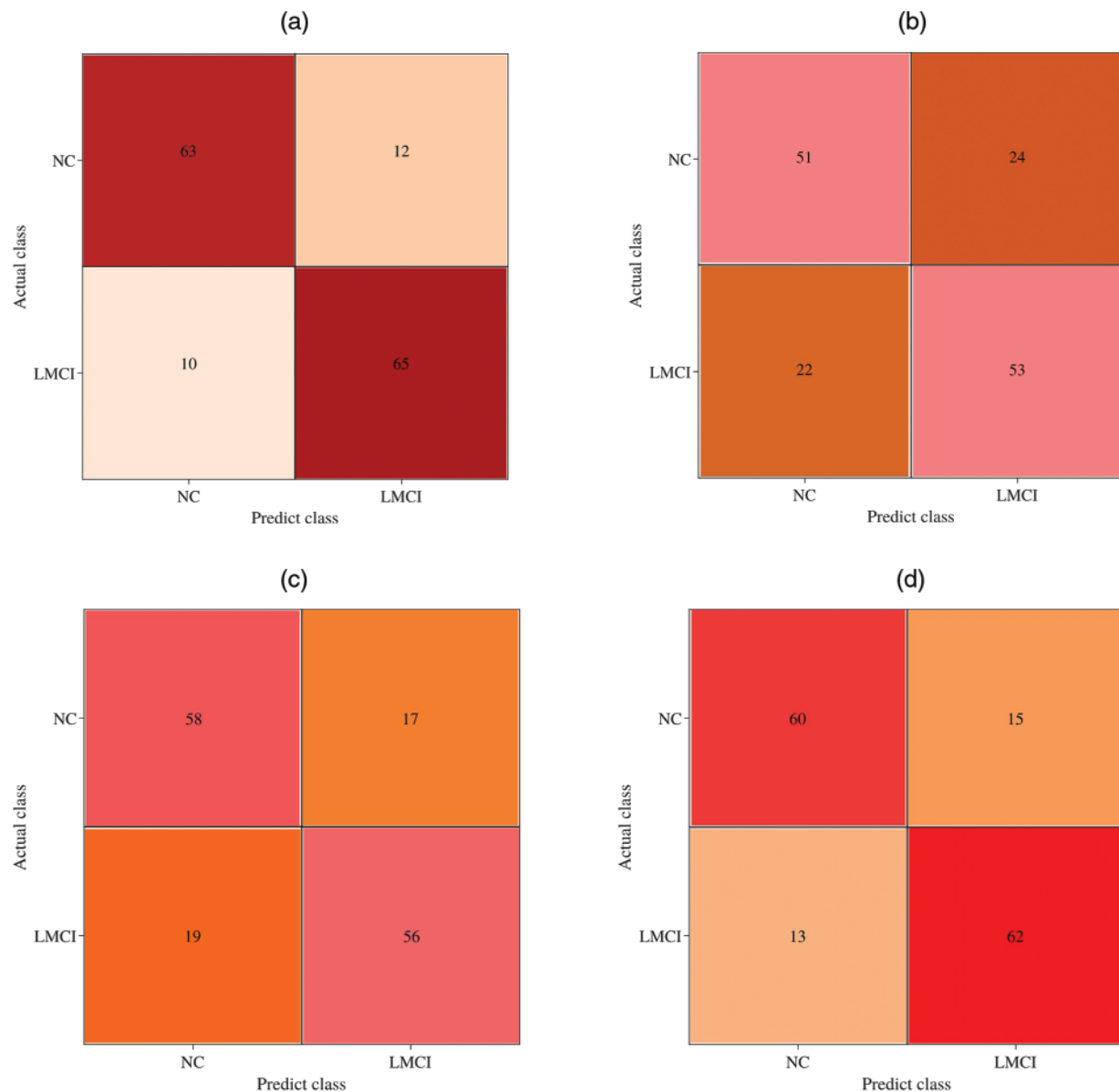


**Figure 9:** Statistical analysis of the original and generated BFN. (A) Cluster coefficient, (B) Node strength, (C) Betweenness centrality, (D) Modularity, (E) Global efficiency, and (F) Local efficiency

#### 4 Discussion

The proposed DAGAE model can generate new BFNs for improving classification performance. Each module in the model contributes to the generation quality of BFNs. To analyze the influence of different modules, we remove the encoder, discriminator, and classifier from the DAGAE model and evaluate the final classification performance. Fig. 10 demonstrates that the encoder significantly impacts the whole model. It drops by 16% in terms of ACC by removing the encoder module. The discriminator ensures the latent node representation in a prior distribution, which also influences the quality of the generated BFNs. This suggests the usefulness of prior label distribution can regularize the latent representation with a stable learning strategy and enhance the BFN classification. Furthermore, we study the dimension  $p$  of the latent node representation  $H$  in the prediction performance. As shown in Fig. 11, the value of ACC and AUC shows relatively stable fluctuation when  $p$  exceeds 32. Considering the computation efficiency, we select  $p = 32$  in the experiment. The transformer generator is essential for the generated BFN quality. We study two variations of the designed transformer generator to test its effectiveness. (1) Remove the DU layer in the generator (No-DU), which means the input and output dimension is the same as the dimension of latent representation  $H$ ; (2) remove the connectivity transformer (No-CT), which simplifies the generator into two layer perceptrons. The results are illustrated in Table 2; it can be seen that the combination of DU and CT achieves the best

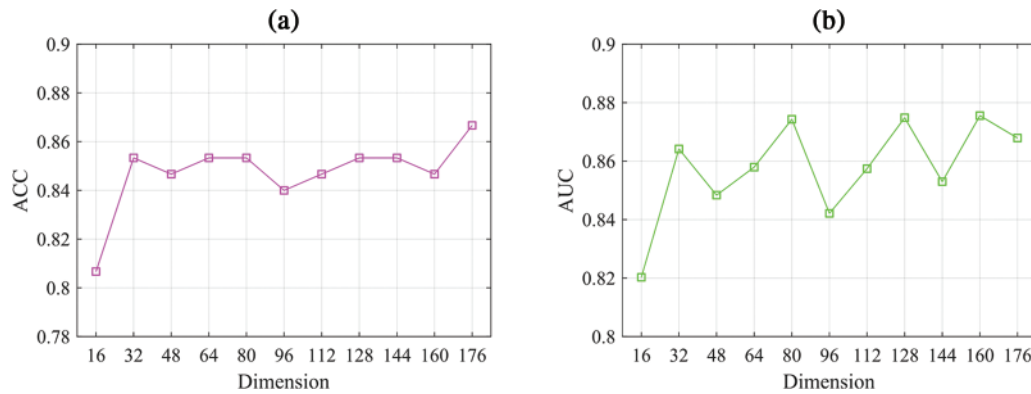
prediction performance. The classification results demonstrate the effectiveness of the transformer generator in the model. This can be explained by that the transformer-based network in the generator preserves the main topological properties and captures more discriminative features.



**Figure 10:** Influence of different modules on the classification performance. (a) The proposed DAGAE, (b) DAGAE without encoder, (c) DAGAE without discriminator, and (d) DAGAE without the classifier

Data-driven models achieve better performance by using large amounts of data. To investigate how much generated data influences the prediction performance, we generate  $k \in \{1, 2, \dots, 8\}$  times the original training set, combine the original BFNs and generated BFNs to train the SVM-based and GCN-based classifier. The mean ACC is estimated by predicting five-fold original test sets separately.

The metric  $\Delta ACC$  is defined as  $ACC_k - ACC_0$ . Here,  $ACC_0$  refers to the predicted results of the classifier trained using the original train set, and the  $ACC_k$  refers to the predicted results of the classifier trained using the original and generated data. Fig. 12 gives the different quantities of generated BFNs that maximizes classification performance in both classifiers. The best quantity of generated data is about five times the original data, with the largest  $\Delta ACC$  value of 19.3% and 13.3% for the SVM and GCN classifier, respectively. The reason why more data degrades classification performance may be that the generated BFNs bring a lot of noise. Besides, compared with the SVM classifier, the GCN-based classifier increases by 5.3% on the best ACC value. This marginal increase may be explained by considering the topological information among the brain regions, which can characterize the disease-related features of the BFNs. Thus, in the BFN augmentation experiment, it is better to choose the GCN-based classifier to evaluate the prediction performance, and different quantities of generated BFNs should be explored to maximize the effect of data augmentation.

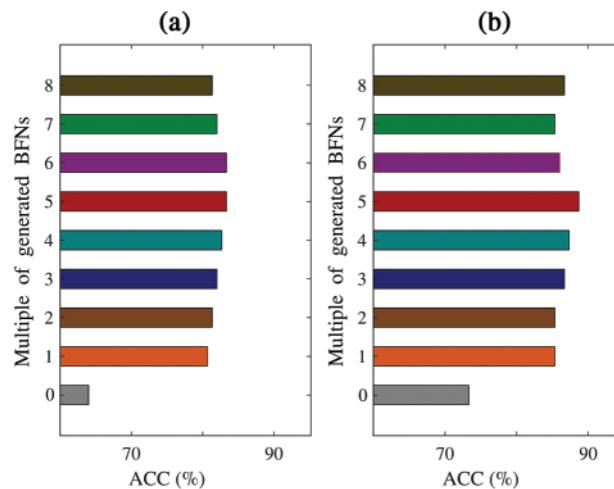


**Figure 11:** Impact of the dimension  $p$  of the learned node representation  $H$  on the ACC and AUC

**Table 2:** Effect of different structures in the transformer generator

Method	ACC	SEN	SPE	AUC
Our model	<b>85.33%</b>	84.00%	<b>86.67%</b>	<b>86.42%</b>
No-DU	81.33%	84.00%	78.67%	82.84%
No-CT	78.67%	82.67%	74.67%	82.63%

Although the proposed DAGAE is promising in augmenting the BFN data for disease prediction, there are still two limitations that have not been considered. (1) The label distribution is estimated using the limited training data, which can not add other prior knowledge. We will introduce disease-related anatomical brain knowledge into the model for performance evaluation. (2) The data in this work are deliberately picked out to maintain category balance. The real condition in the category distribution is always imbalanced. In the following study, we will try to apply the proposed DAGAE in category-imbalanced datasets for other brain disorder diagnosis.



**Figure 12:** Classification accuracy analysis using different amounts of generated data by our model using (a) SVM and (b) GCN classifier, respectively

## 5 Conclusions

This study proposes a novel DAGAE model to augment new BFNs for dementia diagnosis. The BFN augmentation is different from traditional image synthesis, where the latter only extracts local patterns and ignores the topological information buried in the brain network. Our model is novel in two aspects. One is that the estimated label distribution can regularize the latent space and make the learning process stable. Another one is that the transformer generator is devised to map the node representations into node-to-node connections by exploring the long-term dependence of highly-correlated distant brain regions, which preserves the main topological properties and more discriminative features. Testing on the Alzheimer's Disease Neuroimaging Initiative (ADNI) public dataset, the proposed DAGAE can generate similar and high-quality BFNs. The classification results show that adding generated data can achieve higher accuracy values of 85.33%, 83.33%, and 80.67% than the original method using GCN, DNN, and SVM classifiers, respectively. The proposed model also performs better than related augmented models, providing new insight for improving cognitive disease diagnosis accuracy.

**Funding Statement:** This paper is partially supported by the British Heart Foundation Accelerator Award, UK (AA\18\3\34220); Royal Society International Exchanges Cost Share Award, UK (RP202G0230); Hope Foundation for Cancer Research, UK (RM60G0680); Medical Research Council Confidence in Concept Award, UK (MC\_PC\_17171); Sino-UK Industrial Fund, UK (RP202G0289); Global Challenges Research Fund (GCRF), UK (P202PF11); LIAS Pioneering Partnerships Award, UK (P202ED10); Data Science Enhancement Fund, UK (P202RE237); Fight for Sight, UK (24NN201); Sino-UK Education Fund, UK (OP202006); Biotechnology and Biological Sciences Research Council, UK (RM32G0178B8); LIAS Seed Corn, UK (P202RE969).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.



## References

1. Samarasinghe, R. A., Miranda, O. A., Buth, J. E., Mitchell, S., Ferando, I. et al. (2021). Identification of neural oscillations and epileptiform changes in human brain organoids. *Nature Neuroscience*, 24(10), 1488–1500.
2. Suárez, L. E., Markello, R. D., Betzel, R. F., Misic, B. (2020). Linking structure and function in macroscale brain networks. *Trends in Cognitive Sciences*, 24(4), 302–315.
3. Schumacher, J., Peraza, L. R., Firbank, M., Thomas, A. J., Kaiser, M. et al. (2019). Dynamic functional connectivity changes in dementia with lewy bodies and alzheimer's disease. *NeuroImage: Clinical*, 22(37), 101812.
4. Borelli, C. M., Grennan, D., Muth, C. C. (2020). Causes of memory loss in elderly persons. *JAMA*, 323(5), 486–486.
5. Berron, D., Vogel, J. W., Insel, P. S., Pereira, J. B., Xie, L. et al. (2021). Early stages of tau pathology and its associations with functional connectivity, atrophy and memory. *Brain*, 144(9), 2771–2783.
6. Fu, Z., Caprihan, A., Chen, J., Du, Y., Adair, J. C. et al. (2019). Altered static and dynamic functional network connectivity in alzheimer's disease and subcortical ischemic vascular disease: Shared and specific brain connectivity abnormalities. *Human Brain Mapping*, 40(11), 3203–3221.
7. Chen, M., Li, H., Wang, J., Dillman, J. R., Parikh, N. A. et al. (2019). A multichannel deep neural network model analyzing multiscale functional brain connectome data for attention deficit hyperactivity disorder detection. *Radiology: Artificial Intelligence*, 2(1), e190012.
8. Ma, W. Y., Yao, Q., Hu, G. J., Ge, H. L., Xue, C. et al. (2022). Reorganization of rich clubs in functional brain networks of dementia with lewy bodies and alzheimer's disease. *NeuroImage: Clinical*, 33(12), 102930.
9. Zuo, Q., Lei, B., Shen, Y., Liu, Y., Feng, Z. et al. (2021). Multimodal representations learning and adversarial hypergraph fusion for early alzheimer's disease prediction. *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, Beijing, China, Springer.
10. Teipel, S. J., Wohler, A., Metzger, C., Grimmer, T., Sorg, C. et al. (2017). Multicenter stability of resting state fMRI in the detection of alzheimer's disease and amnesic MCI. *NeuroImage: Clinical*, 14(Suppl. 3), 183–194.
11. Ahmed, M. R., Zhang, Y., Feng, Z., Lo, B., Inan, O. T. et al. (2018). Neuroimaging and machine learning for dementia diagnosis: Recent advancements and future prospects. *IEEE Reviews in Biomedical Engineering*, 12, 19–33.
12. Wang, S., Wang, H., Shen, Y., Wang, X. (2018). Automatic recognition of mild cognitive impairment and alzheimers disease using ensemble based 3D densely connected convolutional networks. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, USA, IEEE.
13. Hong, J., Wang, S. H., Cheng, H., Liu, J. (2020). Classification of cerebral microbleeds based on fully-optimized convolutional neural network. *Multimedia Tools and Applications*, 79(21), 15151–15169.
14. Yu, W., Lei, B., Wang, S., Liu, Y., Feng, Z. et al. (2022). Morphological feature visualization of alzheimer's disease via multidirectional perception GAN. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15. USA, IEEE. <https://doi.org/10.1109/TNNLS.2021.3118369>
15. Zong, Y., Jing, C., Zuo, Q. (2022). Multiscale autoencoder with structural-functional attention network for alzheimer's disease prediction. *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, Shenzhen, China, Springer.
16. Sheykhivand, S., Rezaii, T. Y., Mousavi, Z., Delpak, A., Farzamnia, A. (2020). Automatic identification of epileptic seizures from EEG signals using sparse representation-based classification. *IEEE Access*, 8, 138834–138845.
17. Meier, T. B., Desphande, A. S., Vergun, S., Nair, V. A., Song, J. et al. (2012). Support vector machine classification and characterization of age-related reorganization of functional brain networks. *NeuroImage*, 60(1), 601–613.

18. Yu, R., Qiao, L., Chen, M., Lee, S. W., Fei, X. et al. (2019). Weighted graph regularized sparse brain network construction for MCI identification. *Pattern Recognition*, 90, 220–231.
19. Bi, X., Zhao, X., Huang, H., Chen, D., Ma, Y. (2020). Functional brain network classification for alzheimer's disease detection with deep features and extreme learning machine. *Cognitive Computation*, 12(3), 513–527.
20. Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V. et al. (2017). BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146(4), 1038–1049.
21. Yu, S., Wang, S., Xiao, X., Cao, J., Yue, G. et al. (2020). Multi-scale enhanced graph convolutional network for early mild cognitive impairment detection. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Lima, Peru, Springer.
22. Xiao, L., Wang, J., Kassani, P. H., Zhang, Y., Bai, Y. et al. (2019). Multi-hypergraph learning-based brain functional connectivity analysis in fMRI data. *IEEE Transactions on Medical Imaging*, 39(5), 1746–1758.
23. Ji, J., Xing, X., Yao, Y., Li, J., Zhang, X. (2021). Convolutional kernels with an element-wise weighting mechanism for identifying abnormal brain connectivity patterns. *Pattern Recognition*, 109(16), 107570.
24. Zuo, Q., Lu, L., Wang, L., Zuo, J., Ouyang, T. (2022). Constructing brain functional network by adversarial temporal-spatial aligned transformer for early AD analysis. *Frontiers in Neuroscience*, 16, 108815.
25. Ren, Z., Zhang, Y., Wang, S. (2022). Lcdae: Data augmented ensemble framework for lung cancer classification. *Technology in Cancer Research & Treatment*, 21(5), 15330338221124372.
26. Ramaraj, K., Govindaraj, V., Zhang, Y. D., Murugan, P. R., Thiyagarajan, A. (2021). Brain anomaly prediction with the intervention of fuzzy based clustering and optimization techniques for augmenting clinical diagnosis. *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, Greater Noida, India, IEEE.
27. Hong, J., Feng, Z., Wang, S. H., Peet, A., Zhang, Y. D. et al. (2020). Brain age prediction of children using routine brain MR images via deep learning. *Frontiers in Neurology*, 11, 584682.
28. Hu, S., Lei, B., Wang, S., Wang, Y., Feng, Z. et al. (2021). Bidirectional mapping generative adversarial networks for brain MR to pet synthesis. *IEEE Transactions on Medical Imaging*, 41(1), 145–157.
29. Meszlényi, R. J., Buza, K., Vidnyánszky, Z. (2017). Resting state fMRI functional connectivity-based classification using a convolutional neural network architecture. *Frontiers in Neuroinformatics*, 11, 61.
30. Gao, R., Peng, J., Nguyen, L., Liang, Y., Thng, S. et al. (2019). Classification of non-tumorous facial pigmentation disorders using deep learning and smote. *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, Sapporo, Japan, IEEE.
31. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D. et al. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139–144.
32. Hong, J., Yu, S. C. H., Chen, W. (2022). Unsupervised domain adaptation for cross-modality liver segmentation via joint adversarial learning and self-learning. *Applied Soft Computing*, 121, 108729.
33. Hu, S., Shen, Y., Wang, S., Lei, B. (2020). Brain MR to pet synthesis via bidirectional generative adversarial network. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham, Springer.
34. Hu, S., Yuan, J., Wang, S. (2019). Cross-modality synthesis from MRI to pet using adversarial U-Net with different normalization. *2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, Shenzhen, China, IEEE.
35. Hu, B., Lei, B., Shen, Y., Liu, Y., Wang, S. (2021). A point cloud generative model via tree-structured graph convolutions for 3D brain shape reconstruction. *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, Beijing, China, Springer.

36. You, S., Lei, B., Wang, S., Chui, C. K., Cheung, A. C. et al. (2022). Fine perceptive gans for brain MR image super-resolution in wavelet domain. *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13. USA, IEEE. <https://doi.org/10.1109/TNNLS.2022.3153088>
37. Wang, S., Wang, X., Hu, Y., Shen, Y., Yang, Z. et al. (2020). Diabetic retinopathy diagnosis using multichannel generative adversarial network with semisupervision. *IEEE Transactions on Automation Science and Engineering*, 18(2), 574–585.
38. Yu, W., Lei, B., Ng, M. K., Cheung, A. C., Shen, Y. et al. (2021). Tensorizing gan with high-order pooling for alzheimer's disease assessment. *IEEE Transactions on Neural Networks and Learning Systems*, 33(9), 4945–4959.
39. Sheykhivand, S., Mousavi, Z., Mojtahedi, S., Rezaii, T. Y., Farzamnia, A. et al. (2021). Developing an efficient deep neural network for automatic detection of COVID-19 using chest x-ray images. *Alexandria Engineering Journal*, 60(3), 2885–2903.
40. Lei, B., Liang, E., Yang, M., Yang, P., Zhou, F. et al. (2022). Predicting clinical scores for alzheimer's disease based on joint and deep learning. *Expert Systems with Applications*, 187(3), 115966.
41. Hu, D., Zhang, H., Wu, Z., Wang, F., Wang, L. et al. (2020). Disentangled-multimodal adversarial autoencoder: Application to infant age prediction with incomplete multimodal neuroimages. *IEEE Transactions on Medical Imaging*, 39(12), 4137–4149.
42. Hong, J., Zhang, Y. D., Chen, W. (2022). Source-free unsupervised domain adaptation for cross-modality abdominal multi-organ segmentation. *Knowledge-Based Systems*, 250, 109155.
43. Zuo, Q., Lei, B., Wang, S., Liu, Y., Wang, B. et al. (2021). A prior guided adversarial representation learning and hypergraph perceptual network for predicting abnormal connections of alzheimer's disease. arXiv preprint arXiv:2110.09302.
44. Pan, S., Hu, R., Long, G., Jiang, J., Yao, L. et al. (2018). Adversarially regularized graph autoencoder for graph embedding. arXiv preprint arXiv:1802.04407.
45. Tan, Y. F., Ting, C. M., Noman, F., Phan, R. C. W., Ombao, H. (2022). Graph-regularized manifold-aware conditional wasserstein gan for brain functional connectivity generation. arXiv preprint arXiv:2212.05316.
46. Jiang, Y., Chang, S., Wang, Z. (2021). Transgan: Two pure transformers can make one strong gan, and that can scale up. *Advances in Neural Information Processing Systems*, 34, 14745–14758.
47. Wang, J., Wang, X., Xia, M., Liao, X., Evans, A. et al. (2015). GRETNA: A graph theoretical network analysis toolbox for imaging connectomics. *Frontiers in Human Neuroscience*, 9, 386.
48. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O. et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1), 273–289.
49. Mahajan, S., Pandit, A. K. (2021). Hybrid method to supervise feature selection using signal processing and complex algebra techniques. *Multimedia Tools and Applications*, 82(4), 1–22.
50. Mahajan, S., Abualigah, L., Pandit, A. K., Altalhi, M. (2022). Hybrid aquila optimizer with arithmetic optimization algorithm for global optimization tasks. *Soft Computing*, 26(10), 4863–4881.
51. Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, 13(4), 18–28.
52. Kong, Y., Gao, J., Xu, Y., Pan, Y., Wang, J. et al. (2019). Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier. *Neurocomputing*, 324, 63–68.
53. Kipf, T. N., Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
54. van der Maaten, L., Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(11), 2579–2605.