



# Predicting miRNA-Disease Associations Based on Heterogeneous Graph Attention Networks

Cunmei Ji<sup>1\*</sup>, Yutian Wang<sup>1</sup>, Jiancheng Ni<sup>1</sup>, Chunhou Zheng<sup>2</sup> and Yansen Su<sup>2\*</sup>

<sup>1</sup> School of Cyber Science and Engineering, Qufu Normal University, Qufu, China, <sup>2</sup> School of Artificial Intelligence, Anhui University, Hefei, China

## OPEN ACCESS

### Edited by:

Hongmin Cai,  
South China University of Technology,  
China

### Reviewed by:

Xing Chen,  
China University of Mining and  
Technology, China  
Qi Zhao,  
University of Science and Technology  
Liaoning, China  
Wei Lan,  
Guangxi University, China

### \*Correspondence:

Cunmei Ji  
cunmeiji@126.com  
Yansen Su  
suyansen@ahu.edu.cn

### Specialty section:

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 19 June 2021

**Accepted:** 02 August 2021

**Published:** 25 August 2021

### Citation:

Ji C, Wang Y, Ni J, Zheng C and Su Y  
(2021) Predicting miRNA-Disease  
Associations Based on  
Heterogeneous Graph Attention  
Networks. *Front. Genet.* 12:727744.  
doi: 10.3389/fgene.2021.727744

In recent years, more and more evidence has shown that microRNAs (miRNAs) play an important role in the regulation of post-transcriptional gene expression, and are closely related to human diseases. Many studies have also revealed that miRNAs can be served as promising biomarkers for the potential diagnosis and treatment of human diseases. The interactions between miRNA and human disease have rarely been demonstrated, and the underlying mechanism of miRNA is not clear. Therefore, computational approaches has attracted the attention of researchers, which can not only save time and money, but also improve the efficiency and accuracy of biological experiments. In this work, we proposed a Heterogeneous Graph Attention Networks (GAT) based method for miRNA-disease associations prediction, named HGATMDA. We constructed a heterogeneous graph for miRNAs and diseases, introduced weighted DeepWalk and GAT methods to extract features of miRNAs and diseases from the graph. Moreover, a fully-connected neural networks is used to predict correlation scores between miRNA-disease pairs. Experimental results under five-fold cross validation (five-fold CV) showed that HGATMDA achieved better prediction performance than other state-of-the-art methods. In addition, we performed three case studies on breast neoplasms, lung neoplasms and kidney neoplasms. The results showed that for the three diseases mentioned above, 50 out of top 50 candidates were confirmed by the validation datasets. Therefore, HGATMDA is suitable as an effective tool to identify potential diseases-related miRNAs.

**Keywords:** disease, miRNA, graph attention networks, miRNA-disease association, DeepWalk

## 1. INTRODUCTION

MicoRNAs (miRNAs) are a class of endogenous non-coding RNAs with a length of about 21–25 nucleotides, which play an important role in the regulation of post-transcriptional gene expression in organisms (Ambros, 2001, 2004; Bartel, 2004, 2018). Over the past decades, researchers have identified hundreds of miRNAs in humans and shown that many of them interact with most human mRNAs (Friedman et al., 2009). Recent studies have discovered that miRNAs down-regulate gene expression by degrading or silencing targeting mRNAs, thereby affecting many cellular processes, such as growth, development, differentiation, and death (Ambros, 2003; He and Hannon, 2004; Bartel, 2009). Furthermore, many researches have found that human miRNAs are involved in many

human diseases (Croce, 2008; Li et al., 2014; Chou et al., 2016; Huang et al., 2019b). Therefore, miRNAs are promising biomarkers for diagnosis and treatment of human diseases.

Based on previous biological experiments, verifying the association between miRNA and disease is time-consuming and expensive. Computational methods can efficiently select the promising disease-associated miRNAs for further experimental verification. Many computational methods have been proposed for predicting the miRNA-disease associations, which can be roughly divided into two categories: one is based on similarity networks, the other is based on machine learning. Jiang et al. (2010) have firstly introduced similarity network to compute the score between disease and miRNA. They constructed functional miRNA network by computing the overlap of the target genes, and applied the hypergeometric distribution to calculate the association score between disease and miRNA. Since then, various network based approaches have been proposed (Chen et al., 2016, 2018c; Pan et al., 2019; Yu et al., 2019). You et al. (2017) have constructed a heterogeneous network, then proposed a novel path based method named PBMDA for inferring the disease-related miRNAs. However, this method only uses sub-graph information for prediction, which can be improved by considering the global information in the heterogeneous graph. Chen et al. (2012) have firstly developed random walk with restart for predicting miRNA-disease associations (RWRMDA). However, the requirement for at least one known related miRNA in inference may limit the application of RWRMDA. Subsequently, many methods based on random walk were proposed to improve the prediction performance.

Machine learning methods have also been introduced in this field. Matrix completion has been widely used in recommendation systems (Koren et al., 2009). Inspired by these, Li et al. (2017) developed a matrix completion based model to predict the disease-related miRNAs (MCMMDA). However, it will be failed to predict new disease (or miRNA) that has no connections with the known miRNAs (or diseases). Chen et al. (2018b) proposed a effective method of inductive matrix completion for predicting, which can be applied for new diseases (or miRNAs), named IMCMDA, it achieved better prediction performance than MCMMDA. Almost at the same time, Xiao et al. (2018) proposed a novel method called GRNMF, which combined graph Laplacian regularization with non-negative matrix factorization for miRNA-disease associations prediction in the integrated heterogeneous networks. Chen et al. (2018e) introduced low rank matrix decomposition to reduce noises in the datasets, then inferred the associations between miRNAs and diseases in the integrated heterogeneous graph, including disease semantic network, miRNA functional network, the relative GIP kernel networks, and miRNA-disease associations. Chen et al. (2018d) used bipartite recommendation algorithm to generate the association score between disease and miRNA. In addition, there are many other studies based on matrix completion and matrix factorization to predict potential connections between miRNAs and diseases (Shen et al., 2017; Zhong et al., 2018; Yan et al., 2019; Yu et al., 2019). Furthermore, lots of supervised learning algorithms have been introduced in this field. Chen et al. (2019) used decision tree to infer the miRNA-disease

associations. SVM-based methods have been used for predicting the potential relations between miRNAs and diseases (Xu et al., 2011; Chen et al., 2018a). Chen et al. (2015) introduced restricted Boltzmann machine for predicting associations and different types between miRNAs and diseases. Chen et al. (2021) integrated matrix completions with neighborhood constraint for prediction.

Recent decades, deep learning based methods have been gradually used in this field. Fu and Peng (2017) proposed a deep ensemble method and adopted stacked autoencoder for obtaining high-level features from integrated similarities, then a three layers neural networks (NN) was used for prediction. Peng et al. (2019) developed a CNN based model named MDA-CNN for prediction. They constructed disease-gene-miRNA networks, and obtained features of diseases and miRNAs by measuring Pearson correlation coefficient with genes. Then auto-encoder was used for feature selection, followed by a convolutional neural networks (CNN) model for further feature extraction between miRNA-disease pairs. Finally, two fully-connected layers was introduced for classification. Xuan et al. (2018) integrated two CNN models to predict correlation score of the miRNA-disease pair. In order to reduce the impact of negative sample missing on the prediction performance, Zhang et al. (2019) constructed two spliced matrices with the known miRNA-disease associations, disease similarity and miRNA similarity, then used variational autoencoder to calculate the unknown values of miRNA-disease pairs. Ji et al. (2021) used two regression models to learn dense vectors from integrated disease and miRNA similarities, applied the reconstruction probability of an autoencoder model for inferring.

Graph neural networks (GNNs) are a class of models that can effectively extract information from its neighborhood in the graph, it has achieved great success in social networks, knowledge graph, biology and so on (Zhou et al., 2020). DimiG considers the tissue-specific expressions as the features of miRNAs and protein coding genes, and then applies graph convolutional network (GCN) in the protein-coding genes and miRNAs networks (Pan and Shen, 2019). Li et al. (2020a) combined GCN and IMC algorithm to identify disease-related miRNAs. However, the input features of diseases and miRNAs were initialized randomly, which reduced the ability of GCN. Long et al. (2021) developed a novel computational model to predict microbe-disease association (GATMDA). It firstly constructed the input features by integrating similarities of diseases and microbes, and a bipartite network of known microbe-disease associations. Then, it used graph attention networks (GAT) for further learning representations of nodes from the graph. Finally, a decoder with inductive matrix completion (IMC) was selected for prediction. Long et al. (2020) also proposed GNN based model to predict human microbe-drug connections. Unlike GATMDA, they adopted RWR method for initial features learning, and then used GNN with random field. Additionally, negative samples were needed for regression training. GAEMDA integrated similarities of diseases and miRNAs as features of nodes, applied a GCN model for further feature extraction, and then used a bilinear decoder for identification (Li et al., 2020b).

In this paper, we proposed a novel computational model (HGATMDA) that combines graph embedding and graph

attention networks to infer disease-related miRNAs. Specifically, we first constructed a miRNA-disease heterogeneous graph. Then, we adopted weighted DeepWalk to learn dense representations of miRNAs and diseases from miRNA-miRNA sub-graph and disease-disease sub-graph, respectively. In addition, we utilized graph attention networks (GAT) to further learn the graph structure information from the miRNA-disease sub-graph. Furthermore, we presented a fully-connected neural networks for inferring the potential associations between miRNAs and diseases. Finally, we evaluated the performance of HGATMDA under five-fold cross validation (CV). The results showed that HGATMDA achieved the best area under ROC curve (AUC) of  $94.54 \pm 0.34\%$ , the area under P-R curve (AUPR) of  $94.05 \pm 0.18\%$ , Accuracy of  $87.02\%$ , Precision of  $94.07\%$ , Recall of  $90.04\%$ , F1-score of  $87.39\%$ . We conducted case studies on three common human diseases such as breast neoplasms, lung neoplasms, and kidney neoplasms, to further evaluate the performance of HGATMDA. For the three diseases, 50 out of top 50 candidates were confirmed by the validation datasets. Experimental results showed that HGATMDA performed better performance than other state-of-the-art methods, and can be used as an efficient and accurate tool to identify the underlying associations between miRNAs and diseases.

## 2. MATERIALS AND METHODS

### 2.1. Overview

In this study, we proposed a new computational method, HGATMDA, which combines graph embedding and graph attention network (GAT) methods to identify miRNA-disease associations. As shows in **Figure 1**, HGATMDA consists of four parts. Firstly, HGATMDA builds a heterogeneous graph including disease-disease sub-graph, miRNA-miRNA sub-graph and miRNA-disease sub-graph. Secondly, HGATMDA adopts a novel weighted DeepWalk to obtain dense representations of diseases and miRNAs from the disease-disease and miRNA-miRNA sub-graphs. Thirdly, it applies GAT to learn node features from the miRNA-disease sub-graph. Lastly, HGATMDA introduces a fully-connected (FC) neural network as an effective classifier for inferring the potential disease-related miRNAs.

### 2.2. Human miRNA Disease Database

HMDD v2.0 (Wang et al., 2010) and HMDD v3.2 (Huang et al., 2019a) databases, the experimental verified human miRNA-disease associations, are used in this paper. We directly downloaded data from <http://www.cuilab.cn/hmdd>, and constructed a miRNA-disease graph with known associations between miRNAs and diseases. We denote the association matrix as  $A^{N_m \times N_d}$ , and  $N_d$  is the number of diseases,  $N_m$  is the number of miRNAs. The value of  $A_{ij} \in \{0, 1\}$  indicates whether there is a known connections existed between miRNA  $m_i$  and disease  $d_j$ . Here,  $N_d = 383$  and  $N_m = 495$ , and 5,430 known interactions between miRNAs and diseases in the HMDD v2.0. We found that the names in HMDD v3.2, including diseases and miRNAs, did not exactly match those in HMDD V2.0. For consistency, we mapped the

items in HMDD v3.2 to HMDD v2.0 as the previous work (Wang et al., 2010).

### 2.3. Disease Semantic Similarity

The Medical Subject Headings (MeSH) dataset contains hierarchical relationships between diseases. Tree numbers in the Mesh headings denotes parent-child associations between nodes in the networks. By using the disease name as nodes and tree numbers as edges, we can build a directed acyclic graph (DAG) with the headings for each disease (<https://www.nlm.nih.gov/mesh/meshhome.html>). We define  $DAG_d = (T_d, E_d)$  as disease  $d$ , and  $T_d$  and  $E_d$  are the nodes and edges in the DAG. Then, we can compute the contribution of disease  $d_i \in T_d$  to disease  $d$  as follows:

$$\begin{cases} D_d(d_i) = 1 & \text{if } d = d_i \\ D_d(d_i) = \max\{\Delta D_d(d'_i) | d'_i \in \text{children of } d_i\} & \text{if } d \neq d_i \end{cases} \quad (1)$$

where  $\Delta$  is a weight decay parameter and is set as 0.5 in this paper. We define the semantic value of disease  $d$  by the following formula:

$$DV(d) = \sum_{d' \in T_d} D_d(d') \quad (2)$$

Based on the assumption, the more overlap between two DAGs, the more similar the two diseases are. The similarity between disease  $d_i$  and  $d_j$  can be calculated as follows:

$$SS(d_i, d_j) = \frac{\sum_{d' \in T_{d_i} \cap T_{d_j}} (D_{d_i}(d') + D_{d_j}(d'))}{DV(d_i) + DV(d_j)} \quad (3)$$

### 2.4. miRNA Functional Similarity

The basic assumption is that miRNAs with similar functions tend to be connected with similar diseases, and vice versa (Lu et al., 2008; Wang et al., 2010). Wang et al. have proposed miRNA MISIM functional similarity, that is, to calculate the similarity score by the related disease DAG between two miRNAs. Thanks to this excellent work, we can directly download the data from website (<http://www.cuilab.cn/files/images/cuilab/misim.zip>). We then constructed the miRNA functional network, which is denoted as  $FS$  with the shape of  $N_m \times N_m$ , where  $N_m$  is the number of miRNAs. The element  $FS(m_i, m_j)$  represents the similarity score between miRNA  $m_i$  and miRNA  $m_j$ .

### 2.5. Gaussian Interaction Profile Kernel Similarity for Disease and miRNA

In order to overcome the sparsity in disease semantic similarity  $SS$  and miRNA functional similarity  $FS$ , we introduced the Gaussian interaction profile (GIP) kernel similarity. Firstly, we used column  $A_i$  to represent a disease  $d_i$  and row  $A_j$  to represent miRNA  $m_j$ . Then, GIP similarities between two diseases or two miRNAs are defined as follows:

$$KD(d_i, d_j) = \exp(-\gamma_d \|A_i - A_j\|^2) \quad (4)$$

$$KM(m_i, m_j) = \exp(-\gamma_m \|A_i - A_j\|^2) \quad (5)$$

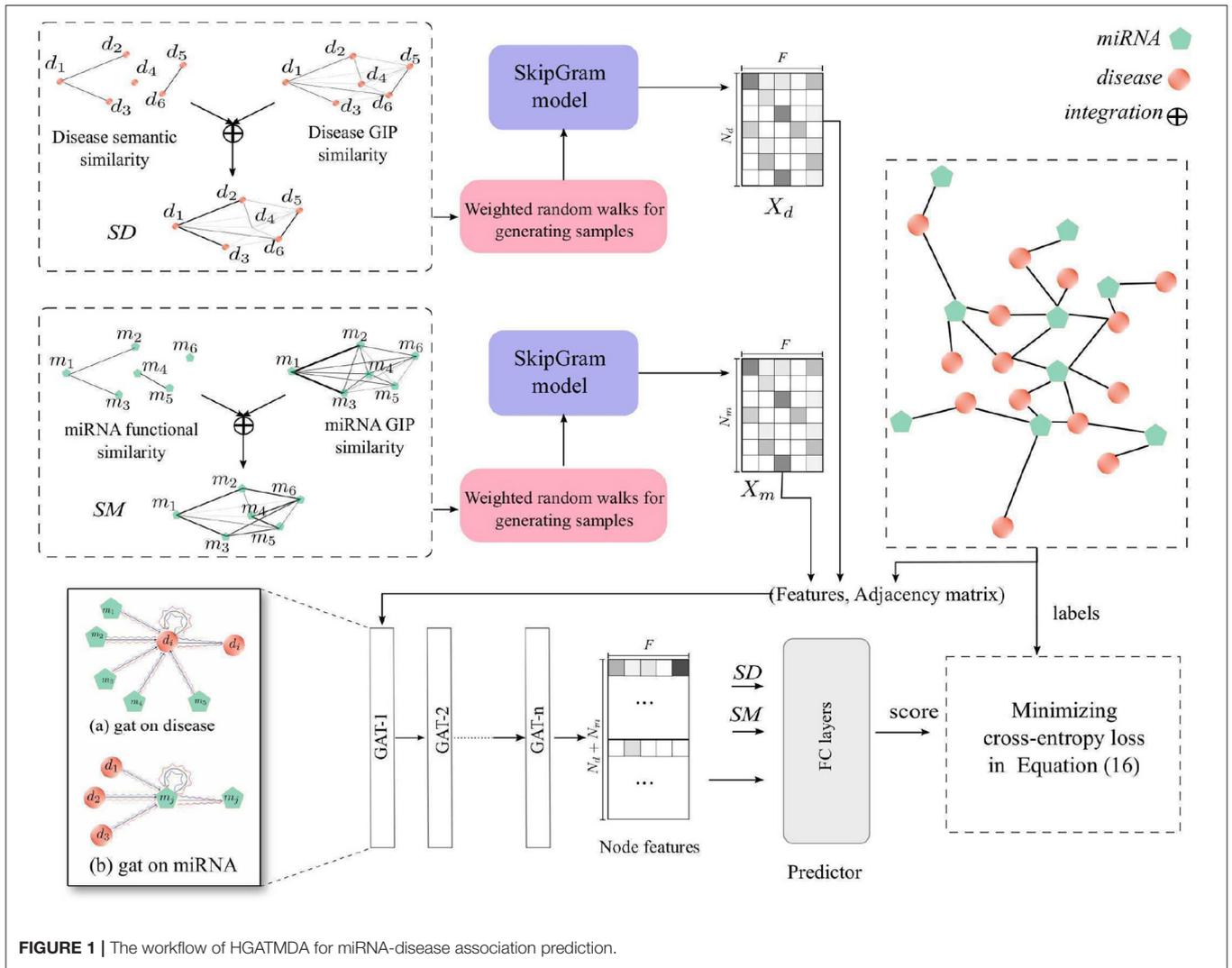


FIGURE 1 | The workflow of HGATMDA for miRNA-disease association prediction.

here,  $\gamma_d$  and  $\gamma_m$  are parameters and can be calculated by following forums:

$$\gamma_d = \frac{\gamma'_d}{\frac{1}{N_d} \sum_{i=1}^{N_d} \|A_{\cdot i}\|^2} \quad (6)$$

$$\gamma_m = \frac{\gamma'_m}{\frac{1}{N_m} \sum_{i=1}^{N_m} \|A_{\cdot i}\|^2} \quad (7)$$

Based on the previous work (van Laarhoven et al., 2011), we set the bandwidth parameters  $\gamma'_d$  and  $\gamma'_m$  to 1 in this paper.

### 2.6. Integrated Similarities for Disease and miRNA

We observed that there was no overlap between DAGs for many diseases, resulting in zero for many elements in disease semantic similarity  $SS$  and miRNA functional similarity  $FS$ . While all entries in the GIP similarities  $KD$  and  $KM$  are non-zero. Therefore, we integrated GIP similarities with  $SS$  and  $FS$

as follows:

$$SD(d_i, d_j) = \frac{SS(d_i, d_j) + KD(d_i, d_j)}{2} \quad (8)$$

$$SM(m_i, m_j) = \frac{FS(m_i, m_j) + KM(m_i, m_j)}{2}. \quad (9)$$

### 2.7. Weighted DeepWalk for Node Representation

DeepWalk (Perozzi et al., 2014) is an algorithm that can learn the representations of vertices in graphs, inspired by the well-known unsupervised feature learning framework word2vec (Mikolov et al., 2013). Given a graph  $G = (V, E)$ ,  $V$  denotes all vertices in the graph, and  $E$  represents the edge or transaction matrix in the graph. In order to generate the corpus in graphs, a vertex  $v_i \in V$  is uniformly selected as the root, and then a random sampling is used from the neighbors as the next hop. After generating samples, DeepWalk applies SkipGram model to learn the representation of each vertex  $v_i$ , denotes as  $\Phi(v_i) \in \mathbb{R}^d$ . Usually, the dimension size  $d$  is used as a hyper-parameter.

We applied a variant DeepWalk, called weighted DeepWalk, to learn disease and miRNA representations. The element in  $SD$  and  $SM$  can be used as the weight of an edge between two vertices of the disease-disease sub-graph or the miRNA-miRNA sub-graph. Here we used the element in  $SD$  and  $SM$  to denote the probability of the current vertex walking to other vertices in the sub-graph. Given a vertex  $v_i$ , we define a walk sequence  $W_{v_i} = \{v_i, \dots\}$  to represent the vertices passed by a weighted random walk starting at the vertex  $v_i$ . We denote the transition possibility of next hop as follows:

$$p(v_j|v_i) = p_{ij} \tag{10}$$

where  $v_i, v_j \in V$  are the vertices in the sub-graph, and  $v_j \in \text{Ner}_{v_i}$  is in the neighbors of vertex  $v_i$ .  $p_{ij} \in SD$  or  $p_{ij} \in SM$  is the weight of an edge from  $v_i$  to  $v_j$ .

After this processing, we obtained two samples  $\{W_{d_1}, W_{d_2}, \dots\}$  and  $\{W_{m_1}, W_{m_2}, \dots\}$  as the corpus, Each sample is a sentence, and each vertex is a word. For learning the representations of diseases and miRNAs, we used DeepWalk to construct two SkipGram models for learning node representations. There are hundreds of diseases and miRNAs in the respective corpus. Therefore, hierarchical softmax was used for accelerating the training process. We set the dimension size as  $F$  in both disease and miRNA representations, and the learned representations of diseases and miRNAs can be denoted as  $X_d = \{d_1, d_2, \dots, d_{N_d}\}, d_i \in \mathbb{R}^F, X_m = \{m_1, m_2, \dots, m_{N_m}\}, m_i \in \mathbb{R}^F$ .  $F$  is the dimension size.

### 2.8. Graph Attention Networks for Node Feature Aggregation

In this section, we further constructed a GAT model in the miRNA-disease sub-graph with node features learned from the previous section, to refine dense vectors of miRNAs and diseases. In consequence, these node features contain global heterogeneous graph structure information. We denote the miRNA-disease sub-graph as  $G_{m-d} = (V, E)$ , and the number of nodes is  $N_d + N_m$ , node  $v_i \in V = \{v_1, \dots, v_{N_d}, v_{N_d+1}, \dots, v_{N_d+N_m}\}$ , edges  $(v_i, v_j) \in E = \{A_{ij} = 1 \in A\}$ . The initial node features in the miRNA-disease sub-graph are defined as  $X = [X_d, X_m]$ , and  $x_i \in \mathbb{R}^F$  is the feature of the  $i$ -th node in the graph. GAT uses multi-head attention mechanism to compute the contributions of neighbors of vertex  $v_i$  to itself. We denote the input of  $l$ -layer of our GAT model as  $H^{(l)} = \{h_1^{(l)}, h_2^{(l)}, \dots, h_N^{(l)}, h_i^{(l)} \in \mathbb{R}^{F^{(l)}}$ , where  $N = N_d + N_m$  is the number of nodes, and  $F^{(l)}$  denotes the input dimension size of each node. Here,  $H^{(0)} = X$  is the input features of the GAT model. We define the output of  $l$ -layer as  $H^{(l+1)} = \{h_1^{(l+1)}, h_2^{(l+1)}, \dots, h_N^{(l+1)}, h_i^{(l+1)} \in \mathbb{R}^{F^{(l+1)}}$ . For each node, we first compute the important score from the neighbor node  $j$  to node  $i$

$$e_{ij}^{(l)} = a(\mathbf{W}^{(l)}h_i^{(l)}, \mathbf{W}^{(l)}h_j^{(l)}) \tag{11}$$

where  $\mathbf{W}^{(l)}$  is a parameter matrix  $W^{(l)} \in \mathbb{R}^{F^{(l+1)} \times F^{(l)}}$ , and  $a$  is a one layer feed-forward neural network. Then, we normalize the

neighborhood important scores of node  $i$  by the softmax function as follows:

$$\alpha_{ij}^{(l)} = \text{softmax}_j^{(l)}(e_{ij}^{(l)}) = \frac{\exp(e_{ij}^{(l)})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik}^{(l)})} \tag{12}$$

where  $\mathcal{N}_i$  is the neighborhood node set of node  $i$ , including node  $i$  itself. Finally, we can use these scores to calculate the new features of node  $i$  by aggregating information from its neighbors:

$$h_i^{(l+1)} = \sigma \left( \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(l)} \mathbf{W}^{(l)} h_j^{(l)} \right) \tag{13}$$

where  $\sigma$  denotes a non-linear activation function, such as LeakyReLU. The power of GAT is benefit from the multi-head attention mechanism, we apply  $K$  independent attention of node  $i$  on its neighborhood, and the output of node features is as follows:

$$h_i^{(l+1)} = \sigma \left( \bigwedge_{k=1}^K \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(l,k)} \mathbf{W}^{(l,k)} h_j^{(l)} \right) \tag{14}$$

where  $\bigwedge$  denotes concatenation or averaging. In our paper, all GAT layers are used concatenation, except averaging for last layer. A GAT carries out the first-order neighborhood information aggregation, and the graph convolution on multi-layers realizes multi-order neighborhood aggregation. As the number of training iterations increases, the node representations can obtain the structure information of the miRNA-disease sub-graph. Together with the node features obtained from miRNA-miRNA and disease-disease sub-graphs, the final node features of  $m_i \in \mathbb{R}^F$  and  $d_j \in \mathbb{R}^F$  contain rich structure information of the global heterogeneous graph.

### 2.9. Potential miRNA-Disease Associations Prediction

At last, we designed a scoring function that can calculate the correlation score between a pair of miRNA and disease. We first integrated node features with raw features in  $SD$  and  $SM$ . In order to transform the raw features in  $SD$  and  $SM$  to the same dimensions as the node features, we introduced projection parameters  $W_d \in \mathbb{R}^{N_d \times F}$  and  $W_m \in \mathbb{R}^{N_m \times F}$ . Then, we can define the correlation score as follows:

$$f(d_i, m_j) = FC([g(d_i, (SD \times W_d)_i), g(m_j, (SM \times W_m)_j)]) \tag{15}$$

where  $FC$  is a fully-connected neural networks, and the details will be discussed in the section 3.3.  $g(\cdot)$  represents an accumulation function, such as  $concat(\cdot)$ , i.e., a concatenation of node features and raw features, or  $sum(\cdot)$ , i.e., summation of node features and raw features. Our model is trained by minimizing the cross-entropy loss and L2 regularization.

$$\begin{aligned} \mathcal{L}(X, Y, \Theta) = & - \sum_{(d_i, m_j) \in \{G^+ \cup G^-\}} y \log f(d_i, m_j) \\ & + (1 - y) \log(1 - f(d_i, m_j)) + \lambda \|\Theta\|^2 \end{aligned} \tag{16}$$

**TABLE 1** | Results of five-fold CV based on HMDD v2.0.

Test fold	Metrics					
	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)	AUPR (%)
1	87.15	94.08	89.87	87.49	94.57	94.07
2	87.57	94.08	89.41	87.79	94.50	94.07
3	86.69	93.75	86.65	86.69	94.09	93.74
4	87.29	94.21	95.30	88.24	95.15	94.20
5	86.37	94.25	88.95	86.71	94.45	94.24
Average	87.02	94.07	90.04	87.39	94.55	94.07

where  $\Theta$  denotes parameters of our GAT model,  $\mathcal{G}^+$  is the set of known associations between miRNA and disease, and  $\mathcal{G}^-$  is the same number set of unknown associations between miRNA and disease from negative sampling, which we will discuss the impact in the later section.  $X$  denotes the vector form of the miRNA-disease pairs in  $\mathcal{G}^+ \cup \mathcal{G}^-$  and  $y \in Y$  represents the corresponding label.

### 3. RESULTS

#### 3.1. Datasets and Experimental Details

We used benchmark datasets of known miRNA-disease associations, including HMDD, dbDEMC v2.0 (Yang et al., 2017), and miR2Disease (Jiang et al., 2008). Mesh dataset is also used for computing disease semantic similarity. To evaluate the prediction performance of our method, we compared the experimental results with other state-of-the-art methods. Moreover, we also performed case studies on three common human diseases such as breast neoplasms, lung neoplasms and kidney neoplasms for further evaluation. HMDD v2.0 is used for training, and dbDEMC v2.0, HMDD v3.2, and miR2Disease are used for validation. We carried out experimental analysis using five-fold cross validation. Specially, all known associations in HMDD v2.0 are taken as positive samples. We first selected equal number of negative samples from a uniform distribution of unknown associations between miRNAs and diseases in HMDD v2.0 dataset. Then, we randomly shuffled all the samples and divided into five equal parts, four of which are in turn used for training and the rest for validation. In addition, we applied several metrics for evaluation, such as Accuracy, Precision, Recall, F1-score. Moreover, we plotted the receiver operating characteristic curve (ROC) and precision-recall curve (PR). Furthermore, the area under the ROC curve (AUC), and the area under the P-R curve (AUPR) are used for analysis the prediction performance quantitatively.

The code of our proposed method is implemented based on the machine learning library PyTorch v1.6.0 (Paszke et al., 2019). We trained the weighted DeepWalk models using GenSim library (Rehurek and Sojka, 2010). For GAT, we used PyTorch Geometric deep learning library (Fey and Lenssen, 2019). Our experiments are run on the Ubuntu 16.04 operating system, with two Intel Xeon CPUs (2.30 GHz, 16 cores), and two Tesla V100 GPUs. The

Adam optimizer is used for training (Kingma and Ba, 2017), with a learning rate of 0.001 and the weight decay is 0.00005.

#### 3.2. Predictive Performance Analysis

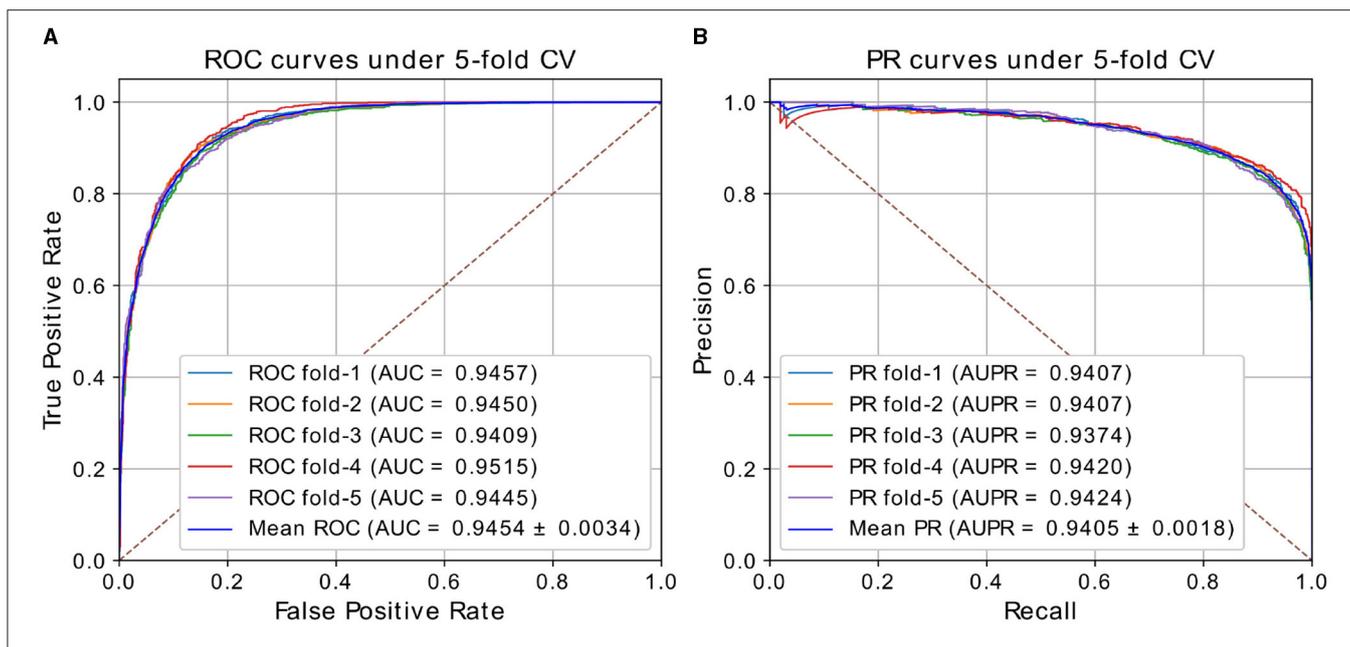
In our experiments, we performed ablation study to analyze the effect of architectures and hyper-parameters, then selected the appropriate parameters. Details of the choices will be discussed in the section 3.3. We chose the model with window size of 5, walk length of 20 in weighted Deepwalk, 2 layers in GAT, and 3-layer FC as predictor. The results of our best model under five-fold CV based on HMDD v2.0 dataset are shown in **Table 1** and **Figure 2**. HGATMDA achieves the average prediction Accuracy of 87.02%, Precision of 94.07%, Recall of 90.04%, and F1-score of 87.39%. The threshold used in these metrics is 0.5. The mean values of AUC and AUPR are  $94.54 \pm 0.34$  and  $94.05 \pm 0.18\%$ , respectively. In summary, the results showed that HGATMDA can significantly promote the performance of predicting miRNA-disease associations.

#### 3.3. Ablation Experiments

In order to further demonstrate the prediction performance of our model, we carried out extensive experiments under different architectures and hyper-parameters, analyzed how the design of sub-model and the choice of hyper-parameters have different effect on the final performance of the proposed model. Five-fold CV on HMDD v2.0 dataset was used to evaluate the model sensitivity of architectures and hyper-parameters in the experiments. The following discussion followed the structure in **Table 2** and **Figures 3, 4**. Dot product Predictor denotes the vector dot product of miRNA  $m_i$  and disease  $d_j$  as the predictor. FC Predictor represents a fully-connected neural networks, as shown in the Equation (15). GAT (untrained) uses randomly weights of neighbors without training, and Raw features denotes integrated similarities of miRNA and disease.

##### 3.3.1. Feature Embedding

Recall that our model adopted weighted DeepWalk to learn representations of miRNAs and diseases from miRNA-miRNA sub-graph and disease-disease sub-graph. We experimented with different feature size of {32,64,128,256} and window size {3,5,7,9} for comparison. Large training windows tend to learn more information from more nodes in the walk, while small training windows do the opposite. HGATMDA obtains the best



**TABLE 2 |** Ablation experiments of different architectures.

Test model	AUC (%)	AUPR (%)
Weighted DeepWalk + Dot product Predictor	73.48	77.97
Weighted DeepWalk + GAT (untrained) + Dot product Predictor	89.78	88.59
Weighted DeepWalk + GAT + Dot product Predictor	92.17	91.52
Raw features + GAT + Dot product Predictor	72.68	69.27
Weight DeepWalk + Raw features + GAT + Dot product Predictor	93.52	93.09
Weight DeepWalk + Raw features + GAT + FC Predictor	<b>94.03</b>	<b>93.51</b>

The variants of the model are listed in the first column. We test each model variant for 10 times under five-fold CV on HMDD v2.0 dataset and report averages. Bold values indicates highlight our results.

performance at window size of 5, and the performance slightly decreases with the increase or decrease of the window size (Figure 3A).

Feature size not only affects the representations of miRNAs and diseases in the SkipGram model, but also affects the final prediction performance together with other parameters. We evaluated the effect of feature size using five-fold CV, and mainly focused on values of AUC and AUPR. As shown in Figure 3B, HGATMDA performs best performance when feature size is set to 64 dimensions.

### 3.3.2. Graph Attention Networks

Representations of miRNAs and diseases are further extract by GAT from miRNA-disease heterogeneous sub-graph. The power of GAT is a multi-head mechanism, we conducted experiments

with different number of attention heads. We chose a 2-layer of GAT model, and each layer used same number of heads for comparison. The results are shown in Figure 3C. We found that more heads slightly increased the AUC and AUPR values.

The number of layer is another factor that affects the performance of HGATMDA. We investigated the depth of GAT model used in HGATMDA. Similar to the phenomenon observed in the classification experiments (Kipf and Welling, 2017), we found that the performance of HGATMDA decreased when the number of layer was set to greater than 2. Therefore, we chose 2-layer of GAT with 16 heads as our default.

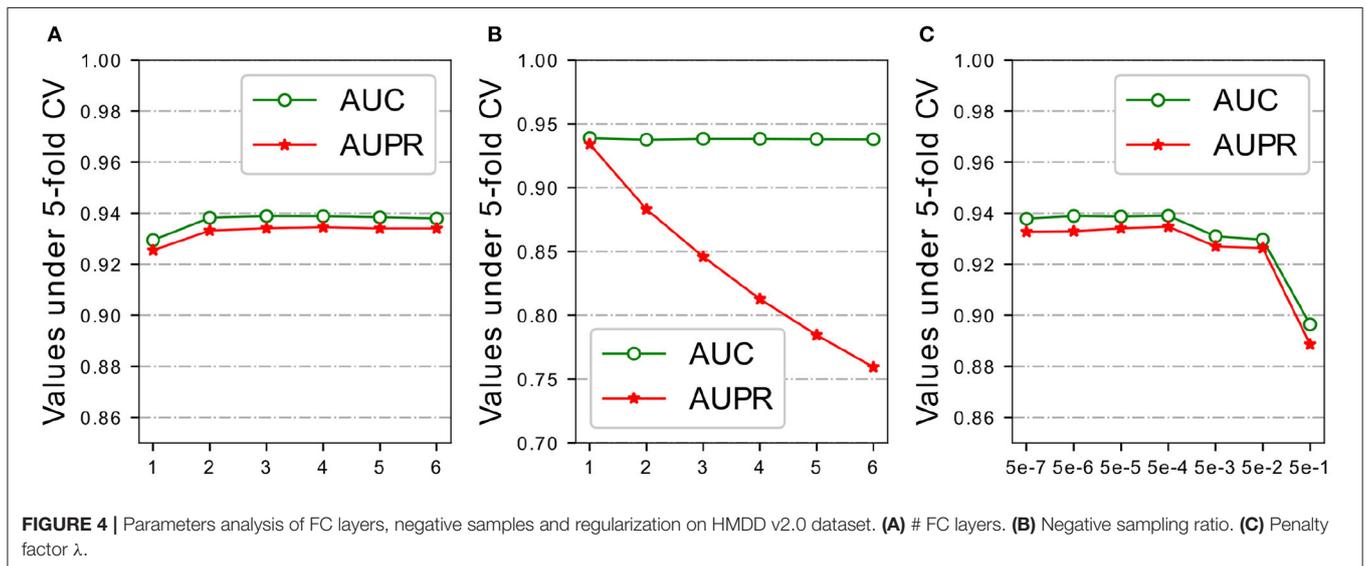
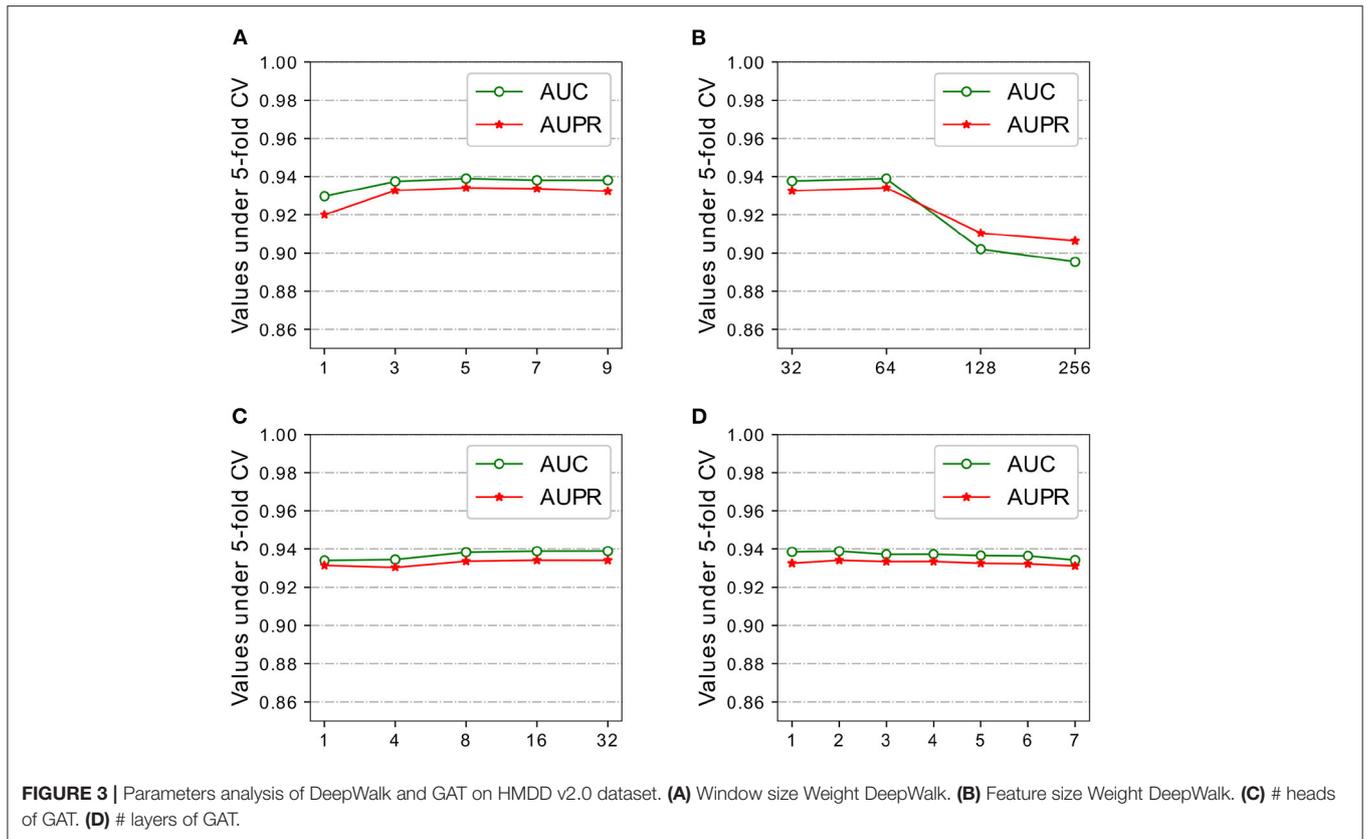
In the Figure 3D, we can see that using only an untrained GAT architecture greatly improves the prediction performance, achieves 89.78% of AUC and 88.59% of AUPR, more than 10% higher than without GAT.

### 3.3.3. FC Layers

The predictor used in HGATMDA is a fully-connected neural networks. Given a pair miRNA and disease, the concatenation of representations of the miRNA and disease is fed into the FC layers, to calculate the correlation score between them. We investigated the influence of number of layers in the predictor. The results are shown in Figure 4A, we can see that HGATMDA achieves the better performance when the number of layers is greater than two. Too deep neural network can not improve the predictive performance of HGATMDA. Considering the computation cost, we chose a three-layer neural networks as the predictor in HGATMDA.

### 3.3.4. Negative Samples

There is lack of prior experimental evidence of non-associations between miRNAs and diseases in the HMDD dataset, we



implemented a sampling strategy to randomly generate negative samples from unknown associations in the dataset of HMDD v2.0. As sampling analysis used in the previous study (Long et al., 2020), we experimentally used different sampling ratios to evaluate the effect. Recall that the number of known associations is 5,430, then the number of negative samples is

{1,0,2,0,3,0,4,0,5,0,6,0} times the number of positive samples. As shown in **Figure 4B**, AUC value could not be improved by sampling more negative samples, and the AUPR value is decreased rapidly. More negative samples may result in an imbalance between positive and negative samples, and P-R curve is more sensitive to the difference of category

**TABLE 3 |** The AUCs comparison with other state-of-the-art methods under five-fold CV based on HMDD v2.0.

Models	AUC (%)
PBMDA (You et al., 2017)	91.72
GRNMF (Xiao et al., 2018)	86.90
MDHGI (Chen et al., 2018e)	87.94
BNPMDA (Chen et al., 2018d)	89.80
MCLPMDA (Yu et al., 2019)	93.25
NIMCGCN (Li et al., 2020a)	92.91
GAEMDA (Li et al., 2020b)	93.56
Our-best	<b>94.52</b>
Our-average	<b>93.88</b>

*Bold values indicates highlight our results.*

distribution. Therefore, we set the sampling ratio to 1.0 in HGATMDA.

### 3.3.5. Regularization

Recall that in Equation (16), we introduced the L2 regularization with a penalty factor  $\lambda$  to improve the generalization ability of the model. We conducted experiments with different values of  $\lambda$  to test the effect on prediction performance of the model under five-fold CV. The area under ROC curve and the area under P-R curve are selected as metrics for evaluation. The values we used in this experiment are {0.5, 0.05, 0.005, 0.0005, 0.00005, 0.000005, 0.0000005}. As shown in **Figure 4C**, We can see that  $\lambda$  is set greater than 0.005, the AUC value declines significantly. Especially, AUC value and AUPR value are both drop below 90% if  $\lambda$  is set as 0.5. When  $\lambda$  is set to 0.0005 or 0.00005, our model achieves better performance. Our model obtains best AUPR at 0.0005, so we finally choose it as our default value.

## 3.4. Comparison With Other Methods

In this section, we compared the prediction performance of our model with other state-of-the-art methods, including PBMDA (You et al., 2017), GRNMF (Xiao et al., 2018), MDHGI (Chen et al., 2018e), BNPMDA (Chen et al., 2018d), MCLPMDA (Yu et al., 2019), NIMCGCN (Li et al., 2020a), and GAEMDA (Li et al., 2020b). We noted that different evaluation metrics and datasets are used in these methods. For fair comparison, AUC values under five-fold CV are selected based on HMDD v2.0 in all these studies for comparison. It is worth noticing that the AUCs reported in these papers are the best values. Therefore, we ran the five-fold CV for 100 times and picked up the best and average AUCs for comparison. The results are shown in **Table 3**. We can see that our model achieves the best AUC performance among these methods, with the best AUC and average AUC of 94.52 and 93.88%, respectively. In particular, NIMCGCN and GAEMDA are GCN-based methods, and AUCs of our best model are 0.96 and 1.61% higher than these two methods, respectively. This further shows that HGATMDA obtains better performance than other methods.

**TABLE 4 |** Top 50 predicted miRNAs related to Breast Neoplasms based on HMDD v2.0.

Rank	miRNA	Evidence	Rank	miRNA	Evidence
1	hsa-mir-17	I; II; III	26	hsa-let-7c	I; II
2	hsa-mir-92a	I; II	27	hsa-mir-125b	I; II; III
3	hsa-mir-1	I; II	28	hsa-mir-30a	I; II; III
4	hsa-mir-34a	I; II	29	hsa-mir-143	I; II; III
5	hsa-mir-29a	I; II	30	hsa-mir-122	I; II
6	hsa-mir-29c	I; II; III	31	hsa-mir-181a	I; II; III
7	hsa-mir-15a	I; II	32	hsa-mir-196a	I; II; III
8	hsa-mir-18a	I; II; III	33	hsa-mir-214	I; II
9	hsa-mir-19b	I; II	34	hsa-mir-7	I; II; III
10	hsa-mir-155	I; II; III	35	hsa-mir-29b	I; II; III
11	hsa-mir-145	I; II; III	36	hsa-mir-222	I; II; III
12	hsa-mir-16	I; II	37	hsa-mir-200c	I; II; III
13	hsa-let-7a	I; II; III	38	hsa-mir-9	I; II
14	hsa-mir-133b	I; II	39	hsa-mir-100	I; II
15	hsa-mir-20a	I; II; III	40	hsa-mir-210	I; II; III
16	hsa-mir-133a	I; II	41	hsa-mir-200b	I; II; III
17	hsa-mir-221	I; II; III	42	hsa-mir-223	I; II
18	hsa-mir-146a	I; II; III	43	hsa-mir-195	I; II; III
19	hsa-mir-126	I; II; III	44	hsa-let-7b	I; II
20	hsa-mir-142	I; II	45	hsa-mir-34c	I; II
21	hsa-mir-21	I; II; III	46	hsa-let-7f	I; II; III
22	hsa-mir-19a	I; II	47	hsa-mir-200a	I; II; III
23	hsa-mir-146b	I; II; III	48	hsa-let-7e	I; II
24	hsa-mir-199a	I; II	49	hsa-let-7d	I; II; III
25	hsa-mir-150	I; II	50	hsa-mir-34b	I; II

*I, II, III denote HMDD v3.2, dbDEMC v2.0, and miR2Disease.*

## 3.5. Case Study

We performed case studies on three common human neoplasms, including *Breast Neoplasms*, *Lung Neoplasms*, *Kidney Neoplasms*, to further evaluate the prediction performance. In these experiments, HMDD v2.0 dataset is used to train our model, and the validation datasets are HMDD v3.2 (Huang et al., 2019a), dbDEMC v2.0 (Yang et al., 2017), and miR2Disease (Jiang et al., 2008), which are used to verify the candidates. For each specific disease, all known associations in HMDD v2.0 are taken as positive samples, while negative samples are chosen from all unknown miRNA-disease associations except the particular disease related. In the prediction phase, all candidate miRNAs are finally ranked by their correlation scores, which are the last layer of our model with sigmoid activation function, to present how much a miRNA associated with the specific disease. In addition, Top 50 candidates are listed and checked whether they are verified in the validation datasets.

We conducted the first case study for *Breast Neoplasms*. It is reported that Breast cancer is the leading cancer among women worldwide, and can arise for a wide number of reasons. It usually happens when cells in breast tissue grow and divide out of control. As shown in **Table 4**, 50 out of the top 50 candidates are confirmed in HMDD v3.2, dbDEMC v2.0, or miR2Disease.

**TABLE 5 |** Top 50 predicted miRNAs related to Lung Neoplasms based on HMDD v2.0.

Rank	miRNA	Evidence	Rank	miRNA	Evidence
1	hsa-mir-20a	I; II; III	26	hsa-mir-9	I; II
2	hsa-mir-125b	I; II; III	27	hsa-mir-203	I; II; III
3	hsa-mir-146b	I; II; III	28	hsa-mir-29c	I; II; III
4	hsa-mir-17	I; II; III	29	hsa-mir-155	I; II; III
5	hsa-mir-181a	I; II; III	30	hsa-mir-34c	I; II; III
6	hsa-mir-145	I; II	31	hsa-mir-142	I; II; III
7	hsa-mir-34a	I; II; III	32	hsa-mir-143	I; II
8	hsa-mir-29b	I; II; III	33	hsa-mir-30a	I; II; III
9	hsa-mir-146a	I; II; III	34	hsa-mir-221	I; II; III
10	hsa-mir-126	I; II; III	35	hsa-mir-133b	I; II; III
11	hsa-let-7a	I; II; III	36	hsa-mir-31	I; II
12	hsa-let-7d	I; II; III	37	hsa-mir-182	I; II; III
13	hsa-mir-92a	I; II; III	38	hsa-mir-34b	I; II; III
14	hsa-let-7i	I; II; III	39	hsa-let-7c	I; II; III
15	hsa-mir-19a	I; II; III	40	hsa-mir-122	I; II; III
16	hsa-mir-19b	I; II	41	hsa-mir-222	I; II; III
17	hsa-let-7b	I; II	42	hsa-mir-200b	I; II; III
18	hsa-let-7g	I; II	43	hsa-mir-18a	I; II
19	hsa-mir-200c	I; II; III	44	hsa-mir-199a	I; II; III
20	hsa-mir-21	I; II	45	hsa-mir-214	I; II
21	hsa-mir-210	I; II	46	hsa-mir-106b	II
22	hsa-let-7f	I; II; III	47	hsa-mir-100	I; II; III
23	hsa-mir-200a	I; II	48	hsa-mir-148a	I; II; III
24	hsa-let-7e	I; II; III	49	hsa-mir-195	I; II; III
25	hsa-mir-16	I; II	50	hsa-mir-101	I; II; III

I, II, III denote HMDD v3.2, dbDEMOC v2.0, and miR2Disease.

**TABLE 6 |** Top 50 predicted miRNAs related to Kidney Neoplasms based on HMDD v2.0.

Rank	miRNA	Evidence	Rank	miRNA	Evidence
1	hsa-mir-21	I; II; III	26	hsa-mir-31	II
2	hsa-mir-146a	II	27	hsa-mir-150	II
3	hsa-mir-155	I; II	28	hsa-let-7a	II; III
4	hsa-mir-34a	I; II	29	hsa-mir-143	II
5	hsa-mir-125b	II	30	hsa-mir-142	II
6	hsa-mir-221	II	31	hsa-mir-200b	II
7	hsa-mir-29a	II; III	32	hsa-mir-181a	II; III
8	hsa-mir-20a	II; III	33	hsa-mir-23a	II
9	hsa-mir-15a	I; II; III	34	hsa-mir-15b	II
10	hsa-mir-16	II	35	hsa-mir-106b	II
11	hsa-mir-17	I; II; III	36	hsa-mir-196a	II; III
12	hsa-mir-29b	II; III	37	hsa-mir-133b	II
13	hsa-mir-92a	II	38	hsa-mir-210	I; II
14	hsa-mir-29c	II; III	39	hsa-mir-181b	II; III
15	hsa-mir-145	II	40	hsa-mir-146b	II
16	hsa-mir-18a	II	41	hsa-mir-30a	II
17	hsa-mir-223	II	42	hsa-mir-24	II
18	hsa-mir-122	II; III	43	hsa-mir-200c	I; II
19	hsa-mir-19b	II; III	44	hsa-mir-9	II; III
20	hsa-mir-126	I; II; III	45	hsa-mir-200a	I; II
21	hsa-mir-199a	I; II; III	46	hsa-mir-182	II
22	hsa-mir-1	II	47	hsa-mir-214	I; II; III
23	hsa-mir-133a	II	48	hsa-mir-148a	II; III
24	hsa-mir-19a	II	49	hsa-mir-195	II
25	hsa-mir-222	II	50	hsa-mir-7	II

I, II, III denote HMDD v3.2, dbDEMOC v2.0, and miR2Disease.

The second case study was implemented for *Lung Neoplasms*. This is a leading cause of cancer deaths among men and women both in United States and the world (Siegel et al., 2020). We applied our model to predict the most relevant miRNAs with *Lung Neoplasms*. The results are shown in the **Table 5**, 50 of top 50 selected miRNAs have been verified in the validation datasets. The last case study we selected was *Kidney Neoplasms*. It is another common disease, and the incidence still continues to increase in the United States (Siegel et al., 2020). We note that there are only 7 known associations related with *Kidney Neoplasms* in HMDD v2.0 dataset. We listed top 50 related miRNAs in **Table 6**, and we can see that 50 out of top 50 candidates are confirmed either in HMDD v3.2, dbDEMOC v2.0 or miR2Disease. In particular, 6 of top 10 related miRNAs are verified in at least two datasets. Therefore, our model can serve as a powerful and effective tool to infer the potential related miRNAs for specific diseases.

#### 4. DISCUSSION

Studies have shown that the occurrence and development of many human diseases are related to the abnormal expression of miRNAs. Traditional biological verification of the interactions

between miRNAs and diseases are time consuming and expensive. Therefore, computational methods of predicting the disease-related miRNAs can accelerate the identification process, and help us understand the potential mechanism of the interactions between miRNAs and diseases.

In this paper, we presented a computational method for miRNA-disease association prediction based on Heterogeneous Graph Attention Networks, which is superior to other state-of-the-art methods. Specially, we first constructed a heterogeneous graph containing miRNAs and diseases using disease semantic similarity, miRNA functional similarity, the GIP kernel similarities, and known associations between miRNAs and diseases. Then, we proposed a novel method based on weighted DeepWalk that can learn dense feature representation of miRNAs and diseases from the miRNA-miRNA and disease-disease sub-graphs. Furthermore, a GAT based model was implemented for further feature extraction from the miRNA-disease heterogeneous sub-graph, followed by a 3-layer fully-connected neural networks as the predictor. We conducted experiments on five-fold CV, case studies and ablation study. The results demonstrated that our proposed method HGATMDA can serve as an efficient and reliable tool for predicting potential relations between miRNAs and diseases, as well as therapeutics and clinical research.

Our model has achieved high predictive performance mainly for the following reasons. First, GAT was applied to extract node features in the heterogeneous miRNA-disease graph, which is very effective and expressive by leveraging multi-head attention mechanism. Second, initial node features extraction were obtained by weighted DeepWalk, combined with similarity integration, to further improve miRNA and disease characterization. Third, FC-based predictor is suitable and reliable for inferring the potential disease-related miRNAs.

However, there is still room for further improvement. Our constructed heterogeneous graph maybe inaccurate due to the computing equations used in disease and miRNA similarities. In the future, we can directly use the Mesh headings and relationships (<https://www.nlm.nih.gov/mesh/meshhome.html>) to build disease-disease sub-graph, and extract node features such as Guo et al. (2021). Furthermore, as our future work, we can introduce more biological information, such as miRNA sequence, miRNA-target, protein-protein, and protein-target interactions, to enrich nodes representations, and construct complex heterogeneous graph (Schlichtkrull et al., 2018) with multiple node and edge types, to enhance the prediction performance of our model.

## REFERENCES

- Ambros, V. (2001). microRNAs: tiny regulators with great potential. *Cell* 107, 823–826. doi: 10.1016/S0092-8674(01)00616-X
- Ambros, V. (2003). MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell* 113, 673–676. doi: 10.1016/S0092-8674(03)00428-8
- Ambros, V. (2004). The functions of animal microRNAs. *Nature* 431, 350–355. doi: 10.1038/nature02871
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/S0092-8674(04)00045-5
- Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002
- Bartel, D. P. (2018). Metazoan microRNAs. *Cell* 173, 20–51. doi: 10.1016/j.cell.2018.03.006
- Chen, X., Clarence Yan, C., Zhang, X., Li, Z., Deng, L., Zhang, Y., et al. (2015). RBMMMDA: predicting multiple types of disease-microRNA associations. *Sci. Rep.* 5:13877. doi: 10.1038/srep13877
- Chen, X., Gong, Y., Zhang, D. H., You, Z. H., and Li, Z. W. (2018a). DRMDA: deep representations-based miRNA-disease association prediction. *J. Cell. Mol. Med.* 22, 472–485. doi: 10.1111/jcmm.13336
- Chen, X., Liu, M. X., and Yan, G. Y. (2012). RWRMDA: Predicting novel human microRNA-disease associations. *Mol. Biosyst.* 8, 2792–2798. doi: 10.1039/c2mb25180a
- Chen, X., Sun, L.-G., and Zhao, Y. (2021). NCMCMDA: miRNA-disease association prediction through neighborhood constraint matrix completion. *Brief. Bioinform.* 22, 485–496. doi: 10.1093/bib/bb159
- Chen, X., Wang, L., Qu, J., Guan, N. N., and Li, J. Q. (2018b). Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* 34, 4256–4265. doi: 10.1093/bioinformatics/bty503
- Chen, X., Wang, L. Y., and Huang, L. (2018c). NDAMDA: Network distance analysis for miRNA-disease association prediction. *J. Cell. Mol. Med.* 22, 2884–2895. doi: 10.1111/jcmm.13583
- Chen, X., Xie, D., Wang, L., Zhao, Q., You, Z. H., and Liu, H. (2018d). BNPMDA: bipartite network projection for miRNA-disease association prediction. *Bioinformatics* 34, 3178–3186. doi: 10.1093/bioinformatics/bty333
- Chen, X., Yan, C. C., Zhang, X., You, Z. H., Huang, Y. A., and Yan, G. Y. (2016). HGMDA: Heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget* 7, 65257–65269. doi: 10.18632/oncotarget.11251

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

CJ and YS conceived the entire project and wrote the paper. CJ and CZ developed the prediction method. CJ and JN collected data and designed the experiments. CJ and YW analyzed the results. All authors read and approved the final manuscript.

## FUNDING

This work was supported by the National Natural Science Foundation of China (grant numbers U19A2064, 61873001, and 61872220), and the Natural Science Foundation of Shandong Province (grant number ZR2020KC022).

- Chen, X., Yin, J., Qu, J., and Huang, L. (2018e). MDHGI: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction. *PLoS Comput. Biol.* 14:e1006418. doi: 10.1371/journal.pcbi.1006418
- Chen, X., Zhu, C. C., and Yin, J. (2019). Ensemble of decision tree reveals potential miRNA-disease associations. *PLoS Comput. Biol.* 15:e1007209. doi: 10.1371/journal.pcbi.1007209
- Chou, C. H., Chang, N. W., Shrestha, S., Hsu, S. D., Lin, Y. L., Lee, W. H., et al. (2016). miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.* 44, D239–D247. doi: 10.1093/nar/gkv1258
- Croce, C. M. (2008). Oncogenes and cancer. *N. Engl. J. Med.* 358, 502–511. doi: 10.1056/NEJMra072367
- Fey, M., and Lenssen, J. E. (2019). Fast graph representation learning with pytorch geometric. *arXiv [Preprint]. arXiv:1903.02428.*
- Friedman, R. C., Farh, K.-H., Burge, C. B., and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research* 19, 92–105. doi: 10.1101/gr.082701.108
- Fu, L., and Peng, Q. (2017). A deep ensemble model to predict miRNA-disease association. *Sci. Rep.* 7, 1–13. doi: 10.1038/s41598-017-15235-6
- Guo, Z. H., You, Z. H., Huang, D. S., Yi, H. C., Zheng, K., Chen, Z. H., et al. (2021). MeSHHeading2vec: a new method for representing MeSH headings as vectors based on graph embedding algorithm. *Brief. Bioinform.* 22, 2085–2095. doi: 10.1093/bib/bba037
- He, L., and Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* 5, 522–531. doi: 10.1038/nrg1379
- Huang, Z., Liu, L., Gao, Y., Shi, J., Cui, Q., Li, J., et al. (2019a). Benchmark of computational methods for predicting microRNA-disease associations. *Genome Biol.* 20, 1–13. doi: 10.1186/s13059-019-1811-3
- Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., et al. (2019b). HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.* 47, D1013–D1017. doi: 10.1093/nar/gky1010
- Ji, C., Gao, Z., Ma, X., Wu, Q., Ni, J., and Zheng, C. (2021). AEMDA: inferring miRNA-disease associations based on deep autoencoder. *Bioinformatics* 37, 66–72. doi: 10.1093/bioinformatics/btaa670
- Jiang, Q., Hao, Y., Wang, G., Juan, L., Zhang, T., Teng, M., et al. (2010). Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst. Biol.* 4(Suppl. 1):S2. doi: 10.1186/1752-0509-4-S1-S2

- Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., et al. (2008). miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* 37(Suppl\_1), D98–D104. doi: 10.1093/nar/gkn714
- Kingma, D. P., and Ba, J. (2017). Adam: a method for stochastic optimization. *arXiv [Preprint]*. arXiv:1412.6980.
- Kipf, T. N., and Welling, M. (2017). Semi-supervised classification with graph convolutional networks. *arXiv [Preprint]*. arXiv:1609.02907.
- Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer* 42, 30–37. doi: 10.1109/MC.2009.263
- Li, J., Zhang, S., Liu, T., Ning, C., Zhang, Z., and Zhou, W. (2020a). Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction. *Bioinformatics* 36, 2538–2546. doi: 10.1093/bioinformatics/btz965
- Li, J. Q., Rong, Z. H., Chen, X., Yan, G. Y., and You, Z. H. (2017). MCMDA: Matrix completion for MiRNA-disease association prediction. *Oncotarget* 8, 21187–21199. doi: 10.18632/oncotarget.15061
- Li, Y., Qiu, C., Tu, J., Geng, B., Yang, J., Jiang, T., et al. (2014). HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* 42, D1070–D1074. doi: 10.1093/nar/gkt1023
- Li, Z., Li, J., Nie, R., You, Z.-H., and Bao, W. (2020b). A graph auto-encoder model for miRNA-disease associations prediction. *Brief. Bioinform.* 22, 1–13. doi: 10.1093/bib/bbaa240
- Long, Y., Luo, J., Zhang, Y., and Xia, Y. (2021). Predicting human microbe-disease associations via graph attention networks with inductive matrix completion. *Brief. Bioinform.* 22, 1–13. doi: 10.1093/bib/bbaa146
- Long, Y., Wu, M., Kwok, C. K., Luo, J., and Li, X. (2020). Predicting human microbe-drug associations via graph convolutional network with conditional random field. *Bioinformatics* 36, 4918–4927. doi: 10.1093/bioinformatics/btaa598
- Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W., et al. (2008). An analysis of human microRNA and disease associations. *PLoS ONE* 3:e3420. doi: 10.1371/journal.pone.0003420
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). “Distributed representations of words and phrases and their compositionality,” in *Advances in Neural Information Processing Systems, NIPS’13* (Red Hook, NY: Curran Associates Inc.), 1–9.
- Pan, X., and Shen, H. B. (2019). Inferring disease-associated microRNAs using semi-supervised multi-label graph convolutional networks. *iScience* 20, 265–277. doi: 10.1016/j.isci.2019.09.013
- Pan, Z., Zhang, H., Liang, C., Li, G., Xiao, Q., Ding, P., et al. (2019). Self-weighted multi-kernel multi-label learning for potential miRNA-disease association prediction. *Mol. Therapy Nucleic Acids* 17, 414–423. doi: 10.1016/j.omtn.2019.06.014
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: an imperative style, high-performance deep learning library. *arXiv [Preprint]*. arXiv:1912.01703.
- Peng, J., Hui, W., Li, Q., Chen, B., Hao, J., Jiang, Q., et al. (2019). A learning-based framework for miRNA-disease association identification using neural networks. *Bioinformatics* 35, 4364–4371. doi: 10.1093/bioinformatics/btz254
- Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). “Deepwalk: online learning of social representations,” in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY). doi: 10.1145/2623330.2623732
- Rehurek, R., and Sojka, P. (2010). “Software framework for topic modelling with large corpora,” in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (Valletta), 45–50.
- Schlichtkrull, M., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., and Welling, M. (2018). “Modeling relational data with graph convolutional networks,” in *Lecture Notes in Computer Science*, eds A. Gangemi, R. Navigli, M.-E. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai, and M. Alam (Heraklion: Springer), 593–607. doi: 10.1007/978-3-319-93417-4\_38
- Shen, Z., Zhang, Y. H., Han, K., Nandi, A. K., Honig, B., and Huang, D. S. (2017). MiRNA-disease association prediction with collaborative matrix factorization. *Complexity* 2017, 1–9. doi: 10.1155/2017/2498957
- Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA Cancer J. Clin.* 70, 7–30. doi: 10.3322/caac.21590
- van Laarhoven, T., Nabuurs, S. B., and Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* 27, 3036–3043. doi: 10.1093/bioinformatics/btr500
- Wang, D., Wang, J., Lu, M., Song, F., and Cui, Q. (2010). Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* 26, 1644–1650. doi: 10.1093/bioinformatics/btq241
- Xiao, Q., Luo, J., Liang, C., Cai, J., and Ding, P. (2018). A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics* 34, 239–248. doi: 10.1093/bioinformatics/btx545
- Xu, J., Li, C. X., Lv, J. Y., Li, Y. S., Xiao, Y., Shao, T. T., et al. (2011). Prioritizing candidate disease miRNAs by topological features in the miRNA target-dysregulated network: case study of prostate cancer. *Mol. Cancer Therapeut.* 10, 1857–1866. doi: 10.1158/1535-7163.MCT-11-0055
- Xuan, P., Dong, Y., Guo, Y., Zhang, T., and Liu, Y. (2018). Dual convolutional neural network based method for predicting disease-related miRNAs. *Int. J. Mol. Sci.* 19:3732. doi: 10.3390/ijms19123732
- Yan, C., Wang, J., Ni, P., Lan, W., Wu, F. X., and Pan, Y. (2019). DNRLMF-MDA: predicting microRNA-disease associations based on similarities of microRNAs and diseases. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 16, 233–243. doi: 10.1109/TCBB.2017.2776101
- Yang, Z., Wu, L., Wang, A., Tang, W., Zhao, Y., Zhao, H., et al. (2017). DbDEM2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res.* 45, D812–D818. doi: 10.1093/nar/gkw1079
- You, Z. H., Huang, Z. A., Zhu, Z., Yan, G. Y., Li, Z. W., Wen, Z., et al. (2017). PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction. *PLoS Comput. Biol.* 13:e1005455. doi: 10.1371/journal.pcbi.1005455
- Yu, S. P., Liang, C., Xiao, Q., Li, G. H., Ding, P. J., and Luo, J. W. (2019). MCLPMDA: A novel method for miRNA-disease association prediction based on matrix completion and label propagation. *J. Cell. Mol. Med.* 23, 1427–1438. doi: 10.1111/jcmm.14048
- Zhang, L., Chen, X., and Yin, J. (2019). Prediction of potential miRNA-disease associations through a novel unsupervised deep learning framework with variational autoencoder. *Cells* 8:1040. doi: 10.3390/cells8091040
- Zhong, Y., Xuan, P., Wang, X., Zhang, T., Li, J., Liu, Y., et al. (2018). A non-negative matrix factorization based method for predicting disease-associated miRNAs in miRNA-disease bilayer network. *Bioinformatics* 34, 267–277. doi: 10.1093/bioinformatics/btx546
- Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., et al. (2020). Graph neural networks: a review of methods and applications. *AI Open* 1, 57–81. doi: 10.1016/j.aiopen.2021.01.001

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Ji, Wang, Ni, Zheng and Su. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.