Meta-path semantic and global-local representation learning enhanced graph convolutional model for disease-related miRNA prediction

Ping Xuan, Xiuju Wang, Hui Cui, Xiangfeng Meng, Toshiya Nakaguchi and Tiangang Zhang *

***Abstract*— Dysregulation of miRNAs is closely related to the progression of various diseases, so identifying disease-related miRNAs is crucial. Most recently proposed methods are based on graph reasoning, while they did not completely exploit the topological structure composed of the higher-order neighbor nodes and the global and local features of miRNA and disease nodes. We proposed a prediction method, MDAP, to learn semantic features of miRNA and disease nodes based on various meta-paths, as well as node features from the entire heterogeneous network perspective, and node pair attributes. Firstly, for both the miRNA and disease nodes, node category-wise meta-paths were constructed to integrate the similarity and association connection relationships. Each target node has its specific neighbor nodes for each meta-path, and the neighbors of longer meta-paths constitute its higher-order neighbor topological structure. Secondly, we constructed a meta-path specific graph convolutional network module to integrate the features of higher-order neighbors and their topology, and then learned the semantic representations of nodes. Thirdly, for the entire miRNA-disease heterogeneous network, a global-aware graph convolutional autoencoder was built to learn the network-view feature representations of nodes. We also designed semantic-level and representation-level attentions to obtain informative semantic features and node representations. Finally, the strategy based on the parallel convolutional-deconvolutional neural networks were designed to enhance the local feature learning for a pair of miRNA and disease nodes. The experiment results showed that MDAP outperformed other state-of-the-art methods, and the ablation experiments demonstrated the effectiveness of MDAP's major innovations. MDAP's ability in discovering potential disease-related miRNAs was further analyzed by the case studies over three diseases.**

***Index Terms*— semantic feature learning based on various meta-paths, higher-order neighbor topological structure integration, pairwise local feature learning, disease-related miRNA prediction, deep learning.**

The work was supported by the Natural Science Foundation of China (62172143, 62372282); STU Scientific Research Initiation Grant (NTF22032); and the Natural Science Foundation of Heilongjiang Province (LH2023F044).

Ping Xuan is with the Department of Computer Science, Shantou University, Shantou, 515063, China.

Xiuju Wang is with the School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China.

Hui Cui is with the Department of Computer Science and Information Technology, La Trobe University, Melbourne, 3083, Australia.

Xiangfeng Meng is with the School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China.

Toshiya Nakaguchi is with the Center for Frontier Medical Engineering, Chiba University, Chiba 2638522, Japan.

Tiangang Zhang is the corresponding author and he is with the School of Computer Science and Technology and the School of Mathematical Science, Heilongjiang University, Harbin, 150080, China (e-mail: zhang@hlju.edu.cn).

I. Introduction

MICRORNAS (miRNAs) are small non-coding RNAs with a length of 22–24 nucleotides [1], [2]. Abnormalities in miRNAs have been associated with the development of various diseases [3], [4]. For this reason, identifying disease-associated miRNAs can help explore the pathogenesis of diseases and facilitate their diagnosis and treatment. Calculating the likelihood of predicting miRNAs associated with diseases can assist in the selection of candidate miRNAs for each disease, reducing the cost and time required for biological experiments.

The three major categories of existing miRNA-disease association prediction methods are as follows. The biological premise underlying the first group of methods is that miRNAs are more likely to be associated with comparable diseases when their functions are similar [5]. Thus, the functional similarity of miRNAs is determined using the diseases related to the two miRNAs [6]. Xuan *et al.* [7] utilized the miRNA family and cluster information to measure the miRNA similarities and identified the potential disease-related miRNAs by the top k most similar neighbors. The prediction models based on random walk on the similarity networks were constructed [8], [9]. However, it was difficult for them to predict the candidate miRNAs for the new diseases.

The strategies in the secondary category add extra diseases information to establish miRNA-disease heterogeneous networks to incorporate miRNA-disease association, miRNA similarity and disease similarity. To infer disease-associated candidate miRNAs, Chen *et al.* proposed an association score estimation strategy based on k-nearest neighbors [10], and You *et al.* proposed a depth-first search method [11]. Wang *et al.* introduced miRNA sequence information into their model and used a logic tree classifier for miRNA-disease association prediction [12]. Certain matrix decomposition-based techniques are also available [13]–[17]. However, the methods mentioned

above did not fully integrate the context relationships among the various connections in the miRNA-disease heterogeneous network. The miRNA-disease association score can also be obtained using random walk [18]–[21], regularized least squares [22], miRNA-disease network projection [23], [24] and label propagation [25]. These methods built the shallow prediction models and they failed to learn the deep and representative features of the miRNA and disease nodes.

The third category uses deep learning techniques to combine data on miRNAs and diseases to more precisely identify potential miRNA candidates linked to diseases. Convolutional neural network (CNN) based prediction models were proposed by Peng *et al.* [26] and Xuan *et al.* [27]. These models did not completely take advantage of the more local and detailed features in the feature map of a pair of miRNA and disease nodes. In addition, graph convolutional neural networks (GCN) [28]–[30], generative adversarial networks [31], fully connected autoencoder [32], variational graph autoencoder [33], and deep belief networks [34] have been used to model miRNA-disease association candidate prediction. These prediction models achieved decent prediction performances as they learned the deep and representative features of the miRNA and disease nodes. However, they rarely focused on integrating the meta-path semantic information which is specific to each type of nodes, and learning the global context relationships within the miRNA-disease network.

We suggest a novel prediction method in this study called MDAP for learning meta-path semantics of multiple types of nodes and higher-order structures in a miRNA-disease heterogeneous network. The method leverages topological information of the entire network and diverse node-pair attribute information. The following is a summary of our model's contributions.

- A heterogeneous network is firstly constructed and it consists of the miRNA and disease nodes, and the similarity connections and the association ones among them. There are long-distance dependencies among the nodes in the miRNA-disease network. Thus, we build multiple meta-paths to form the topological structures composed of high-order neighbors and learn the dependencies among these neighbors.
- The heterogenous network contains the miRNA nodes and the disease ones, and each type of nodes has their own meta-paths. The multiple meta-paths for the miRNA nodes and the ones for the disease nodes are established respectively, and these meta-paths reflect the biological characteristics that the miRNAs are more likely to associate with the diseases. A meta-path-sensitive learning module is constructed based on graph convolutional networks, and it may learn the semantic features of each miRNA (disease) node for each meta-path. Since the multiple meta-paths of the miRNA (disease) nodes have different contributions to the semantic feature learning, the semantic-level attention is designed to discriminate their contributions.
- The entire heterogeneous network contains the various similarity and association connections and these connections have global context relationships. We constructed a global-sensitive graph convolutional autoencoder to encode these relationships and form the network-view representations of the miRNA and disease nodes. Furthermore, the attention at the representation level is proposed to adaptively fuse the semantic representations and the network-view representations of multiple types of nodes.
- The detailed features from the local regions of the feature map of a pair of miRNA-disease nodes are important auxiliary information for miRNA-disease association prediction. A module based on convolution and deconvolution networks is constructed to stretch the original feature map and learn the more local and detailed pairwise features. A comprehensive comparison with several recent prediction methods and the case studies over 3 diseases indicate our method's superior performance and its ability in retrieving the potential disease-related miRNAs.

II. Materials and Methods

We proposed an association prediction model, MDAP, to predict potential miRNAs linked to a certain disease (Figure 1). First, we learned semantic information about the meta-paths specific to miRNA- and disease-type nodes, as well as the higher-order structure of the heterogeneous network based on meta-path-specific graph convolution (GCN) (Figure 1(a)). A global-aware graph convolutional autoencoder was constructed to integrate the global context relationships of all the nodes in the entire heterogeneous network (Figure 1(b)). A node representation-level attention mechanism was designed to self-adapt to fuse meta-path semantic representations and network-view representations (Figure 1(c)). In addition, we designed a parallel convolution and deconvolution network to encode the attributes of a pair of miRNA and disease nodes (Figure 1(d)).

A. Dataset

The Human MicroRNA Disease Database contains 7908 miRNA-disease correlations [35], and these associations cover 793 miRNAs and 341 diseases. We created directed acyclic graphs (DAGs) using the terminology for diseases from the American Library of Medicine [36]. Based on their DAGs, the semantic similarity between the two diseases was determined [6].

B. Double-layer heterogeneous network and matrix

Definition 1. MiRNA-disease heterogeneous network. A miRNA-disease heterogeneous network is defined as $G=(V,E,W)$. The set of nodes $V=\left\{V^{mir}\cup V^{dis}\right\}$, where V^{mir} denotes N_m miRNA nodes $\{mir_1, mir_2, ..., mir_{N_m}\}$ and V^{dis} denotes N_d disease nodes $\{dis_1, dis_2, ..., dis_{N_d}\}$. Edge $e_{ij}\in E$ with weight $w_{ij}\in W$ links a pair of nodes, $v_i, v_j \in V$. There is a node mapping function $\delta: V\rightarrow\vartheta$ and an edge mapping function $\Delta: E\rightarrow\varpi$. ϑ and ϖ represents the node types and the edge types, and $|\vartheta|+|\varpi|>2$.

There are two kinds of connections between nodes in G. One is the intra-layer connection between nodes of the same type, such as miRNA-miRNA similarity connection and

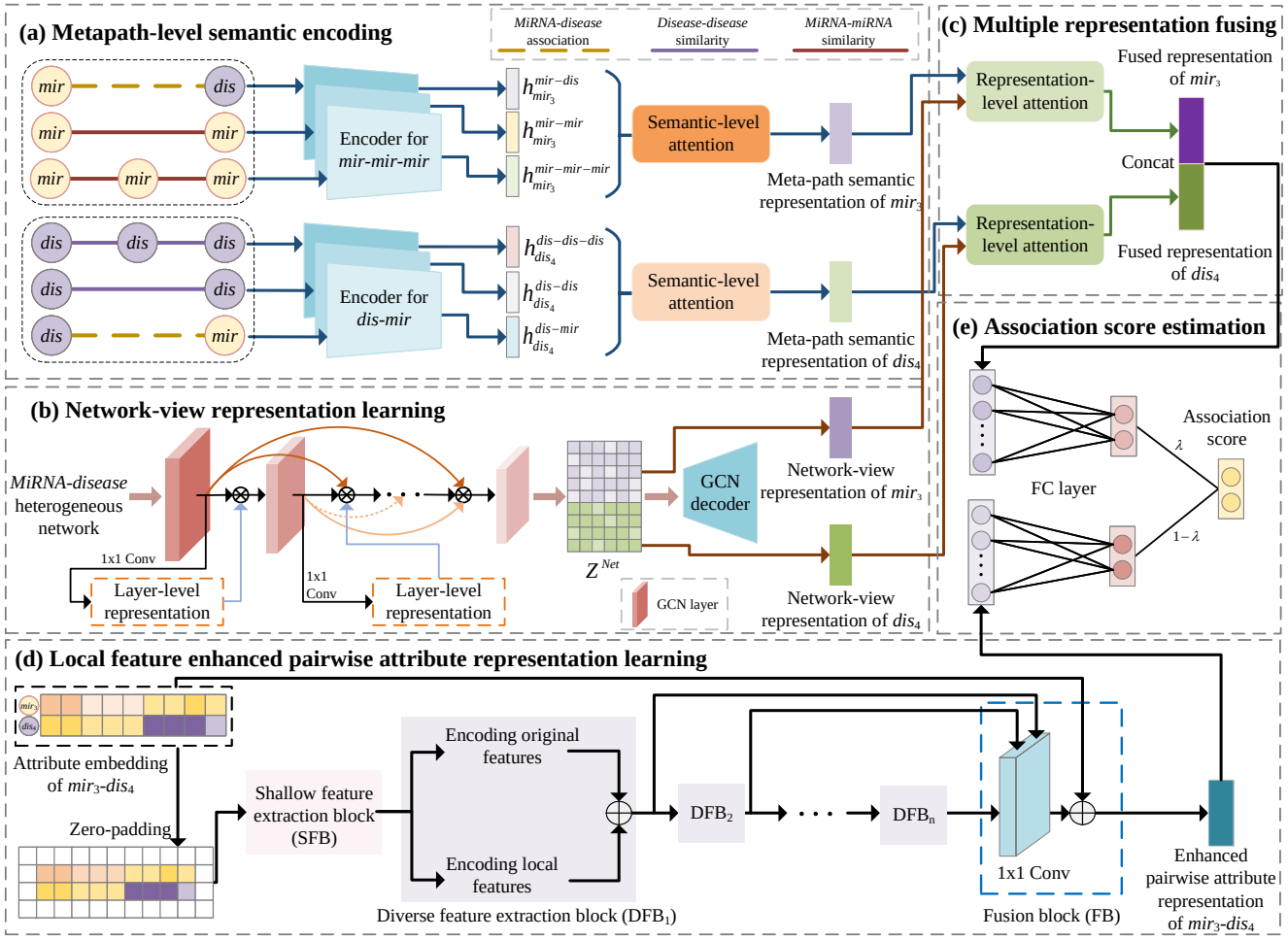

Fig. 1. Framework of the proposed MDAP model. **(a)** learn meta-path semantic representation of each miRNA (disease) node. **(b)** learn the network-view representations of nodes from the entire heterogeneous network perspective. **(c)** fuse multiple representations of nodes. **(d)** learn the local feature enhanced pairwise attribute representation. **(e)** final fusion and association score estimation.

disease-disease similarity connection. The other type is the inter-layer connection, including miRNA-disease association connection. We define $W = \left(A^{sim}, B^{mir-dis}\right)$, where A^{sim} is the intra-layer relationship matrix and $B^{mir-dis}$ is the inter-layer relationship matrix.

A^{sim} includes miRNA similarity and disease similarity matrices, defined as

$$A^{sim} = \begin{cases} A^{mir} \in R^{N_m \times N_m}, if\ v_i, v_j \in V^{mir} \\ A^{dis} \in R^{N_d \times N_d}, if\ v_i, v_j \in V^{dis} \end{cases}, \quad (1)$$

A^{mir} records similarities between N_m miRNAs. The more similar the function of two miRNAs, the more likely they are to be associated with similar diseases. Based on this biological premise and influenced by the strategy of Wang *et al.* [6], similarity A^{mir}_{ij} between mir_i and mir_j is calculated using their association with the two groups of diseases. The DAG of a disease usually consists of the disease and its related terms and their associations [6]. Using the Wang *et al.* technique [6], the semantic similarity A^{dis}_{ij} between two diseases dis_i and dis_j is determined based on their DAGs. The higher the value of A^{mir}_{ij} $\left(A^{dis}_{ij}\right)$, the greater the similarity between mir_i (dis_i) and mir_j (dis_j).

$B^{mir-dis}$ reflects the association between miRNAs and diseases,

$$B^{mir-dis} \in R^{N_m \times N_d}, if\ v_i \in V^{mir}, v_j \in V^{dis}, \quad (2)$$

each row of $B^{mir-dis}$ stands in for one miRNA and each column for one disease. In case mir_i and dis_j are related, then $B^{mir-dis}_{ij} = 1$; otherwise, $B^{mir-dis}_{ij} = 0$.

Given the similarity matrix A^{mir}, A^{dis} and association matrix $B^{mir-dis}$, the weight matrix of all edges in G is

$$W = \begin{bmatrix} A^{mir} & B^{mir-dis} \\ \left(B^{mir-dis}\right)^T & A^{dis} \end{bmatrix} \in R^{N_v \times N_v}, \quad (3)$$

where $\left(B^{mir-dis}\right)^T$ is the transpose matrix for $B^{mir-dis}$ and $N_v = N_m + N_d$. All nodes in G are connected by edges to form the adjacency matrix C, $C = W$. For miRNA or disease nodes, the similarities and associations between them can be considered attributes of the nodes. Therefore, W is considered the attribute matrix of nodes in G and is denoted as X.

C. Learning of semantic information based on various meta-paths

Definition 2. Meta-path. A meta-path is a path with a form of $D_1 \xrightarrow{r_1} D_2 \xrightarrow{r_2} \cdots \xrightarrow{r_l} D_{l+1}$, where $D_1 ... D_{l+1}$ represent the node types and $r_1 ... r_{l+1}$ denote the relation types.

Definition 3. Meta-path instance. For a meta-path q, its instance is a sequence of nodes and the instance follows q's definition. For example, $mir \xrightarrow{r_1} mir \xrightarrow{r_2} dis$ is a meta-path, r_1 denotes the miRNA similarity, and r_2 denotes the miRNA-disease association. $mir_1 \rightarrow mir_3 \rightarrow dis_9$ is one of its instances.

Definition 4. Meta-path-based neighbor. In terms of a meta-path q, the meta-path-based neighbors N_i^D of a node i is a group of nodes that connect with i via meta-path instances of q.

Taking miRNA as an example, we established three meta-paths $miRNA \rightarrow miRNA$, $miRNA \rightarrow disease$ and $miRNA \rightarrow miRNA \rightarrow miRNA$. First, if the functions of two miRNAs are more similar, they are more likely to associate with similar diseases. The first meta-path $(miRNA \rightarrow miRNA)$ for the miRNA nodes was constructed based on this biological premise, and the connection semantic within the meta-path is the similarity between two miRNAs. Second, if two miRNAs have associated with more similar diseases, it is more possible that they would associate with similar diseases. Thus, the second meta-path $(miRNA \rightarrow disease)$ was designed and it contained the semantic reflecting the association between a miRNA and a disease. Third, the third meta-path was also designed based on the same biological premise with that for the first meta-path. Its goal is to obtain the neighbors of a target miRNA's neighbors, and these high-order neighbors have similar functions with the target miRNA. Thus, the third meta-path $(miRNA \rightarrow miRNA \rightarrow miRNA)$ includes two connections and their semantics are the miRNA similarity. Therefore, constructing the meta-paths with various semantic connections could reflect the biological characteristics about the miRNA-disease associations.

The miRNA (disease) node have multiple meta-paths, and each meta-path contains its specific semantic information. For instance, the meta-path $miRNA \rightarrow miRNA$ includes the similarity semantic information among a target miRNA node and its meta-path neighbors, while the meta-path $miRNA \rightarrow disease$ contains the association semantic information among the target miRNA node and its meta-path neighbors. Therefore, constructing a separate learning block for each meta-path is helpful for encoding its specific semantic. We built a meta-path-guided encoder to learn semantic information (Figure 2).

Because the semantic coding process of miRNAs is similar to that of diseases, we used miRNA as the target node to explain this process. The feature vector of mir_i is called $X_i \in X$. First, we projected X_i using the mapping matrix W_m^{MP} of miRNA node types to form the projected feature vector h_i,

$$h_i = \mu\left(W_m^{MP} \cdot X_i + b_m^{MP}\right), \tag{4}$$

where μ is the activation function $tanh$ and b_m^{MP} denotes a bias vector. The type-specific mapping matrix for disease nodes is W_d^{MP}.

Given a meta-path q_m, $N_i^{q_m}$ is the set of neighboring nodes of mir_i corresponding to the meta-path. The topology and node attributes in q_m are deeply fused to obtain the semantic

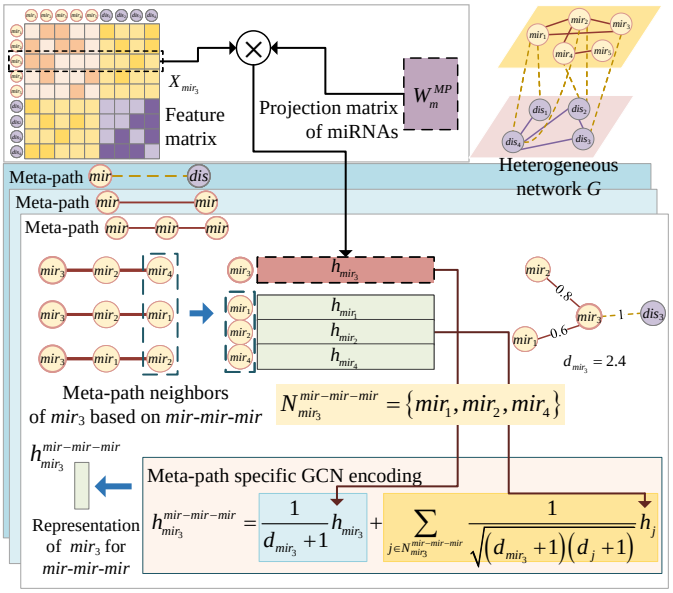

Fig. 2. Illustration of meta-path specific GCN encoding for a miRNA node.

feature $h_i^{q_n}$ of mir_i by meta-path specific GCN [37],

$$h_i^{q_m} = \frac{1}{d_i + 1} h_i + \sum_{j \in N_i^{q_m}} \frac{1}{\sqrt{(d_i + 1)(d_j + 1)}} h_j, \tag{5}$$

where h_i and h_j are the projected feature vectors of mir_i and its neighboring nodes, while d_i and d_j are degrees of the nodes.

The miRNA (disease) nodes have their specific multiple meta-paths, and these meta-paths contain various semantic connections. Given a meta-path set $Q_{mir} = \{q_1, q_2, ..., q_{N_p}\}$, N_p semantic feature vectors $\left\{h_i^{q_1}, ..., h_i^{q_{N_p}}\right\}$ of the target node mir_i are generated by meta-path specific GCN encoder. Because the semantic features implied by $h_i^{q_1}, ..., h_i^{q_{N_p}}$ contribute differently to the semantic representation learning of mir_i for the meta-path, we design a semantic-level attention mechanism. The attention weight of the semantics implied by q_m is α_{i,q_m},

$$\alpha_{i,q_m} = \frac{\exp\left(\rho\left((u_i^{q_m})^T h_i^{q_m}\right)\right)}{\sum_{s=1}^{N_p} \exp\left(\rho\left((u_i^{q_m})^T h_i^{q_s}\right)\right)}, \tag{6}$$

where $u_i^{q_m}$ denotes the trainable attention vector, exp is the exponential function, and ρ is the activation function $LeakyReLU$.

The meta-path semantics representation of mir_i enhanced by attention, $Z^{MP}(i)$, is given as follows.

$$Z^{MP}(i) = \sum_{s=1}^{N_p} (\alpha_{i,q_s} \cdot h_i^{q_s} + h_i^{q_s}), \tag{7}$$

The meta-path semantic representations form matrix Z^{MP}. The meta-path semantic representation of disease node dis_j is $Z^{MP}(N_m + j)$.

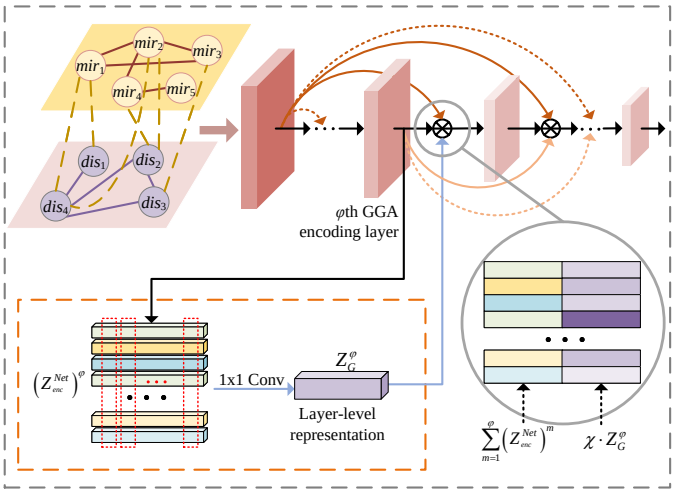

Fig. 3. Illustration of global-aware graph convolutional autoencoder with layer-level representation encoding.

D. Learning of miRNA-disease heterogeneous network information

The heterogeneous network G is composed of the miRNA and disease nodes, and the similarity and association connections have their context relationships. A global context-aware graph convolutional autoencoder (GGA) was designed to encode the relationships and form the network-view representations of nodes (Figure 3).

GGA encoder. C_{ii} denotes the similarity between node v_i and itself, so the values on the diagonal of C are all 1, indicating the connection from each node to itself. C is normalized by Laplace to $\tilde{C}$,

$$\tilde{C} = E^{-\frac{1}{2}} C E^{-\frac{1}{2}}, \tag{8}$$

where E is C's degree matrix, and $E_{ii} = \sum_j C_{ij}$.

We feed $\tilde{C}$ and X into the GGA encoder. Let the output of the φth GGA encoding layer be $\left(Z_{enc}^{Net}\right)^{\varphi}$, where $\varphi = 1, 2, 3, ..., \lambda_{enc}$ and λ_{enc} is the total number of encoding layers. $\left(Z_{enc}^{Net}\right)^{\varphi}$ is fused by 1x1 convolution to obtain the layer-level representation Z_G^{φ} of the φth GGA encoding layer,

$$Z_G^{\varphi} = w_G^{\varphi} \cdot \left(Z_{enc}^{Net}\right)^{\varphi} + b_G^{\varphi}, \tag{9}$$

where w_G^{φ} is the filter of 1x1 convolution and b_G^{φ} is the bias vector. Z_G^{φ} contains the global information of all the miRNA and disease nodes in the φth layer.

The outputs of the 1st,2nd,...,$(\varphi-1)$th GGA encoding layer, $\left(Z_{enc}^{Net}\right)^{1}$, $\left(Z_{enc}^{Net}\right)^{2}$,...,$\left(Z_{enc}^{Net}\right)^{\varphi-1}$, are supplement to the φth layer to form $\sum_{m=1}^{\varphi}\left(Z_{enc}^{Net}\right)^{m}$. $\sum_{m=1}^{\varphi}\left(Z_{enc}^{Net}\right)^{m}$ and the layer-level representation Z_G^{φ} are concatenated to obtain $\sum_{m=1}^{\varphi}\left(Z_{enc}^{Net}\right)^{m} || \left(\chi \cdot Z_G^{\varphi}\right)$, where χ is a learnable vector, and $||$ denotes the concatenation operation. We feed $\sum_{m=1}^{\varphi}\left(Z_{enc}^{Net}\right)^{m} || \left(\chi \cdot Z_G^{\varphi}\right)$ into the $(\varphi+1)$th GGA encoding layer to get the network representation matrix $\left(Z_{enc}^{Net}\right)^{\varphi+1}$ of the $(\varphi+1)$th layer,

$$\left(Z_{enc}^{Net}\right)^{\varphi+1} = \theta\left(\tilde{C}\left(\sum_{m=1}^{\varphi}\left(Z_{enc}^{Net}\right)^{m} || \left(\chi \cdot Z_G^{\varphi}\right)\right) W_{enc}^{\varphi+1}\right), \tag{10}$$

where θ is an activation function $Relu$, $W_{enc}^{\varphi+1}$ is a weight matrix. The network-view representation matrix $\left(Z_{enc}^{Net}\right)^{\lambda_{enc}}$ of each node is obtained by the final layer of GGA encoder and renamed Z^{Net}.

GGA decoder. To obtain a better network-view representation, we employed a GGA decoder to gradually raise Z^{Net}'s dimensionality to its initial level. The output of layer φ of GGA decoder is

$$\left(Z_{dec}^{Net}\right)^{\varphi} = g\left(\tilde{C}\left(Z_{dec}^{Net}\right)^{\varphi-1} W_{dec}^{\varphi}\right), \tag{11}$$

where $\varphi = 1, 2, 3, ..., \lambda_{dec}$. g is $Relu$, an activation function. In addition, the weight matrix of the φth decoding layer is W_{dec}^{φ}, and the total amount of decoding layers is λ_{dec}. The output $\left(Z_{dec}^{Net}\right)^{\lambda_{dec}}$ of the last layer of GGA decoder is obtained and renamed $\hat{X}$.

Optimization. Our optimization goal is to make the decoding result $\hat{X}$ and the original message X as consistent as possible. The Adam function optimizes the loss $Loss_{Net}$ between the two using the following definition:

$$Loss_{Net} = \left\| X - \hat{X} \right\|_F^2, \tag{12}$$

where F is the Frobenius norm.

E. Integrating semantic representation and network-view representation

For a miRNA node mir_i, the meta-path semantic representation and the network-view representation are $Z^{MP}(i)$ and $Z^{Net}(i)$, respectively. Due to the different contributions of $Z^{MP}(i)$ and $Z^{Net}(i)$ to the miRNA-disease association prediction, the representation-level attention was proposed to adaptively fuse these representations. The following formulas are used to determine the attention score $w_{i,v}$ and the accompanying normalized attention weights $\beta_{i,v}$.

$$w_{i,v} = a_m \sigma\left(W_m^{nft} \cdot \left(Z^v(i)\right) + b_m^{nft}\right), \tag{13}$$

$$\beta_{i,v} = \frac{\exp\left(w_{i,v}\right)}{\sum_{t\in\{MP,Net\}} \exp\left(w_{i,t}\right)}, \tag{14}$$

where $v \in \{MP, Net\}$, a_m and W_m^{nft} are the weight vector and weight matrix, respectively, b_m^{nft} represents bias vector,

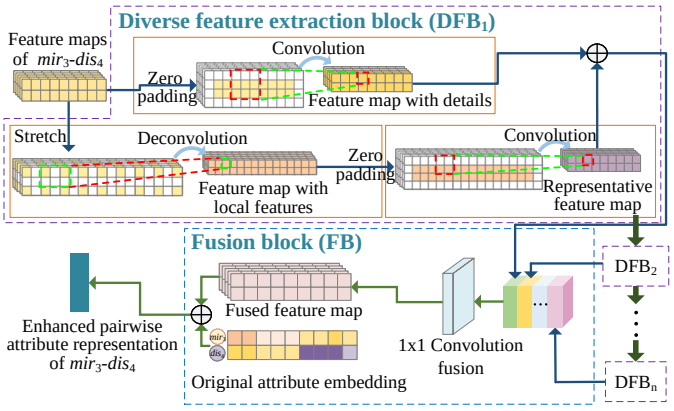

Fig. 4. Learning the local feature enhanced attribute representation for a pair of miRNA and disease nodes.

σ is the activation function $tanh$, and exp is an exponential function.

The node feature fusion representation of mir_i is denoted as $T^{nft}(i)$,

$$T^{nft}(i) = \sum\nolimits_{v\in\{MP,Net\}} \left(w_{i,v} \cdot \left(Z^v(i)\right) + Z^v(i)\right), \quad (15)$$

$T^{nft}(N_m + j)$ is the node feature fusion representation of disease dis_j.

$T^{nft}(i)$ and $T^{nft}(N_m + j)$ are stitched together left and right to form K. Then we obtained the $mir_i - dis_j$ association score s^{nft} by feeding K to the fully connected layer and the $softmax$ layer,

$$s^{nft} = softmax\left(W_{nft} \cdot K + b_{nft}\right), \quad (16)$$

where W_{nft} and b_{nft} denote weight matrix and bias vector of the fully connected layer.

F. Learning of detail and diversity feature enhanced node pair attribute representations

The feature vectors X_i and X_{N_m+j} of miRNA mir_i and disease dis_j were spliced up and down to obtain the paired attribute embedding $H \in R^{2\times N_v}$ of $mir_i - dis_j$. A parallel convolutional-deconvolutional network was built to extract diverse pairwise node attributes. As shown in Figure 1(d), the parallel convolution-deconvolution module includes a shallow feature extraction block (SFB), multiple diversity feature extraction blocks (DFB), and a fusion block (FB).

The SFB consists of a convolutional layer, H is fed into SFB to form the feature map $T^s_{deh} \in R^{2\times N_v}$,

$$T^s_{deh} = \delta\left(W^s_{deh} * H + b^s_{deh}\right), \quad (17)$$

where δ is the activation function $Relu$, weight matrix and bias vector are represented by W^s_{deh} and b^s_{deh}, and ' $*$ ' denotes the convolution operation.

We connected the N_{deh} DFBs in series, and the input of each DFB was the output of the previous feature extraction block. As described in Figure 4, each DFB contains two parallel-connected blocks: a convolution block and deconvolution block. The attribute features of the node pair at the original level are extracted using a convolutional block, which consists of one convolutional layer, and T^s_{deh} is the input to the first convolutional block. To learn the edge information of T^s_{deh}, it was filled with 0.

T^s_{deh} passes through the convolutional layer in the ψ th convolutional block to form the convolutional feature map $T^{c,\psi}_{deh}$,

$$T^{c,\psi}_{deh} = W^{c,\psi}_{deh} * T^{c,\psi-1}_{deh} + b^{c,\psi}_{deh}, 1 \le \psi \le N_{deh}, \quad (18)$$

where ' $*$ ' represents the convolution operation. $W^{c,\psi}_{deh}$ represents the weight matrix.

One deconvolution layer and a convolution layer constitute the deconvolution block. The deconvolution layer performs an inter-element complementary 0 to the original feature map, stretching its size and making the perceptual field more focused on the local information of the original input. Therefore, this deconvolution layer can learn representative features at the local level in the feature map.

The input of the first deconvolution block is T^s_{deh}, which passes through the deconvolution layer of the ψ th deconvolution block to obtain the feature map $T^{dc,\psi}_{deh}$,

$$T^{dc,\psi}_{deh} = W^{dc,\psi}_{deh} \circledast T^{dc,\psi-1}_{deh} + b^{dc,\psi}_{deh}, 1 \le \psi \le N_{deh}, \quad (19)$$

where ' $\circledast$ ' denotes the deconvolution operation. $W^{dc,\psi}_{deh}$ and $b^{dc,\psi}_{deh}$ are the weight matrix and the bias vector, respectively, and $T^{dc,0}_{deh} = T^s_{deh}$.

$T^{dc,\psi}_{deh}$ is then passed through a convolution layer so that it is reprojected to the same dimension as $T^{c,\psi}_{deh}$ to form a deconvolution feature map $T^{d,\psi}_{deh}$,

$$T^{d,\psi}_{deh} = W^{d,\psi}_{deh} * T^{dc,\psi}_{deh} + b^{d,\psi}_{deh}, 1 \le \psi \le N_{deh}, \quad (20)$$

where ' $*$ ' represents the convolution operation, $W^{d,\psi}_{deh}$ represents the weight matrix, and $b^{d,\psi}_{deh}$ is the bias vector. Thus, after the convolution blocks and deconvolution blocks, we learn the diversity features of the node pairs from different levels.

In each DFB, we sum the obtained convolution and deconvolution feature maps before feeding the activation function, and then obtain the ψ th diverse feature map $T^{M,\psi}_{deh}$,

$$T^{M,\psi}_{deh} = g\left(T^{c,\psi}_{deh} + T^{d,\psi}_{deh}\right), \quad (21)$$

where g is the activation function $Relu$. To integrate detailed and diverse node pair features, we established cross-layer connections between individual feature extraction blocks.

We fused $T^{M,1}_{deh}, T^{M,2}_{deh}, ..., T^{M,N_{deh}}_{deh}$ at FB using a 1×1 convolution kernel to obtain the fused diversity feature map $\hat{T}^R_{deh}$,

$$\hat{T}^R_{deh} = W^R_{deh} \cdot \left|T^{M,1}_{deh}, T^{M,2}_{deh}, ..., T^{M,N_{deh}}_{deh}\right| + b^R_{deh}, \quad (22)$$

where $|\cdot|$ is the connection operation, W^R_{deh} and b^R_{deh} are weight matrix and bias vector, respectively. Finally, the original feature map H was supplemented to retain more original detail information, resulting in detailed and diverse feature-enhanced pairwise attribute representation T^R_{deh},

$$T^R_{deh} = \hat{T}^R_{deh} + H \quad (23)$$

We fed T^R_{deh} into the fully connected layer and $softmax$ layer to calculate the association prediction score s^{deh},

$$s^{deh} = softmax\left(W_{deh} \cdot T^R_{deh} + b_{deh}\right), \quad (24)$$

where W_{deh} and b_{deh} represent the weight matrix and the bias vector, respectively.

G. Association prediction score and optimization

When learning the node feature fusion representation, cross-entropy is applied as a loss function between the association prediction score s^{nft} and the real label t, defined as

$$Loss_{nft} = -\sum_{i=1}^{N_{tra}} \sum_{j=1}^{c} t_j \log\left(s^{nft}, j\right), \quad (25)$$

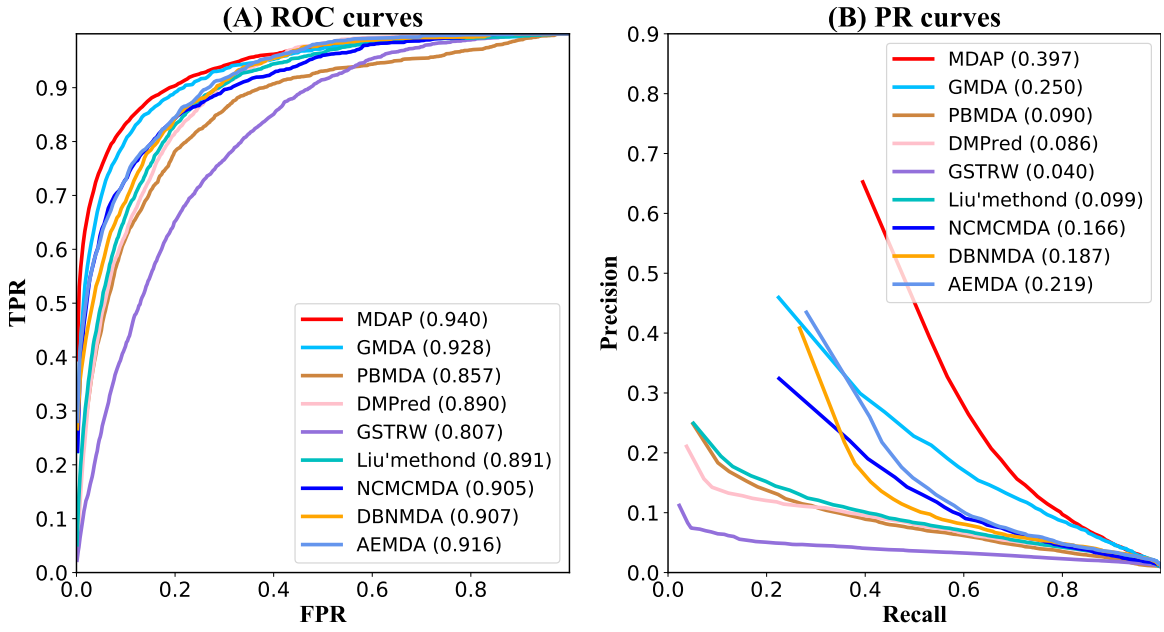

Fig. 5. MDAP and other methods for the area under ROC and PR curves of all diseases.

where N_{tra} denotes the number of training sample sets. When there is a real association between mir_i and dis_j, $t_j = 1$; otherwise, $t_j = 0$

In the learning process of node pair attribute representation with enhanced detail and diversity features, the cross-entropy loss is $Loss_{deh}$,

$$Loss_{deh} = -\sum_{i=1}^{N_{tra}} \sum_{j=1}^{c} t_j \log\left(s^{deh}, j\right), \tag{26}$$

We use Adam algorithm [38] to optimize the loss functions $Loss_{nft}$ and $Loss_{deh}$, respectively.

Finally, s^{nft} and s^{deh} are weighted and integrated to obtain the final predicted score S (Figure 1(e)),

$$S = \lambda \times s^{nft} + (1 - \lambda) \times s^{deh}, \tag{27}$$

where λ is utilized to modify the contributions of the enhanced node pair attribute information and the node feature fusion representation.

III. Experimental Evaluations and Discussions

A. Experimental setup

In this model, we learn the semantic information of meta-path $miRNA - miRNA$, $miRNA - disease$, $miRNA - miRNA - miRNA$, $disease - disease$, $disease - miRNA$, and $disease - disease - disease$. The dimension of the feature mapping was 600 and the number of neighbor nodes was 50 in the meta-path-encoding module. The network-view representation learning module had 3 encoding layers and 3 decoding layers, and output feature dimensions of 600 for all of the encoding layers and 700, 900 and 1134 for the decoding layers. In the convolution-deconvolution module, SFB has a convolutional layer with 3 × 3 filters and the model has two DFB blocks. The deconvolution block in a DFB contains a deconvolution layer and a convolution layer. The deconvolution layer stretches the feature map by complementing a 0 among the original features. It has 2 × 3 filters and the number of filters is 32, and the sliding stride is 1. The convolution layer contains 32 filters with 2 × 2 size and the sliding stride is 2. The convolution block in a DFB has 32 filters with 3 × 3 size and the sliding stride is 1, and the zero-padding is 1. The learning rate was set as 0.0005.

B. Evaluation metrics

The performance of the MDAP and other models used to infer miRNA-associated diseases was assessed using five-fold cross validation. All miRNA-disease associations we know were employed as positive samples, of which 4/5 were used for training and the remaining for testing. Negative samples consisted of all unobserved miRNA-disease associations. Randomly chosen negative samples were added to the training set in an amount equal to the number of positive samples used for training, and the remaining negative samples were used for testing. We used the area under the receiver operating characteristic (ROC) curve (AUC) [31], [39] and the area under the precision-recall (PR) curve (AUPR) as evaluation metrics [40]. Prior to obtaining the average of the five outcomes, we first determined the average AUC and AUPR for each fold. Also, biologists typically choose the candidates with the highest rankings for validation. Therefore, we calculate the recall rates for the top-ranked candidates with the highest prediction results and used it as one of the evaluation metrics.

C. Ablation experiments

To confirm the contribution of meta-path semantic learning, network-view feature learning, node pair attribute learning, representation-level attention mechanism, and semantic-level attention mechanism to predict miRNA-disease associations, we ran ablation experiments. As shown in Table I, the final model MDAP achieved the best performance. Without using the meta-path encoding module to learn meta-path semantic

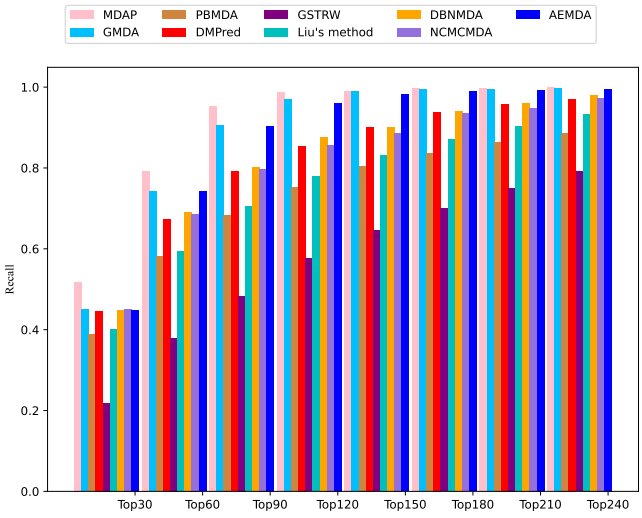

Fig. 6. The recall values at different top k cutoffs.

information, the AUC decreased by 1.2%, and the AUPR was 8.6% lower than that of the final model. It showed that integrating the various semantics of multiple meta-paths and the attributes of their neighbors is necessary for the improved prediction performance. Without using the GGA to learn heterogeneous network information, the AUC and AUPR were 1.1% and 6.3% lower than those of the final model. It demonstrated the contribution of features from the heterogeneous network to obtaining the better prediction results. Compared with the model without pairwise attribute learning by the convolutional-deconvolutional neural networks, the complete model achieved 1.7% of higher AUC and 10.2% of higher AUPR. We replaced the deconvolution with the dilated convolution in our prediction model, and the corresponding prediction performance is shown in the 6th row of Table I. Compared with the dilated convolution strategy, the AUC and AUPR of the deconvolution strategy increased by 1.3% and 11.5%. It indicated learning the more detailed features of the feature map is necessary. Compared to MDAP, the model without the representation-level attention mechanism showed a decrease of 0.5% and 4.1% in AUC and AUPR, respectively, while the model without the semantic-level attention mechanism displayed a 0.3% and 1.6% decrease in AUC and AUPR, respectively. The main reason is that representation-level attention and semantic-level attention assign higher weights to more informative node representations and semantic features, which helps to capture the more important features of nodes and improve the prediction performance. The pairwise attribute learning has the most contribution for the prediction performance improvement. On the one hand, the pairwise attribute embedding contains the direct similarity and association information among the pair of miRNA-disease nodes and all the miRNA and disease nodes. On the other hand, the convolutional-deconvolutional neural network is able to learn the more detailed features of the node pair.

D. Comparison with other methods

MDAP was compared with GMDA [31], AEMDA [32], DBNMDA [34], NCMCMDA [16], GSTRW [21], DMPred [13], PBMDA [11], and Liu's method [19]. By utilizing the best parameter values available for each method, the best performance of MDAP and the other eight methods was presented. Figure 5 displays the ROC and PR curves for the eight approaches for 341 diseases. For the average AUC across all diseases, MDAP yielded the highest AUC of 0.940, which outperformed GMDA by 1.2%, was 4.9% better than Liu's method, was superior to PBMDA by 8.3%, was 5% and 3.3% higher than the values of DMPred and DBNMDA, exceeded NCMCMDA by 3.5%, AEMDA by 2.4%, and outperformed the worst performing GSTRW by 13.3 %. Similarly, MDAP obtained the best AUPR value (AUPR = 0.397), which was 14.7%, 30.7%, 21%, 29.8%, 35.7%, 23.1%, 17.8%, and 31.1% higher than GMDA, PBMDA, DBNMDA, Liu's method, GSTRW, NCMCMDA, AEMDA and DMPred, respectively.

In summary, the proposed MDAP model achieves the best performance, with GMDA and AEMDA in the second and third places, respectively. GMDA is based on generative adversarial networks that integrate the neighboring topology of node pairs. AEMDA uses a deep auto-coding model to learn the potential associations. Both ignore the rich semantic information in meta-paths. DBNMDA is based on a deep belief network that deeply integrates the complex relationships between miRNAs and diseases, but does not consider the details of node-pair attributes and edge information. NCMCMDA is a matrix completion-based model that fully utilizes similarity information. The AUC and AUPR values of Liu's method are similar to those of DMPred, both of which were modeled based on the random walk and non-negative matrix factorization, respectively. The AUPR of PBMDA was a trifle higher than that of DMPred, yet its AUC was 3.3% lower than that of DMPred. GSTRW performed much worse than MDAP, mainly because GSTRW integrated the global similarities among all the miRNAs (diseases) and the information which was obtained by random walking on the miRNA network and the disease network. However, it is a shallow prediction model, and it is difficult for the model to learn the deep and complex features of the miRNA and disease nodes. Most of the compared methods were proposed based on the graph convolutional network, the graph neural network, and the convolutional neural network. They all belong to the deep learning methods and they could mine the deep features of the miRNA and disease nodes. The experimental results demonstrated these methods achieved better prediction performance than GSTRW, which indicates learning the deep features of nodes is necessary. The improvement of MDAP over other methods is mainly because of its in-depth learning of more local and representative features of node-pair attributes and the integration of node features at the network and meta-path levels.

We obtained mean five-fold AUC and AUPR results for 341 diseases for each prediction approach. A paired Wilcoxon test was performed using 341 pairwise results when the two methods were compared. The study outcomes are presented in Table II. MDAP showed better performance in both AUC and AUPR, with a p-value of 0.05.The biologists usually select the top-ranked candidates from the prediction results for further wet-lab experimental validation. Therefore, whether there are

TABLE I
RESULTS OF ABLATION EXPERIMENTS IN OUR METHOD.

meta-path semantic learning	network-view feature learning	node pair attribute learning		representation-level attention	semantic-level attention	Average AUC	Average AUPR
		deconvolution	dilated convolution				
×	✓	✓	×	×	×	0.928	0.311
✓	×	✓	×	×	✓	0.929	0.334
✓	✓	×	×	✓	✓	0.923	0.295
✓	✓	✓	×	×	✓	0.935	0.356
✓	✓	✓	×	✓	×	0.937	0.381
✓	✓	×	✓	✓	✓	0.927	0.282
✓	✓	✓	✓	✓	✓	**0.940**	**0.397**

TABLE II
RESULTS OF THE PAIRED WILCOXON TEST FOR COMPARING MDAP AND ALL THE OTHER METHODS.

	GMDA	AEMDA	Liu's Method	PBMDA	DMPred	DBNMDA	GSTRW	NCMCMDA
p-value of AUC	4.6863e-13	8.6258e-12	8.2599e-17	6.5448e-23	5.3865e-26	7.5159e-22	2.0356e-18	1.8501e-29
p-value of AUPR	2.1259e-06	4.1831e-07	1.9136e-12	3.6607e-20	9.4412e-15	5.3148e-13	3.1410e-09	6.0247e-24

the actual miRNA-disease associations among the top-ranked candidates is crucial for the biologists when they select the candidates. The recall rates for the top-ranked candidates may reflect the proportion of the correctly determined positive samples out the total positive samples. Figure 6 displays the recall rates for various top k values. More positive samples were obtained when the recall rate was higher. Our model outperformed other methods for different values of k because of the comprehensive learning of the semantic information of nodes for multiple meta-paths, miRNA-disease network information, and property details of node pairs.

MDAP obtained the highest recall rates when k was 30, 60, 90, and 120, with 52.3%, 79.6%, 96.1%, and 98.8%, respectively. The GMDA and AEMDA also achieved acceptable recall rates. The former had recall rates of 45%, 74.2%, 90.6%, and 97%, while the latter had 44.9%, 74.25, 90.4%, and 96.1%. When k in the range of 30–120, the values of NCMCMDA were 45.2%, 68.7%, 79.8%, and 85.7%, respectively. The recall rates for DBNMDA and DMPred were very close, with 44.8%, 69.1%, 80.2%, and 87.6% for the former and 44.7%, 67.4%, 79.1%, and 85.4% for the latter corresponding to the rates. When k=30, Liu's method ranked 40.2%, and PBMDA ranked 38.8%. When k ranged from 60 to 120, Liu's method achieved recall rates of 59.5%, 70.5%, and 77.9%, respectively, and PBMDA obtained rates of 58.1%, 68.3%, and 75.2%. The rates of DSTRW were consistently lower than those of the other methods when the k values ranged from 30 to 120, with rates of 21.8%, 37.9%, 48.4%, and 57.6%.

E. Case studies on 3 diseases

To further demonstrate the capability of MDAP to identify potential miRNA-disease associations, we conducted case studies on three diseases, including ovarian neoplasms, lung neoplasms, and breast neoplasms. Using ovarian neoplasms as an example, we obtained the association score of its candidate miRNAs through the model and ranked it in decreasing order. For validation and analysis, the top 50 putative miRNAs for ovarian neoplasms were identified. Supplementary Tables ST1 and ST2 list the top 50 potential miRNAs associated with breast and lung neoplasms, respectively.

The miRCancer database contains 9080 experimentally validated miRNA-disease associations, which were extracted from published literature using text mining techniques by Xie *et al.* [41] and further verified manually. miR2Disease is a database covering 349 miRNAs, 163 diseases, and 3273 miRNA-disease-related information. As indicated in Table III, 37 candidate miRNAs for ovarian neoplasms were identified in miRCancer, and 9 candidate miRNAs were included in miR2Disease. This suggests that the expression of these miRNAs is upregulated or downregulated in ovarian neoplasms.

The integrated database dbDEMC stores differentially expressed miRNAs in cancer, as detected using high-throughput and low-throughput methods [42]. The database includes 2224 miRNAs and 36 cancers. In our model, 46 of the 50 miRNA candidates associated with ovarian neoplasms were supported by dbDEMC, implying that they were aberrantly expressed in ovarian neoplasm cells. Among the 50 candidates, 1 candidate miRNA was not confirmed by the observed evidence; hence, it was labeled as "unconfirmed". Supplementary Table ST1 shows 50 candidate miRNAs for breast neoplasms. miRCancer and miR2Disease contain 33 and 10 candidate miRNAs, respectively, suggesting that these miRNAs may be involved in the development of breast neoplasms. Among the 50 candidate miRNAs for lung neoplasms (Supplementary Table ST2), miRCancer included 43 candidate miRNAs. There were 49 candidates recorded by dbDEMC and 16 candidates by miR2Disease, indicating that these miRNAs were indeed associated with lung neoplasms. All case studies show that MDAP is indeed able to discover potential miRNA-disease associations.

F. Investigation of novel disease-related miRNAs prediction

We utilized MDAP to infer disease-associated miRNA candidates for 341 diseases. In Supplementary Table ST3, the top 50 miRNA candidates for every disease predicted by MDAP are listed.

TABLE III
THE TOP 50 MIRNA CANDIDATES RELATED TO OVARIAN NEOPLASMS.

Rank	MiRNA name	Description	Rank	MiRNA name	Description
1	hsa-mir-155	miRCancer, dbDEMC, miR2Disease	26	hsa-mir-34b	miRCancer, dbDEMC, miR2Disease
2	hsa-mir-145	miRCancer, dbDEMC	27	hsa-mir-139	miRCancer, dbDEMC
3	hsa-mir-92a	miRCancer, dbDEMC	28	hsa-mir-181c	miRCancer, dbDEMC
4	hsa-mir-223	miRCancer, dbDEMC, miR2Disease	29	hsa-mir-326	dbDEMC
5	hsa-mir-222	miRCancer, dbDEMC	30	hsa-mir-141	miRCancer, dbDEMC, miR2Disease
6	hsa-mir-29a	miRCancer, dbDEMC, miR2Disease	31	hsa-mir-10a	miRCancer, dbDEMC
7	hsa-mir-15a	miRCancer, dbDEMC	32	hsa-mir-135a	miRCancer, dbDEMC
8	hsa-mir-146b	miRCancer, dbDEMC, miR2Disease	33	hsa-mir-103	dbDEMC
9	hsa-mir-1	miRCancer, dbDEMC	34	hsa-mir-186	dbDEMC
10	hsa-mir-122	dbDEMC	35	hsa-mir-184	dbDEMC
11	hsa-mir-27b	miRCancer, dbDEMC	36	hsa-mir-328	dbDEMC
12	hsa-mir-18	miRCancer, miR2Disease	37	hsa-mir-33b	miRCancer, dbDEMC
13	hsa-mir-133a	miRCancer, dbDEMC	38	hsa-mir-202	miRCancer, dbDEMC
14	hsa-mir-149	miRCancer, dbDEMC	39	hsa-mir-18b	miRCancer
15	hsa-mir-342	dbDEMC	40	hsa-mir-320	miRCancer
16	hsa-mir-26b	miRCancer, dbDEMC	41	hsa-mir-99b	miRCancer, dbDEMC
17	hsa-mir-30a	miRCancer, dbDEMC, miR2Disease	42	hsa-mir-451a	dbDEMC
18	hsa-mir-486	miRCancer, dbDEMC	43	hsa-mir-301a	dbDEMC
19	hsa-mir-92	unconfirmed	44	hsa-mir-425	dbDEMC
20	hsa-mir-98	miRCancer, dbDEMC	45	hsa-mir-590	miRCancer, dbDEMC
21	hsa-mir-103a	dbDEMC	46	hsa-mir-198	miRCancer, dbDEMC
22	hsa-mir-200b	miRCancer, dbDEMC, miR2Disease	47	hsa-mir-33a	miRCancer, dbDEMC
23	hsa-mir-192	miRCancer, dbDEMC	48	hsa-mir-377	dbDEMC
24	hsa-mir-194	miRCancer, dbDEMC	49	hsa-mir-32	miRCancer, dbDEMC
25	hsa-mir-204	miRCancer, dbDEMC	50	hsa-mir-340	miRCancer, dbDEMC

IV. CONCLUSIONS

We proposed a miRNA-disease association prediction method, MDAP, for encoding and fusing the semantics from diverse meta-paths and the topological structure and attributes from the entire miRNA-disease network. The constructed multiple meta-paths for the miRNA (disease) nodes are helpful for capturing the high-order neighbor topologies about these meta-paths to learn the semantic features of the miRNA (disease) nodes. A meta-path-sensitive graph convolution module was built for each meta-path and we integrated the various semantic features of the miRNA and disease nodes from multiple meta-paths. A graph convolutional autoencoder was constructed to capture the topological structure and the node attributes from the entire miRNA-disease heterogeneous network. The designed strategy based on parallel convolutional-deconvolutional networks is beneficial for the more local and representative features of each miRNA-disease pair. Two attention mechanisms were proposed to assign greater weights for the more informative meta-paths and node representations. Five-fold cross-validation results showed that MDAP achieved higher AUC and AUPR than the compared methods. MDAP also obtained the more actual disease-related miRNAs in the top-ranked prediction candidates. MDAP's ability in discovering the potential miRNA-disease associations was illustrated through the case studies on 3 diseases which includes ovarian neoplasms, lung neoplasms, and breast neoplasms.

REFERENCES

[1] L. F. R. Gebert and I. J. MacRae, "Regulation of microrna function in animals," *NATURE REVIEWS MOLECULAR CELL BIOLOGY*, vol. 20, no. 1, pp. 21–37, 2019.

[2] L. Chen, L. Heikkinen, C. Wang, Y. Yang, H. Sun, and G. Wong, "Trends in the development of mirna bioinformatics tools," *BRIEFINGS IN BIOINFORMATICS*, vol. 20, no. 5, pp. 1836–1852, 2019.

[3] X. Chen, D. Xie, Q. Zhao, and Z. You, "Micrornas and complex diseases: from experimental results to computational models," *BRIEFINGS IN BIOINFORMATICS*, vol. 20, no. 2, pp. 515–539, 2019.

[4] H. Matsuyama and H. I. Suzuki, "Systems and synthetic microrna biology: From biogenesis to disease pathogenesis," *INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES*, vol. 21, no. 1, 2020.

[5] Y. Zhao, X. Chen, and J. Yin, "Adaptive boosting-based computational model for predicting potential mirna-disease associations (vol 47, pg 353, 2019)," *BIOINFORMATICS*, vol. 36, no. 1, p. 330, 2020.

[6] D. Wang, J. Wang, M. Lu, F. Song, and Q. Cui, "Inferring the human microrna functional similarity and functional network based on microrna-associated diseases," *BIOINFORMATICS*, vol. 26, no. 13, pp. 1644–1650, 2010.

[7] P. Xuan, K. Han, M. Guo, Y. Guo, J. Li, J. Ding, Y. Liu, Q. Dai, J. Li, Z. Teng, and Y. Huang, "Prediction of micrornas associated with human diseases based on weighted k most similar neighbors," *PLOS ONE*, vol. 8, no. 8, p. e70204, 2013.

[8] X. Chen, M. Liu, and G. Yan, "Rwrmda: predicting novel human microrna-disease associations," *MOLECULAR BIOSYSTEMS*, vol. 8, no. 10, pp. 2792–2798, 2012.

[9] P. Xuan, K. Han, Y. Guo, J. Li, X. Li, Y. Zhong, Z. Zhang, and J. Ding, "Prediction of potential disease-associated micrornas based on random walk," *BIOINFORMATICS*, vol. 31, no. 11, pp. 1805–1815, 2015.

[10] X. Chen, Q. Wu, and G. Yan, "Rknnmda: Ranking-based knn for mirna-disease association prediction," *RNA BIOLOGY*, vol. 14, no. 7, pp. 952–962, 2017.

[11] Z. You, Z. Huang, Z. Zhu, G. Yan, Z. Li, Z. Wen, and X. Chen, "Pbmda: A novel and effective path-based computational model for mirna-disease association prediction," *PLOS COMPUTATIONAL BIOLOGY*, vol. 13, no. 3, p. e1005455, 2017.

[12] L. Wang, Z. You, X. Chen, Y. Li, Y. Dong, L. Li, and K. Zheng, "Lmtrda: Using logistic model tree to predict mirna-disease associations by fusing multi-source information of sequences and similarities," *PLOS COMPUTATIONAL BIOLOGY*, vol. 15, no. 3, p. e1006865, 2019.

[13] Y. Zhong, P. Xuan, X. Wang, T. Zhang, J. Li, Y. Liu, and W. Zhang, "A non-negative matrix factorization based method for predicting disease-associated mirnas in mirna-disease bilayer network," *BIOINFORMATICS*, vol. 34, no. 2, pp. 267–277, 2018.

[14] Y. Zhao, X. Chen, and J. Yin, "A novel computational method for the identification of potential mirna-disease association based on symmetric non-negative matrix factorization and kronecker regularized least square," *FRONTIERS IN GENETICS*, vol. 9, p. 324, 2018.

[15] P. Xuan, T. Shen, X. Wang, T. Zhang, and W. Zhang, "Inferring disease-associated micrornas in heterogeneous networks with node attributes,"

IEEE-ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, vol. 17, no. 3, pp. 1019–1031, 2020.

[16] X. Chen, L. Sun, and Y. Zhao, "Ncmcmda: mirna-disease association prediction through neighborhood constraint matrix completion," *BRIEFINGS IN BIOINFORMATICS*, vol. 22, no. 1, SI, pp. 485–496, 2021.

[17] X. Chen, J. Yin, J. Qu, and L. Huang, "Mdhgi: Matrix decomposition and heterogeneous graph inference for mirna-disease association prediction," *PLOS COMPUTATIONAL BIOLOGY*, vol. 14, no. 8, p. e1006418, 2018.

[18] A. Li, Y. Deng, Y. Tan, and M. Chen, "A novel mirna-disease association prediction model using dual random walk with restart and space projection federated method," *PLOS ONE*, vol. 16, no. 6, 2021.

[19] Y. Liu, X. Zeng, Z. He, and Q. Zou, "Inferring microrna-disease associations by random walk on a heterogeneous network with multiple data sources," *IEEE-ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS*, vol. 14, no. 4, pp. 905–915, 2017.

[20] J. Luo and Q. Xiao, "A novel approach for predicting microrna-disease associations by unbalanced bi-random walk on heterogeneous network," *JOURNAL OF BIOMEDICAL INFORMATICS*, vol. 66, pp. 194–203, 2017.

[21] M. Chen, B. Liao, and Z. Li, "Global similarity method based on a two-tier random walk for the prediction of microrna-disease association," *SCIENTIFIC REPORTS*, vol. 8, p. 6481, 2018.

[22] X. Chen and G. Yan, "Semi-supervised learning for potential human microrna-disease associations inference," *SCIENTIFIC REPORTS*, vol. 4, no. 1, p. 5501, 2014.

[23] C. Gu, B. Liao, X. Li, and K. Li, "Network consistency projection for human mirna-disease associations inference," *SCIENTIFIC REPORTS*, vol. 6, p. 36054, 2016.

[24] X. Chen, D. Xie, L. Wang, Q. Zhao, Z. You, and H. Liu, "Bnpmda: Bipartite network projection for mirna-disease association prediction," *BIOINFORMATICS*, vol. 34, no. 18, pp. 3178–3186, 2018.

[25] X. Zhang, J. Yin, and X. Zhang, "A semi-supervised learning algorithm for predicting four types mirna-disease associations by mutual information in a heterogeneous network," *GENES*, vol. 9, no. 3, p. 139, 2018.

[26] J. Peng, W. Hui, Q. Li, B. Chen, J. Hao, Q. Jiang, X. Shang, and Z. Wei, "A learning-based framework for mirna-disease association identification using neural networks," *BIOINFORMATICS*, vol. 35, no. 21, pp. 4364–4371, 2019.

[27] P. Xuan, Y. Dong, Y. Guo, T. Zhang, and Y. Liu, "Dual convolutional neural network based method for predicting disease-related mirnas," *INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES*, vol. 19, no. 12, p. 3732, 2018.

[28] J. Li, Z. Li, R. Nie, Z. You, and W. Bao, "Fcgcnmda: predicting mirna-disease associations by applying fully connected graph convolutional networks," *MOLECULAR GENETICS AND GENOMICS*, vol. 295, no. 5, pp. 1197–1209, 2020.

[29] R. Zhu, C. Ji, Y. Wang, Y. Cai, and H. Wu, "Heterogeneous graph convolutional networks and matrix completion for mirna-disease association prediction," *FRONTIERS IN BIOENGINEERING AND BIOTECHNOLOGY*, vol. 8, p. 901, 2020.

[30] J. Li, S. Zhang, T. Liu, C. Ning, Z. Zhang, and W. Zhou, "Neural inductive matrix completion with graph convolutional networks for mirna-disease association prediction," *BIOINFORMATICS*, vol. 36, no. 8, pp. 2538–2546, 2020.

[31] P. Xuan, D. Wang, H. Cui, T. Zhang, and T. Nakaguchi, "Integration of pairwise neighbor topologies and mirna family and cluster attributes for mirna-disease association prediction," *BRIEFINGS IN BIOINFORMATICS*, vol. 23, no. 1, p. bbab428, 2022.

[32] C. Ji, Z. Gao, X. Ma, Q. Wu, J. Ni, and C. Zheng, "Aemda: inferring mirna-disease associations based on deep autoencoder," *BIOINFORMATICS*, vol. 37, no. 1, pp. 66–72, 2021.

[33] Y. Ding, X. Lei, B. Liao, and F. Wu, "Predicting mirna-disease associations based on multi-view variational graph auto-encoder with matrix factorization," *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS*, vol. 26, no. 1, pp. 446–457, 2022.

[34] X. Chen, T. Li, Y. Zhao, C. Wang, and C. Zhu, "Deep-belief network for predicting potential mirna-disease associations," *BRIEFINGS IN BIOINFORMATICS*, vol. 22, no. 3, p. bbaa186, 2021.

[35] Y. Li, C. Qiu, J. Tu, B. Geng, J. Yang, T. Jiang, and Q. Cui, "Hmdd v2.0: a database for experimentally supported human microrna and disease associations," *NUCLEIC ACIDS RESEARCH*, vol. 42, no. D1, pp. D1070–D1074, 2014.

[36] S. Kim, L. Yeganova, and W. J. Wilbur, "Meshable: searching pubmed abstracts by utilizing mesh and mesh-derived topical terms," *BIOINFORMATICS*, vol. 32, no. 19, pp. 3044–3046, 2016.

[37] X. Wang, N. Liu, H. Han, and C. Shi, "Self-supervised heterogeneous graph neural network with co-contrastive learning," 2021.

[38] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Computer Science*, 2014.

[39] K. Hajian-Tilaki, "Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation," *Caspian journal of internal medicine*, vol. 4, no. 2, p. 627, 2013.

[40] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," *PloS one*, vol. 10, no. 3, p. e0118432, 2015.

[41] B. Xie, Q. Ding, H. Han, and D. Wu, "mircancer: a microrna-cancer association database constructed by text mining on literature," *BIOINFORMATICS*, vol. 29, no. 5, pp. 638–644, 2013.

[42] Z. Yang, L. Wu, A. Wang, W. Tang, Y. Zhao, H. Zhao, and A. E. Teschendorff, "dbdemc 2.0: updated database of differentially expressed mirnas in human cancers," *NUCLEIC ACIDS RESEARCH*, vol. 45, no. D1, pp. D812–D818, 2017.