

Predicting miRNA-disease associations via node-level attention graph auto-encoder

Huizhe Zhang[#], Juntao Fang[#], Yuping Sun*, Guobo Xie, Zhiyi Lin, Guosheng Gu

Abstract—Previous studies have confirmed microRNA (miRNA), small single-stranded non-coding RNA, participates in various biological processes and plays vital roles in many complex human diseases. Therefore, developing an efficient method to infer potential miRNA disease associations could greatly help understand operational mechanisms for diseases at the molecular level. However, during these early stages for miRNA disease prediction, traditional biological experiments are laborious and expensive. Therefore, this study proposes a novel method called AGAEMD (node-level Attention Graph Auto-Encoder to predict potential MiRNA Disease associations). We first create a heterogeneous matrix incorporating miRNA similarity, disease similarity, and known miRNA-disease associations. Then the matrix are input into a node-level attention encoder-decoder network which utilizes low dimensional dense embeddings to represent nodes and calculate association scores. To verify the effectiveness of the proposed method, we conduct a series of experiments on two benchmark datasets (the Human MicroRNA Disease Database v2.0 and v3.2) and report the averages over 10 runs in comparison with several state-of-the-art methods. Experimental results have demonstrated the excellent performance of AGAEMD in comparison with other methods. Three important diseases (Colon Neoplasms, Lung Neoplasms, Lupus Vulgaris) were applied in case studies. The results confirm the reliable predictive performance of AGAEMD.

Index Terms—miRNA, disease, deep learning, attention mechanisms, graph auto-encoder

1 INTRODUCTION

MICRORNAS (miRNAs) are a class of short endogenous non-coding RNAs containing approximately 22 nucleotides. Many studies have shown that miRNAs can not only regulate gene expression at the post-transcriptional level, but also have important roles in various biological processes, including cell proliferation, differentiation, apoptosis, signal transduction, etc. [1]. Recent findings indicate that miRNA dysregulation may contribute to human diseases. For example, Burkitt lymphoma tissues always include large quantities of miR-155 precursor [2]; miR-195 and miR-497 expression levels are correlated inversely with human breast tumor malignancy [3]; and some congenital defects and heart diseases are caused by miRNA dysregulation [4]. Known internal relationship between miRNAs and diseases to date remain just barest start of the whole picture. However, traditional experiments to provide critical data or confirmations are often costly and time-consuming. Hence several recent studies have looked for effective computational methods for latent miRNA disease association prediction.

Several new assumptions have been proposed from recent studies to help realize miRNA and disease relationships. One key assumption is that phenotypically similar diseases tend to be associated with miRNAs having similar functions [5]. This has engendered many proposed computational methods, which can be broadly summarized as

network-based or machine learning (ML) methods.

Network-based methods can be simply summarized as performing path-based algorithms on various miRNA (disease) similarity networks. Chen et al. [6] proposed RWR-MDA which performed a random walk on a miRNA-miRNA similarity network and utilized the convergence result to rank candidate miRNAs. You et al. [7] incorporated miRNA functional similarity, disease semantic similarity, Gaussian interaction profile kernel similarity for miRNAs (diseases), and known miRNA-disease associations into a depth-first search algorithm to infer miRNA disease associations. Chen et al. [8] proposed a network distance computation and a score conversion module on similarity networks to calculate association probabilities. Although impressive results have been achieved by these methods in predicting miRNA-disease associations, their performance strongly depends on the quantity of known miRNA-disease interactions and information propagation based on related similarity matrices. Thus, it's difficult for these methods to deal well with miRNA (disease) cases with few or no known related diseases (miRNAs).

In contrast, ML methods tend to use miRNA (disease) features to train supervised or semi-supervised classifiers for miRNA disease association prediction [9] [10]. For example, Jiang et al. [11] constructed feature vectors for miRNA disease associations and applied a support vector machine (SVM) on the vectors to predict potential associations. RLSMDA [12] leveraged a semi-supervised classifier based on the regularized least squares concept to predict miRNA disease associations without negative samples. Although many remarkable ML methods have been proposed, using hand-crafted features greatly limits their representational capacity, impairing predictive accuracy.

Compared with traditional ML methods, deep learning

• Huizhe Zhang, Juntao Fang, Yuping Sun, Guobo Xie, Zhiyi Lin, Guosheng Gu are with School of Computer Science, Guangdong University of Technology, Guangzhou, 510000, China.

• * Corresponding author: Yuping Sun, E-mail: syp@gdut.edu.cn.

• # These authors contributed equally.

Manuscript received xx xx, xx; revised xx xx, xx.

(DL) methods can efficiently transform raw inputs into expressive representations at higher and more abstract levels [13] [14], prompting various studies [15] [16] [17] [18] [19] to achieve excellent miRNA-disease association prediction performance. For example, Peng et al. [20] proposed a convolutional neural network (CNN) architecture to predict latent miRNA disease associations. Graph neural networks (GNNs), which is derived from concepts of convolution neural networks (CNNs) and graph embedding, can efficiently aggregate graph structured information [21]. Benefiting from strong representation learning ability and scalable frameworks, GNNs can generate low dimensional node representations graphically and hence mitigate dependence on similarity matrices. GNNs have achieved excellent performance for some practical applications [22] [23] [24]. For example, Li [25] created an end-to-end graph auto-encoder model to generate association probabilities. Zhou et al. [26] proposed a novel method called NIMCGCN to incorporate GNNs with a neural inductive matrix completion model.

Recently, attention mechanisms [27] [12], which enable networks to focus on task-relevant input portions, have also been introduced into GNNs to identify hidden relationships among miRNAs (diseases). For example, GCAN [28] performed a node-level attention mechanism on each graph convolution layer to predict disease related RNA associations. Luo et al. [29] created a hierarchical attention mechanism that adopted node-level attention to learn node representations as well as a graph-level attention to adjust contributions from different input graphs. Although great progress has been achieved, most current GNN methods have several major drawbacks that severely restrain predictive performance, in particular over-fitting and over-smoothing. Developing an appropriate GNN framework that can adequately fetch valuable information from heterogeneous networks in miRNA-disease association prediction requires further investigation.

Motivated by the aforementioned issues, we propose a novel computational method, called node-level Attention Graph Auto-Encoder to predict potential MiRNA Disease associations (called AGAEMD for short). The encoder-decoder framework of the proposed AGAEMD is inspired by previous work by Kipf et al. [30]. At first, we construct an miRNA functional similarity matrix and feature matrix containing heterogeneous graph information based on the disease semantic similarity matrix and miRNA disease adjacency matrix. Then, we feed the feature matrix into a deep graph attention network. Embeddings from each layer are input into a jumping knowledge module, and the final embedding is output as a weighted sum. Finally, an inner product decoder reconstructs the predictive association matrix by miRNA and disease embedding. AGAMED provides better predictive ability for nodes with few or even no known associations in comparison with many existing network-based and DL methods. Incorporating residual module fusing, jump knowledge module, and node-level attention mechanisms allows nodes to aggregate embedding messages from increasingly distant neighborhoods. We deployed AGAEMD on the latest Human microRNA Disease Database (HMDD) dataset, which contains more miRNAs (diseases) with few known associations, and then obtained average AUC values of 0.9173 and 0.9261 on

HMDD V2.0 and v3.2, respectively. Most of top 30 disease related miRNAs were verified in three case studies, confirming AGAEMD is a promising tool for discovering potential miRNA-disease associations.

2 MATERIALS AND METHODS

2.1 Human miRNA-disease associations

The Human MicroRNA Disease Database (HMDD) is a widely-used benchmark database that contains experimentally verified miRNA-disease associations. Among the versions of this database, v2.0 and v3.2 are chosen for conducting experiments in this paper. HMDD v2.0, which is widely used in many previous studies, includes 495 miRNAs, 383 diseases, and 5430 human miRNA-disease association entries verified by biological experiments, and can be directly downloaded from <https://www.cuilab.cn/hmdd>. Recently, a more challenging version of the database, HMDD v3.2 [31] has been released, which covers more entries than HMDD v2.0, with the number of associations, miRNAs, and diseases to be 35547, 1206, and 893, respectively. We downloaded the latest MeSH disease descriptors from <https://www.nlm.nih.gov/databases/download/mesh.html> for experiments. There are 554 common diseases that can be found from Mesh and HMDD v3.2. Besides, we integrated miRNAs that produce the same mature miRNA into one group, whose related disease information are also integrated. After such preprocessing, we constructed a dataset derived from the original HMDD v3.2 for the following experiments. Table 1 shows summary statistics for the two benchmark datasets.

TABLE 1: Statistics for the data used in the experiments

dataset	miRNAs	Diseases	Associations
HMDD v2.0	495	383	5430
HMDD v3.2	915	554	11742

2.2 Disease semantic similarity

The computation of disease semantic similarity is derived from Wang's study [32]. According to MeSH descriptors, the hierarchical relations of diseases can be easily transformed into a Directed Acyclic Graph (DAG), where nodes represent diseases and edges represent 'is-a' relationship between diseases. Then relative position of two diseases on the DAG can be utilized to compute similarity coefficients for them. For example, suppose $DAG_d = (A_d, E_d)$ is the DAG for disease d , where A_d represents a set of its all ancestor nodes (including d itself), and E_d denotes a set of corresponding edges between ancestor nodes and their children. Disease semantic similarity defines the semantic value S_d for a disease d as the sum of contributions of all its ancestors,

$$S_d = \sum_{d_i \in A_d} C_d(d_i) \quad (1)$$

$$C_d(d_i) = \begin{cases} 1, & d_i = d \\ 0.5^{lp(d, d_i)}, & d_i \neq d \end{cases}, \quad (2)$$

where $C_d(d_i)$ denotes the contribution of disease d_i in DAG to the semantics value of disease d , $lp(d, d_i)$ denotes the

shortest path length between d_i and d . The disease semantic similarity scores assumes that diseases with more common ancestors will have larger similarity scores. Hence given any two disease d'' and d' , their corresponding disease semantic similarity S_D can be expressed as

$$S_D(d', d'') = \frac{\sum_{d_j \in A_{d'} \cap A_{d''}} C_{d'}(d_j) + C_{d''}(d_j)}{S_{d'} + S_{d''}}. \quad (3)$$

Applying equations 1,2,3 on data of HMDD v2.0 and v3.2 yields disease semantic similarity matrices with dimension 383×383 and 554×554 respectively.

2.3 MiRNA functional similarity

Many previous studies obtained miRNA functional similarity directly from <https://www.cuilab.cn/files/images/cuilab/misim.zip> [32], even though all miRNA similarity information in the file comes from HMDD v1.0. Thus, we cannot directly obtain functional similarity for any miRNA newly added in HMDD v2.0 and v3.2, and cannot adopt the simple way to fill all absent entries in functional similarity matrix with zeros as it will seriously impairs predictive accuracy. To solve the aforementioned problem, we adopt an algorithm to recalculate functional similarity between two certain miRNAs e_i and e_j [33] as follows:

$$S_M(e_i, e_j) = \frac{\sum_{d \in D(e_i)} S_D(d, d_j^*) + \sum_{d \in D(e_j)} S_D(d, d_i^*)}{|D(e_i)| + |D(e_j)|}, \quad (4)$$

where $D(e_i)$ is a set of diseases connected with e_i , $|D(e_i)|$ is the number of elements in $D(e_i)$, S_M is the miRNA functional similarity matrix, and d_i^* denotes the disease in $D(e_i)$ that have the largest semantic similarity with d which is calculated as follows:

$$d_i^* = \operatorname{argmax}_{d_i \in D(e_i)} S_D(d, d_i). \quad (5)$$

2.4 AGAEMD

Graph attention networks have been widely used in many prediction tasks [27], [34], and some studies in the field of bioinformatic, such as LAGCN [35] and EGATMDA [29], have achieved impressive predictive performance. Encouraged by these achievements, we propose AGAEMD, a node-level attention graph auto-encoder model to predict latent miRNA disease associations. The proposed AGAEMD method, outlined in Fig.1, mainly consists of three steps, including the construction of miRNA-disease heterogeneous network, a node-level attention graph encoder and an inner product decoder. Each step will be detailed in the following subsections.

2.4.1 MiRNA-disease heterogeneous network

We integrate miRNA disease associations, miRNA-miRNA similarity, and disease-disease similarity to create a heterogeneous network. Specifically, we construct an $M \times N$ binary adjacency matrix A to represent all possible miRNA-disease pairs, where M and N are the number of miRNAs and diseases, respectively. The nonzero with value 1 occurs at the entry with row index i and column index j if miRNA

m_i is experimentally verified to be connected with disease d_j . We also construct disease and miRNA similarity matrices S_D and S_M by following the calculations in Sections 2.2 and 2.3 respectively. Then the heterogeneous network H can be calculated as follows:

$$H = \begin{bmatrix} S_M & A \\ A^T & S_D \end{bmatrix}, \quad (6)$$

where $H \in \mathbb{R}^{(M+N) \times (M+N)}$. The adjacency matrix G on the heterogeneous network is denoted as

$$G = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \in \mathbb{R}^{(M+N) \times (M+N)}. \quad (7)$$

2.4.2 Node-level attention graph auto-encoder

In the next, we construct a graph auto-encoder with the application of an node-level attention mechanism and deploy this auto-encoder on H to learn low-dimensional representation for nodes. In contrast with some graph attention methods which output a weighted sum of each layer embedding [36], AGAEMD not only keeps the graph substructure information contained in each embedding, but also adaptively captures representations of neighboring nodes for each node. AGAEMD treats H as a feature matrix, and we design a corresponding projection matrix $W \in \mathbb{R}^{(M+N) \times F}$ to project node features of H onto a F -dimensional feature space as follows:

$$H' = HW, \quad (8)$$

where F is the hidden embedding dimension of the above feature space. Let $h_i \in \mathbb{R}^{M+N}$ and $h_i' \in \mathbb{R}^F$ denotes the i -th row of the matrix H and H' respectively, each of which corresponds to a certain node. And then we perform the node-level attention mechanism on these projected nodes. Given a projected node h_i' , the set of its first-order neighborhood which is determined by edge indices selected from G is denoted as N_i and the weight for the node's every related edge is denoted as e_{ij} :

$$e_{ij} = a^T \sigma(W(h_i' + h_j')), \forall j \in N_i, \quad (9)$$

where $\sigma(\cdot)$ is the leaky ReLU function which performs activation operation for every element of the input vector, and $a \in \mathbb{R}^F$ is a shared attention row vector that can be treated as a single-layer feedforward neural network to extract structure information from the corresponding node pair. In the next, we normalize the weight e_{ij} to get attention coefficient by using the softmax function as follows

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})}. \quad (10)$$

Thus, we can integrate all embeddings corresponding to the first-order neighborhood as a weighted sum

$$h_i'' = \sigma\left(\sum_{j \in N_i} \alpha_{ij} h_j'\right), \quad (11)$$

where α_{ij} is the attention coefficient for the projected feature h_j' . Vaswani et al. [37] made a modification to a single attention model by applying multiple independent attention

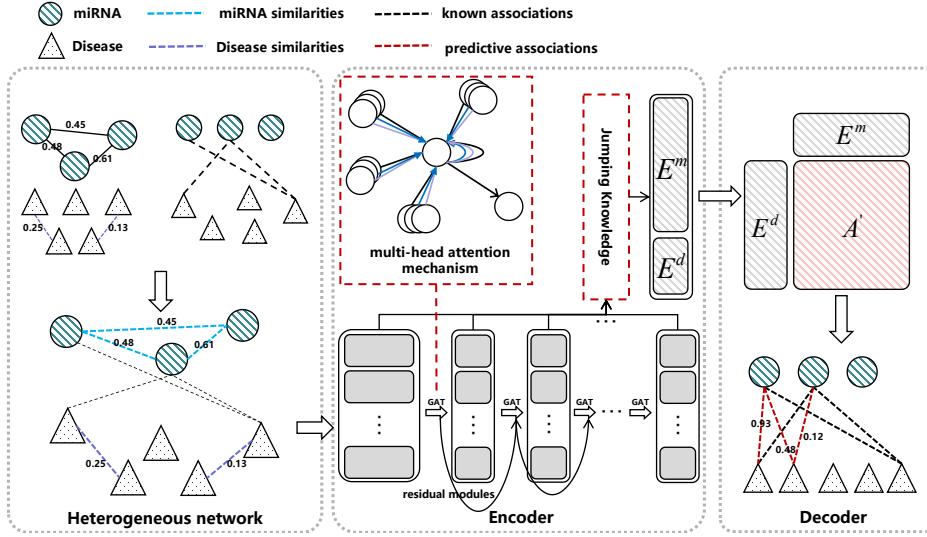


Fig. 1: The workflow of a node-level attention graph auto-encoder model. As depicted in the figure, AGAEMD consists of three modules including heterogeneous network construction, encoder and decoder. At first, miRNA disease associations, miRNA-miRNA similarity, and disease-disease similarity are integrated as a heterogeneous matrix H . Secondly, H and its corresponding adjacent matrix G , are treated as the input data of encoder. Thirdly, our node-level attention auto-encoder aggregates message from first-order neighborhoods and output a fusion of embeddings generated by each layer. Finally, AGAEMD calculates the miRNA-disease associations score matrix A' through the inner product operation between a miRNA embedding E^m and a disease embedding E^d .

functions to a model simultaneously. It's observed that such multi-head attention allows the model to jointly attend to information from different representation subspaces at different positions, which will help to achieve stable performance for the model. The encoder outputs the final embeddings by averaging the results from T attention heads

$$h_i^* = \sigma\left(\frac{1}{T} \sum_{t=1}^T h_i^{(t)}\right). \quad (12)$$

Residual modules We proposed a deep GNN framework which includes both residual and weighted sum modules and can be treated as a variant of DeepGCN [38]. With the employment of residual modules, convolution layers with residual structure tend to learn a residual mapping $F(x) = H(x) - x$ instead of the standard mapping $H(x)$, which makes the network easier to train. We incorporate residual modules with the embedding updating process for our GNN framework as follows,

$$\begin{aligned} H^{l+1} &= H_{agg}^l + H_{res}^l \\ &= \text{Aggregate}(H^l, N_H) + \text{Linear}(H^l), \end{aligned} \quad (13)$$

where H_{agg}^l , H_{res}^l denote the l -th aggregated representations and a projective representation respectively, N_H is the set of neighboring nodes. We adopt a averaging aggregator in corporation with a node-level attention mechanism to update node representations in our experiment.

Jumping knowledge Instead of directly obtaining embedding on output layer, AGAEMD calculates a weighted sum of embeddings extracted from every layer by feeding these embedding into a jumping knowledge module [39]. Under the premise of having the same size of receptive field, the

information that could be aggregated for a node from its neighborhood relies on its location in the heterogeneous network. For example, vertices located near the topological boundary of the graph may require a deeper network to aggregate more messages and generate informative representations; while a deeper network for nodes located nearby the topological center of the graph may cause severe over-squashing. Therefore, for different nodes, the importances of embeddings generated by each layer are different.

In addition, we also observe that a deeper embedding can capture the wider neighborhood information based on the previous embedding. It means that there are connections between embeddings. Recurrent Neural Networks (RNNs) could be the desired tool for dealing with the output embedding sequence from AGAEMD. But the traditional RNNs have difficulties in bringing long time lags.

Specifically, the core of LSTM-attention layer is a bidirectional LSTM, which is a variant of RNN introduced by Hochreiter and Schmidhuber [40]. A memory cell rather than the repeating module in traditional RNNs is utilized in the bidirectional LSTM. The memory cell outputs hidden and cell states and enables the model to learn long-term dependencies. As shown in Fig.2, the bidirectional LSTM can further learn hidden connections between embeddings along backward and forward directions of the embedding sequence. And the jumping knowledge integrates the output features from the bidirectional LSTM with the embedding sequence to generate the final embedding. For the embedding extracted from i -th layer H^i , the jumping knowledge module will create its backward features L^i and forward features R^i by utilizing a bi-directional LSTM, and perform a linear mapping function to turn $L^i + R^i$ into a weight vector. Fig.2 shows the structure of the jump-

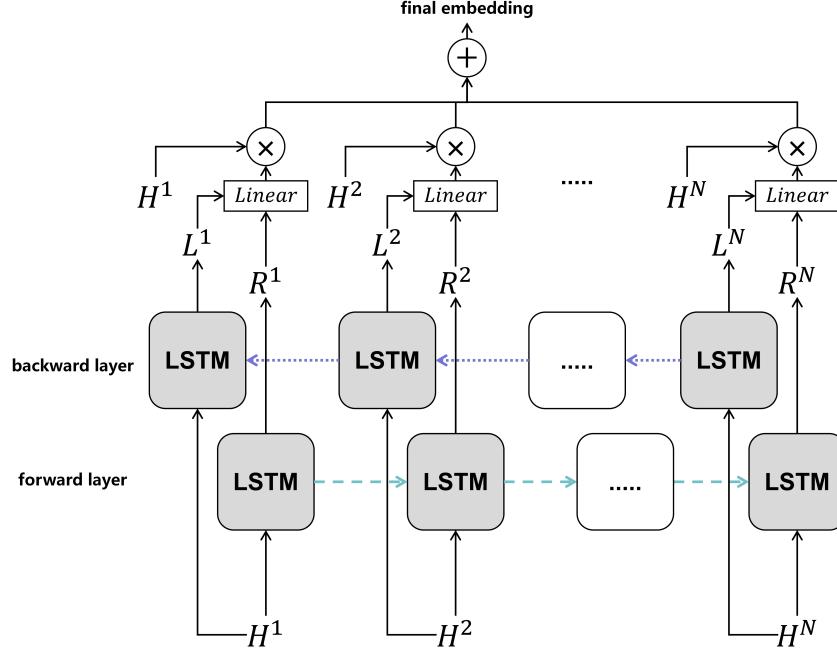


Fig. 2: The structure of the jumping knowledge module.

ing knowledge module. The process of applying jumping knowledge module is formulated as follows:

$$\{(L^1, R^1), (L^2, R^2), \dots, (L^N, R^N)\} = LSTM(H^1, H^2, \dots, H^N) \quad (14)$$

$$H^n = \begin{bmatrix} H_m \\ H_d \end{bmatrix} = \sum_{i=1}^N a_i H^i = \sum_{i=1}^N Linear([L^i || R^i]) H^i \quad (15)$$

where a_i is the attention coefficient for embedding from the i -th layer.

Dropout The performance of many traditional deep GNN network is severely limited by the overfitting problem. Therefore, we introduce a regular dropout technique [41] to alleviate this problem and improve predictive performance of our model. By randomly dropping units in the network, dropout makes network train a number of subnetworks sampled from the original framework, which provides a simple and efficient way to approximately combine different results from different subnetwork architectures. In our implementation, the dropout is set to 0.6.

2.4.3 Inner product decoder

After the execution of the aforementioned steps, we could obtain the final embeddings as:

$$H^n = \begin{bmatrix} h_1^n \\ h_2^n \\ \vdots \\ h_M^n \\ h_{M+1}^n \\ \vdots \\ h_{M+N-1}^n \\ h_{M+N}^n \end{bmatrix} = \begin{bmatrix} H_m \\ H_d \end{bmatrix}, \quad (16)$$

where $H_m \in \mathbb{R}^{M \times F}$ and $H_d \in \mathbb{R}^{N \times F}$ denote the final miRNA and disease embeddings, respectively. We implement an inner product decoder to reconstruct the final adjacency matrix. Since miRNA-disease association prediction can be treated as a specific binary classification problem, we choose the sigmoid function as the activation function in decoder, which is widely used for binary classification tasks [25]. After decoding, AGAEMD outputs a predictive score matrix as follows,

$$A' = \text{sigmoid} \left(H_m H_d^T \right) \quad (17)$$

where A'_{ij} denotes the predictive association score between a miRNA m_i and a disease d_j . In the experiments, we used all the known miRNA-disease associations as positive examples and treated others as negative examples, where the sets of positive and negative instances are represented as y^+ and y^- , respectively. Unfortunately, so far, few miRNA-disease associations have been recorded, leading to the fact that the number of positive samples is much less than that of negative ones. To mitigate the impact of class imbalance, we set a factor λ to increase the weight for known associations,

$$Loss = -\frac{1}{M \times N} \times \left(\lambda \times \sum_{(i,j) \in y^+} \log A'_{ij} + \sum_{(i,j) \in y^-} \log (1 - A'_{ij}) \right), \quad (18)$$

where $\lambda = \frac{|y^-|}{|y^+|}$, $|y^+|$ and $|y^-|$ are the number of samples in y^+ and y^- , respectively.

3 RESULTS

We implemented AGAEMD on PyTorch and the source codes are available at <https://github.com/Zhhuizhe/AGAEMD>. At the beginning of the training process, Xaiver initialization technique [42] is applied to initialize all trainable weight matrices, and the loss function is minimized using Adam optimizer [43]. Grid search is used to select the best hyperparameters, including the number of heads T , the hidden embedding dimension F , the number of hidden layers n , the number of epochs e , the learning rate lr and the weight decay wd . By performing the grid search on HMDD v3.2, we set optimal hyperparameters lr as 10^{-5} , e as 2000, wd as 10^{-3} , T as 4, F as 200, n as 4 and the selection scheme for T , F , n is discussed in Section 3.2. In our experiments, AGAEMD obtained optimal predictive result on HMDD v2.0 by setting lr as 10^{-4} and other hyperparameters as the same as those of HMDD v3.2.

In experiments, we randomly split all the used associations into 4 types: training message, training supervision, validation, and test edges. Each type of the associations consists of certain percentage of positive samples (20%, 16%, 16% and 48% as test, validation, training supervision and training message edges respectively) together with randomly selected negative samples with the same number of positive ones to keep the class balance all the time. We used training message edges to predict training supervision edges, and training message and training supervision edges to predict validation edges. All the association prediction experiments are repeated 10 times to obtain the averages of AUPR, AUC, F1-score, and Accuracy to evaluate performance of the models. In addition, We use 5-fold cross-validation (5-fold CV) based on training and validation edges to modify AGAEMD structure and optimize parameter settings.

3.1 Comparison with other methods

We compared AGAEMD with several state-of-the-art methods [44] [45] [46] [33] [47] [48] [48] [49] [50] by conducting a series of experiments on HMDD v2.0 and v3.2. All the compared methods were specified to use only disease semantic and miRNA functional similarity calculated dynamically during training. We optimized the learning rate, number of epochs, and dropout score for all GNN-based methods by using grid search technique. See Table 2 for the performance comparison. AGAEMD achieved the highest AUPR, AUC of 91.726%, 91.535% on HMDD v2.0 and the highest AUPR, AUC scores of 93.003%, 92.614% on HMDD v3.2. The AUC score of AGAEMD is 1% and 2.9% higher than TDRC, despite AGAEMD slightly trailing TDRC with respect to F1-score and Accuracy. Compared with three path-based methods (WBPMD, GLNMDA, MCLPMDA), AGAEMD have a obvious advantage in all evaluation metrics. Besides, AGAEMD significantly outperforms the other two GNN-based methods (GAEMDA and MMGCN) and effectively predicts unverified miRNA disease associations. The superiority of AGAEMD over these methods can be attributed to its being able to partly alleviate two common problems about dealing with less-annotated nodes and the strong dependence of functional similarity.

3.2 predictive results for less-annotated or none-annotated miRNA nodes

As mentioned above, AGAEMD achieves outstanding predictive performance for less-annotated or none-annotated nodes by learning low dimensional representations for nodes during the training process. To demonstrate the superior predictive performance of AGAEMD for these nodes, we consider the case of miRNA to investigate the influence of the number of associations on predictive results. At first, we randomly select 80% of the whole positive samples into the training set, together with randomly chosen unknown associations as negative ones, making the quantities of two categories equal for each miRNA. The rest positive samples with the balanced amount of randomly chosen negative samples for each miRNA are used as the testing set. Based on this training/testing partition (denoted as P), we assign those less-annotated miRNAs into 3 disjoint groups based on degree of each miRNA, which is natural because degree here is equivalent to the number of diseases related to this miRNA. Besides, we construct a dot product (denoted as Baseline hereafter) between miRNA functional similarity matrix and miRNA disease adjacency matrix as benchmark, and we also choose PMFMDA [46] as the other method for comparison. We adopt average accuracy of 10 runs as the evaluation metric because some miRNAs do not have negative and positive samples simultaneously in the test set. Table.3 summarizes the experiment results for the three methods regarding the less-annotated or none-annotated miRNAs.

Group L_1-L_3 represent miRNAs with few or even no known associations in the network. As shown in Table.3, accuracy scores of AGAEMD on all groups are much higher than other comparison methods. To further explore the influence of the degree on the prediction accuracy, we further adjust the partition P into two different new partitions by increasing the degree of each miRNA for each less-annotated group. The degree variation (denoted as $dV_m(\alpha)$) for each miRNA m can be defined as the rounding down of $\alpha\%$ of the quantity of related positive samples in the testing set, where α are set to 25 and 50 for the first (P_1) and second (P_2) adjustment respectively. For example, in order to adjust P into P_1 for group L_2 , for every miRNA m belonging to L_2 , $dV_m(25)$ positive samples and $dV_m(25)$ negative samples that are related to miRNA m are randomly selected from the testing set and then moved to the training set, while all other samples remain unchanged. Table.4 summarizes the experimental results regarding the change of degree for less-annotated or no-annotated miRNAs.

It can be observed from Table.4 that the more the degrees of less-annotated miRNAs increase, the better accuracy scores of all compared methods could obtain. In comparison with Baseline and PMFMDA, the performance of AGAEMD is not that sensitive to the degree of miRNA, showing its great potential in discovering associations for miRNAs (diseases) with few or even no annotations.

3.3 Effects of parameters

This section analyzes AGAEMD parameter sensitivity. we conducted a series of experiments on HMDD v3.2 to investigate influence of parameters T , F , n on average AUC scores

TABLE 2: Comparison with other methods on HMDD v2.0 and HMDD v3.2

dataset	Comparison methods	AUPR	AUC	F ₁ -score	Accuracy
HMDD v2.0	GAEMDA [25]	88.402% ± 0.00831	89.069% ± 0.00608	77.06% ± 0.02149	79.664% ± 0.01365
	PMFMDA [46]	91.535% ± 0.00465	90.476% ± 0.00276	80.833% ± 0.011	82.859% ± 0.00872
	TDRC [33]	89.952% ± 0.00307	90.581% ± 0.00533	83.952% ± 0.00653	84.263% ± 0.00625
	MMGCN [51]	77.179% ± 0.02434	82.79% ± 0.0192	79.529% ± 0.00992	78.034% ± 0.01529
	WBPMD [47]	86.289% ± 0.00359	86.037% ± 0.00774	78.012% ± 0.00539	78.167% ± 0.00698
	EDTMDA [48]	85.741% ± 0.00408	86.452% ± 0.00165	77.991% ± 0.00122	74.344% ± 0.00279
	GLNMDA [49]	90.603% ± 0.00699	90.241% ± 0.00303	83.04% ± 0.00224	82.394% ± 0.00296
	MCLPMDA [50]	82.61% ± 0.00286	82.975% ± 0.00463	76.913% ± 0.00597	74.658% ± 0.00209
	AGAEMD	91.726% ± 0.00387	91.535% ± 0.00406	83.748% ± 0.0087	83.766% ± 0.00866
HMDD v3.2	GAEMDA [25]	88.992% ± 0.00626	89.023% ± 0.00237	82.83% ± 0.00672	83.492% ± 0.00553
	PMFMDA [46]	92.495% ± 0.00256	91.226% ± 0.00333	80.853% ± 0.00574	83.079% ± 0.00478
	TDRC [33]	92.216% ± 0.00055	89.742% ± 0.00117	85.95% ± 0.00152	86.082% ± 0.00175
	MMGCN [51]	77.933% ± 0.00872	83.491% ± 0.00724	80.337% ± 0.00697	79.084% ± 0.00679
	AGAEMD	93.003% ± 0.00321	92.614% ± 0.00285	85.068% ± 0.00603	85.085% ± 0.00477

TABLE 3: The accuracy scores for testing samples on different groups.

Interval of degree	L_1 (0)	L_2 (1-2)	L_3 (3-6)
Baseline	0.483	0.573	0.591
PMFMDA [46]	0.376	0.454	0.53
AGAEMD	0.732	0.733	0.744

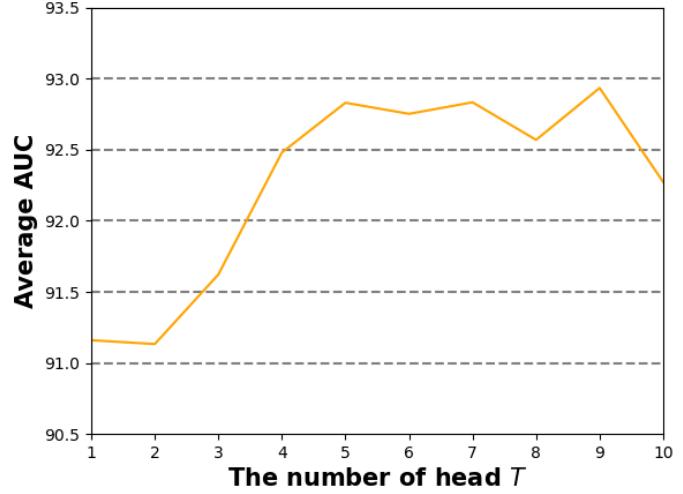
over 10 runs, by alternatively adjusting one parameter while fixing the other two.

(1) Effects of varying the number of heads T : AGAEMD adopts a node-level attention mechanism with multi-heads, hence there are various attention vectors to assign weights for different neighbors. For a given node, AGAEMD can use different views to aggregate neighbor information, mitigating the problem of neglecting important nodes to some extent, but also causing explosive growth in the number of parameters.

(2) Effects of varying the hidden embedding dimension F : Embedding dimension F also decides the dimensionality for trainable weighting matrices W in each layers. The larger F is, the longer the model training time would be. Thus we limit that F is less than or equal to 400, and adopt grid search for the decision of optimal dimension.

(3) Effects of varying the number of hidden layers n : Increasing the number of layers can allow nodes to collect more structure information from the heterogeneous network, but also exponentially increases the number of neighbors and tends to increase loss of neighbor information during information aggregation. AGAEMD adopts a deep GNN framework to relieve this problem. Therefore, we conduct a test to analyze the influence of the number of layers in the proposed method.

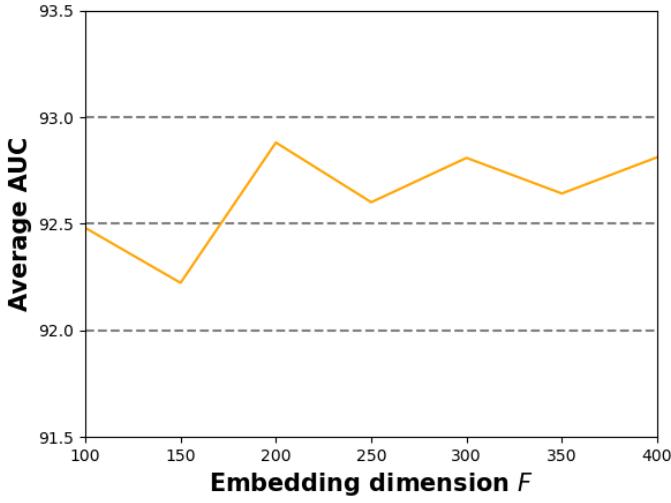
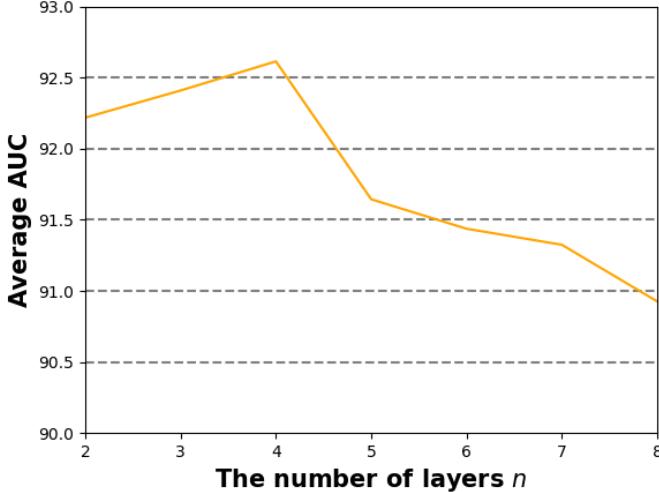
In our implementation, we first use grid search to obtain an approximate range for one parameter with the other two parameters fixed and then use a narrower and more

Fig. 3: Parameter sensitivity for the number of heads T

precise interval to find the optimal parameter value. Fig.3 shows that best average AUC can be obtained when n is 4. In many traditional GNN methods, as the network goes deeper, the embeddings of nodes within same connected components gradually become similar and even much the same after certain layers. Introducing the node-level attention mechanism and the jumping knowledge module can partly alleviate over-smoothing. It enables the depth of our model can be further increased to 4. The performance of our method drops much when n is larger than 4. As illustrated in Fig.4, a small F may limit the performance of AGAEMD. Besides, AUC growth slows when F exceeds 200. Similarly, the larger T is, the larger average AUC score will be obtained within a certain range (see Fig.3), and gradually flattens when T is greater than 5. Therefore, we set T as 5, F as 512, n as 4 to obtain an optimal performance. In

TABLE 4: The accuracy scores for testing samples of different groups with the varying degree conditions (%)

	L_1			L_2			L_3		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
Baseline	0.483	0.723	0.689	0.573	0.728	0.761	0.591	0.698	0.692
PMFMDA [46]	0.376	0.547	0.566	0.454	0.634	0.667	0.53	0.655	0.691
AGAEMD	0.732	0.727	0.712	0.733	0.751	0.758	0.744	0.753	0.762

Fig. 4: Parameter sensitivity for embedding dimension F Fig. 5: Parameter sensitivity for the number of layers n

experience, the classification performance using this scheme is stable regardless the randomness of sampling. It can be observed in experiments that the performance of AGAEMD is stable regardless the parameter fluctuation, showing its strong robustness.

3.4 Ablation experiments

AGAEMD includes three key components: node-level attention mechanism, residual modules, and jump knowledge. We performed ablation experiments on HMDD v3.2 to

further verify the contributions of these components to the predictive ability of our method, and the averages of AUPR, AUC, F1-score, and Accuracy on 10 runs are reported. We rebuild three different comparison models by removing one or two of the three key components of AGAEMD or replacing them with normal operations.

- AGAEMD-RES model retained the backbone graph attention layer and the jumping knowledge module and removed the residual structure from the original framework.
- AGAEMD-JK model retained the GAT backbone and residual modules and replaced the jumping knowledge module with a simple concatenation operation.
- AGAEMD-GAT model retained residual and jumping knowledge modules and replaced the GAT backbone with a normal GCN backbone.

To have a fair comparison, we kept the experimental setting same for all the compared methods and adopted a grid search to obtain optimal performance for them. The results are demonstrated in Table 5. It can be observed that the performance of AGAEMD-RES is worse than that of AGAEMD as it suffer from one major drawback of original GNN framework, i.e. a large network depth making the model hard to be trained. By introducing the residual structures, AGAEMD-JK could implicitly learn a residual mapping and outputs a residual node representation at each layer. Such modifications significantly alleviate the vanishing gradient problem and improve its representation learning ability. Hence, AGAEMD-JK achieved slight improvement in F1-score in comparison with AGAEMD-RES. The performance of AGAEMD-JK is still worse than that of AGAEMD, because embedding a jumping knowledge module in the deep framework ensures that the representations for nodes with few or no associations can still be well processed by AGAEMD. AGAEMD-GAT performed the worst among all, implying the effectiveness of introducing node-level attention mechanism. For node-level attention mechanism, attention coefficients are used to adjust the priority of messages sent by the different neighboring nodes in order to aggregate and update node embeddings, which is beneficial for GAT's achieving better predictive performance than the normal GCN.

3.5 Case study

We select three different diseases as case studies: Colon Neoplasms, Lung Neoplasms and Lupus Vulgaris. Many experiments show that cancers are strongly related with miRNAs' deregulation, and uncovering the roles of miRNAs in these neoplasms may develop the understanding of neoplasms' pathogenesis, diagnosis and so on [52] [53] [54]. For example, the serum exosomal levels of several miRNAs like let-7a, miR-1229 in primary colorectal cancer cases

TABLE 5: Ablation Experiments Results of four different models on HMDD v3.2

Comparison methods	AUPR	AUC	F ₁ -score	Accuracy
AGAEMD	93.003% ± 0.00321	92.614% ± 0.00285	85.068% ± 0.00603	85.085% ± 0.00477
AGAEMD-RES	92.588% ± 0.00599	91.832% ± 0.00742	85.063% ± 0.00466	85.532% ± 0.00464
AGAEMD-JK	92.015% ± 0.00253	90.847% ± 0.00314	85.271% ± 0.00338	85.328% ± 0.00333
AGAEMD-GAT	91.974% ± 0.00437	91.474% ± 0.00496	83.84% ± 0.00617	83.30% ± 0.00843

TABLE 6: Top 30 predicted miRNAs associated with Colon Neoplasms

miRNA	Evidence	miRNA	Evidence
hsa-mir-21	dbDEMC 3.0;HMDD v3.2	hsa-mir-141	dbDEMC 3.0;HMDD v3.2
hsa-mir-155	dbDEMC 3.0;HMDD v3.2	hsa-mir-1	dbDEMC 3.0;HMDD v3.2
hsa-mir-206	dbDEMC 3.0	hsa-mir-16	HMDD v3.2
hsa-mir-486	dbDEMC 3.0;HMDD v3.2	hsa-mir-373	HMDD v3.2
hsa-mir-145	dbDEMC 3.0;HMDD v3.2	hsa-mir-214	dbDEMC 3.0;HMDD v3.2
hsa-mir-7	dbDEMC 3.0	hsa-mir-34c	HMDD v3.2
hsa-mir-199a	HMDD v3.2	hsa-mir-200b	dbDEMC 3.0;HMDD v3.2
hsa-mir-210	dbDEMC 3.0;HMDD v3.2	hsa-mir-182	dbDEMC 3.0;HMDD v3.2
hsa-mir-203	dbDEMC 3.0;HMDD v3.2	hsa-mir-192	dbDEMC 3.0;HMDD v3.2
hsa-mir-34a	dbDEMC 3.0;HMDD v3.2	hsa-mir-224	dbDEMC 3.0;HMDD v3.2
hsa-mir-223	HMDD v3.2	hsa-mir-148a	dbDEMC 3.0;HMDD v3.2
hsa-mir-195	dbDEMC 3.0;HMDD v3.2	hsa-mir-27a	dbDEMC 3.0;HMDD v3.2
hsa-mir-25	dbDEMC 3.0;HMDD v3.2	hsa-mir-101	HMDD v3.2
hsa-mir-125b	dbDEMC 3.0;HMDD v3.2	hsa-mir-150	dbDEMC 3.0;HMDD v3.2
hsa-mir-221	dbDEMC 3.0;HMDD v3.2	hsa-mir-193b	dbDEMC 3.0

are higher than in healthy controls [55]. Besides, through downregulating Cdc42 and Cdk6 expression, miR-137 has an influence on behaviors of lung cancer cell [56]. Therefore, Colon Neoplasms and Lung Neoplasms, which are common causes of cancer death, are selected for case study. Except for above two diseases which were studied widely, we also focus on those diseases which only has few related miRNAs verified by experiments. Lupus Vulgaris meets this requirement. Similarly, existing experiments also show that miRNAs may become potential diagnosis biomarkers of Lupus Vulgaris [57].

In experiments, predictive results are verified by independent sources and/or public reports, hence we choose reliable datasets dbDEMC 3.0 <https://www.picb.ac.cn/dbDEMC/> and HMDD v3.2 [19] www.cuilab.cn/hmdd as verification sets. We first trained AGAEMA on HMDD v2.0 using 5430 known miRNA disease associations, and then used the trained model to calculate association scores for specific diseases. We ranked the prediction scores in descending order, screened out the top 30 disease related miRNAs, and verified using dbDEMC 3.0 and HMDD v3.2. AGAEMD identified 30, 30, and 21 of the top 30 Colon Neoplasms (Table 6), Lung Neoplasms (Table 7), and Lupus Vulgaris (Table 8) related miRNAs, respectively, verified by dbDEMC 3.0 and HMDD v3.2. Thus, what has been confirmed is that AGAEMD can achieve impressive and stable predictive performance for practical diseases.

4 CONCLUSION

Many previous studies have shown that identifying relationships between miRNAs and diseases may help human understand diseases pathogenesis and develop better treatments for complex diseases. However, traditional biological experiments are always costly and time-consuming. Therefore, it becomes urgent to develop a more effective and robust computational approach for identifying undiscovered miRNA-disease interactions. This study proposed AGAEMD, which embeds a node-level attention graph layer into a deep GNN framework. By introducing the attention mechanism, AGAEMD can adaptively capture internal information among nodes in heterogeneous network. The deep GNN framework which consists of residual modules and a jumping knowledge module not only alleviates over-squashing and over-smoothing, but also makes model pay attention to the representations of less-annotated nodes. We conducted a series of experiments to verify the effectiveness and robustness of our proposed AGAEMD. Experimental results showed that AGAEMD outperforms many state-of-the-art computational models with several metrics on both HMDD v2.0 and HMDD v3.2. What's more, in case study, 30, 30 and 21 of top 30 disease-related miRNAs discovered by AGAEMD can be confirmed by dbDEMC 3.0 and HMDD v3.2.

Although AGAEMD is effective at uncovering potential miRNA disease associations, there is still considerable room for improvement. As we all know, the quantity of the known miRNA-disease associations is very important

TABLE 7: Top 30 predicted miRNAs associated with Lung Neoplasms

miRNA	Evidence	miRNA	Evidence
hsa-mir-17	dbDEMC 3.0;HMDD v3.2	hsa-let-7a	dbDEMC 3.0;HMDD v3.2
hsa-mir-92a	dbDEMC 3.0;HMDD v3.2	hsa-mir-143	dbDEMC 3.0;HMDD v3.2
hsa-mir-21	dbDEMC 3.0;HMDD v3.2	hsa-mir-145	dbDEMC 3.0;HMDD v3.2
hsa-mir-19a	dbDEMC 3.0;HMDD v3.2	hsa-mir-181b	dbDEMC 3.0;HMDD v3.2
hsa-mir-20a	dbDEMC 3.0;HMDD v3.2	hsa-mir-29b	dbDEMC 3.0;HMDD v3.2
hsa-mir-34a	dbDEMC 3.0;HMDD v3.2	hsa-mir-222	dbDEMC 3.0;HMDD v3.2
hsa-mir-146a	dbDEMC 3.0;HMDD v3.2	hsa-let-7c	dbDEMC 3.0;HMDD v3.2
hsa-mir-18a	dbDEMC 3.0;HMDD v3.2	hsa-mir-126	dbDEMC 3.0;HMDD v3.2
hsa-mir-221	dbDEMC 3.0;HMDD v3.2	hsa-mir-106b	dbDEMC 3.0;HMDD v3.2
hsa-mir-125b	dbDEMC 3.0;HMDD v3.2	hsa-mir-155	dbDEMC 3.0;HMDD v3.2
hsa-mir-29a	dbDEMC 3.0;HMDD v3.2	hsa-let-7b	dbDEMC 3.0;HMDD v3.2
hsa-mir-15a	dbDEMC 3.0;HMDD v3.2	hsa-mir-182	dbDEMC 3.0;HMDD v3.2
hsa-mir-181a	dbDEMC 3.0;HMDD v3.2	hsa-mir-200b	dbDEMC 3.0;HMDD v3.2
hsa-mir-19b	dbDEMC 3.0;HMDD v3.2	hsa-let-7g	dbDEMC 3.0;HMDD v3.2
hsa-mir-16	dbDEMC 3.0;HMDD v3.2	hsa-mir-34c	dbDEMC 3.0;HMDD v3.2

TABLE 8: Top 30 predicted miRNAs associated with Lupus Vulgaris

miRNA	Evidence	miRNA	Evidence
hsa-mir-30a	HMDD v3.2	hsa-mir-296	HMDD v3.2
hsa-mir-146a	HMDD v3.2	hsa-mir-99a	HMDD v3.2
hsa-mir-612	HMDD v3.2	hsa-mir-223	HMDD v3.2
hsa-mir-181a	Unconfirmed	hsa-mir-663a	HMDD v3.2
hsa-let-7f	Unconfirmed	hsa-mir-302a	Unconfirmed
hsa-mir-365b	HMDD v3.2	hsa-mir-654	HMDD v3.2
hsa-mir-575	HMDD v3.2	hsa-mir-516a	HMDD v3.2
hsa-let-7i	Unconfirmed	hsa-mir-30d	HMDD v3.2
hsa-mir-302c	Unconfirmed	hsa-mir-106b	Unconfirmed
hsa-mir-601	HMDD v3.2	hsa-mir-516b	Unconfirmed
hsa-mir-658	HMDD v3.2	hsa-mir-20b	Unconfirmed
hsa-mir-185	HMDD v3.2	hsa-let-7e	HMDD v3.2
hsa-mir-608	HMDD v3.2	hsa-mir-532	HMDD v3.2
hsa-mir-197	HMDD v3.2	hsa-mir-637	HMDD v3.2
hsa-mir-138	Unconfirmed	hsa-mir-500a	HMDD v3.2

for training a powerful prediction method. In the future, when more miRNA-disease associations are experimentally verified and added to the current database together with abundant and accurate supplementary data, the predictive ability of AGAEMD will be improved. Although the proposed AGAEMD could recommend related diseases (miRNAs) for miRNAs (diseases) with few or even no confirmed associations, we have found that the embedding generation technique of AGAEMD could be modified to be more effective in comparison with some state-of-the-art representation learning methods.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (62002070, 61802072), the Natural Science Foundation of Guangdong Province (2018A030313389),

the Opening Project of Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University (Grant No. 2021013), and the Science and Technology Plan Project of Guangzhou City (Grant No. 201902020006, 201902020012, 202102021236). Yuping Sun is the corresponding author.

REFERENCES

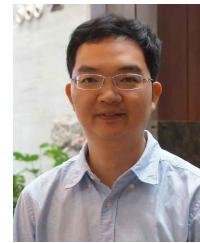
- [1] David P Bartel. Micrornas: genomics, biogenesis, mechanism, and function. *cell*, 116(2):281–297, 2004.
- [2] Lin He and Gregory J. Hannon. Micrornas: small rnas with a big role in gene regulation. *Nature Reviews Genetics*, 5:522—531, 2004.
- [3] Dan Li, Yulan Zhao, Changxing Liu, Xiaona Chen, Yanting Qi, Yue Jiang, Chao Zou, Xiaolong Zhang, Shunying Liu, Xuejing Wang, Dan Zhao, Qiang Sun, Zhenbing Zeng, Andreas Dress, Marie C. Lin, Hsiang-Fu Kung, Hallgeir Rui, Ling-Zhi Liu, Feng

- Mao, Bing-Hua Jiang, and Lihui Lai. Analysis of mir-195 and mir-497 expression, regulation and role in breast cancer. *Clinical Cancer Research*, 17(7):1722–1730, 2011.
- [4] Michael V.G. Latronico, Daniele Catalucci, and Gianluigi Condorelli. Emerging role of microRNAs in cardiovascular biology. *Circulation Research*, 101(12):1225–1236, 2007.
- [5] Ming Lu, Qipeng Zhang, Min Deng, Jing Miao, Yanhong Guo, Wei Gao, and Qinghua Cui. An analysis of human microRNA and disease associations. *PLOS ONE*, 3(10):1–5, 10 2008.
- [6] Xing Chen, Ming-Xi Liu, and Gui-Ying Yan. Rwrmda: predicting novel human microRNA–disease associations. *Molecular BioSystems*, 8(10):2792–2798, 2012.
- [7] Zhu-Hong You, Zhi-An Huang, Zexuan Zhu, Gui-Ying Yan, Zheng-Wei Li, Zhenkun Wen, and Xing Chen. Pbmda: A novel and effective path-based computational model for microRNA–disease association prediction. *PLOS Computational Biology*, 13(3):1–22, 03 2017.
- [8] Xing Chen, Le-Yi Wang, and Li Huang. Ndamda: Network distance analysis for microRNA–disease association prediction. *Journal of Cellular and Molecular Medicine*, 22(5):2884–2895, 2018.
- [9] Xing Chen and Gui-Ying Yan. Semi-supervised learning for potential human microRNA–disease associations inference. *Scientific reports*, 4(1):1–10, 2014.
- [10] Xing Chen, Qiao-Feng Wu, and Gui-Ying Yan. Rknnmda: Ranking-based knn for microRNA–disease association prediction. *RNA Biology*, 14(7):952–962, 2017. PMID: 28421868.
- [11] Qinghua Jiang, Guohua Wang, Shuilin Jin, Yu Li, and Yadong Wang. Predicting human microRNA–disease associations based on support vector machine. *International journal of data mining and bioinformatics*, 8(3):282–293, 2013.
- [12] Kiran K. Thekumparampil, Chong Wang, Sewoong Oh, and Li-Jia Li. Attention-based graph neural network for semi-supervised learning, 2018.
- [13] Christof Angermueller, Tanel Pärnmaa, Leopold Parts, and Oliver Stegle. Deep learning for computational biology. *Molecular Systems Biology*, 12(7):878, 2016.
- [14] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [15] Li Zhang, Xing Chen, and Jun Yin. Prediction of potential microRNA–disease associations through a novel unsupervised deep learning framework with variational autoencoder. *Cells*, 8(9), 2019.
- [16] Laiyi Fu and Qinke Peng. A deep ensemble model to predict microRNA–disease association. *Scientific Reports*, 7(1):14482, Nov 2017.
- [17] Cunmei Ji, Zhen Gao, Xu Ma, Qingwen Wu, Jiancheng Ni, and Chunhou Zheng. AEMDA: inferring miRNA–disease associations based on deep autoencoder. *Bioinformatics*, 37(1):66–72, 07 2020.
- [18] Xing Chen, Yao Gong, De-Hong Zhang, Zhu-Hong You, and Zheng-Wei Li. Drmda: deep representations-based microRNA–disease association prediction. *Journal of cellular and molecular medicine*, 22(1):472–485, 2018.
- [19] Li Zhang, Xing Chen, and Jun Yin. Prediction of potential microRNA–disease associations through a novel unsupervised deep learning framework with variational autoencoder. *Cells*, 8(9):1040, 2019.
- [20] Jiajie Peng, Weiwei Hui, Qianqian Li, Bolin Chen, Jianye Hao, Qinghua Jiang, Xuequn Shang, and Zhongyu Wei. A learning-based framework for microRNA–disease association identification using neural networks. *Bioinformatics*, 35(21):4364–4371, 2019.
- [21] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI Open*, 1:57–81, 2020.
- [22] Yanyi Chu, Xuhong Wang, Qiuying Dai, Yanjing Wang, Qiankun Wang, Shaoliang Peng, Xiaoyong Wei, Jingfei Qiu, Dennis Russell Salahub, Yi Xiong, and Dong-Qing Wei. MDA-GCNFTG: identifying miRNA–disease associations based on graph convolutional networks via graph sampling through the feature and topology graph. *Briefings in Bioinformatics*, 05 2021. bbab165.
- [23] Yulian Ding, Li-Ping Tian, Xiujuan Lei, Bo Liao, and Fang-Xiang Wu. Variational graph auto-encoders for microRNA–disease association prediction. *Methods*, 2020.
- [24] Cunmei Ji, Yu-Tian Wang, Zhen Gao, Lei Li, Jian-Cheng Ni, and Chun-Hou Zheng. A semi-supervised learning method for microRNA–disease association prediction based on variational autoencoder. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pages 1–1, 2021.
- [25] Zhengwei Li, Jiashu Li, Ru Nie, Zhu-Hong You, and Wenzheng Bao. A graph auto-encoder model for microRNA–disease associations prediction. *Briefings in Bioinformatics*, 22(4), 2021.
- [26] Jin Li, Sai Zhang, Tao Liu, Chenxi Ning, Zhuoxuan Zhang, and Wei Zhou. Neural inductive matrix completion with graph convolutional networks for microRNA–disease association prediction. *Bioinformatics*, 36(8):2538–2546, 2020.
- [27] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.
- [28] Jinli Zhang, Xiaohua Hu, Zongli Jiang, Bo Song, Wei Quan, and Zheng Chen. Predicting disease-related RNA associations based on graph convolutional attention network. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 177–182. IEEE, 2019.
- [29] Yahui Long, Min Wu, Yong Liu, Chee Keong Kwoh, Jiawei Luo, and Xiaoli Li. Ensembling graph attention networks for human microbe–drug association prediction. *Bioinformatics*, 36:i779–i786, 12 2020.
- [30] Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.
- [31] Zhou Huang, Jiangcheng Shi, Yuanxu Gao, Chunmei Cui, Shan Zhang, Jianwei Li, Yuan Zhou, and Qinghua Cui. Hmdd v3. 0: a database for experimentally supported human microRNA–disease associations. *Nucleic acids research*, 47(D1):D1013–D1017, 2019.
- [32] Dong Wang, Juan Wang, Ming Lu, Fei Song, and Qinghua Cui. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics*, 26(13):1644–1650, 2010.
- [33] Feng Huang, Xiang Yue, Zhankun Xiong, Zhouxin Yu, Shichao Liu, and Wen Zhang. Tensor decomposition with relational constraints for predicting multiple types of microRNA–disease associations. *Briefings in Bioinformatics*, 22(3):bbaa140, 2021.
- [34] Dan Busbridge, Dane Sherburn, Pietro Cavallo, and Nils Y. Hammerla. Relational graph attention networks, 2019.
- [35] Zhouxin Yu, Feng Huang, Xiaohan Zhao, Wenjie Xiao, and Wen Zhang. Predicting drug–disease associations through layer attention graph convolutional network. *Briefings in Bioinformatics*, 22(4):bbaa243, 2021.
- [36] Zhouxin Yu, Feng Huang, Xiaohan Zhao, Wenjie Xiao, and Wen Zhang. Predicting drug–disease associations through layer attention graph convolutional network. *Briefings in Bioinformatics*, 10 2020. bbba243.
- [37] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.
- [38] Guohao Li, Matthias Müller, Ali Thabet, and Bernard Ghanem. DeepGCNs: Can GCNs go as deep as CNNs? In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9267–9276, 2019.
- [39] Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi, and Stefanie Jegelka. Representation learning on graphs with jumping knowledge networks. In *International Conference on Machine Learning*, pages 5453–5462. PMLR, 2018.
- [40] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [41] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [42] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.
- [43] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [44] Zhengwei Li, Jiashu Li, Ru Nie, Zhu-Hong You, and Wenzheng Bao. A graph auto-encoder model for microRNA–disease associations prediction. *Briefings in Bioinformatics*, 2020.
- [45] Xinru Tang, Jiawei Luo, Cong Shen, and Zihan Lai. Multi-view multichannel attention graph convolutional network for microRNA–disease association prediction. *Briefings in Bioinformatics*, 2021.
- [46] Junlin Xu, Lijun Cai, Bo Liao, Wen Zhu, Peng Wang, Yajie Meng, Jidong Lang, Geng Tian, and Jialiang Yang. Identifying potential microRNAs–disease associations with probability matrix factorization. *Frontiers in genetics*, 10:1234, 2019.

- [47] Guobo Xie, Zhiliang Fan, Yiping Sun, Cuiming Wu, and Lei Ma. Wbnpmid: weighted bipartite network projection for microRNA-disease association prediction. *Journal of translational medicine*, 17(1):1–11, 2019.
- [48] Xing Chen, Chi-Chi Zhu, and Jun Yin. Ensemble of decision tree reveals potential miRNA-disease associations. *PLOS Computational Biology*, 15(7):1–24, 07 2019.
- [49] Sheng-Peng Yu, Cheng Liang, Qiu Xiao, Guang-Hui Li, Ping-Jian Ding, and Jia-Wei Luo. Glnmda: a novel method for miRNA-disease association prediction based on global linear neighborhoods. *RNA biology*, 15(9):1215–1227, 2018.
- [50] Sheng-Peng Yu, Cheng Liang, Qiu Xiao, Guang-Hui Li, Ping-Jian Ding, and Jia-Wei Luo. Melpmda: A novel method for miRNA-disease association prediction based on matrix completion and label propagation. *Journal of cellular and molecular medicine*, 23(2):1427–1438, 2019.
- [51] Zhao Li, Zhanlin Liu, Jiaming Huang, Geyu Tang, Yucong Duan, Zhiqiang Zhang, and Yifan Yang. Mv-gcn: multi-view graph convolutional networks for link prediction. *IEEE Access*, 7:176317–176328, 2019.
- [52] Diana Castro, Márcia Moreira, Alexandra Monteiro Gouveia, Daniel Humberto Pozza, and Ramon Andrade De Mello. MicroRNAs in lung cancer. *Oncotarget*, 8(46):81679, 2017.
- [53] Zhen Yang, Liangcai Wu, Anqiang Wang, Wei Tang, Yi Zhao, Haitao Zhao, and Andrew E Teschendorff. dbdemic 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic acids research*, 45(D1):D812–D818, 2017.
- [54] Raquel Díaz, Javier Silva, José M García, Yolanda Lorenzo, Vanesa García, Cristina Peña, Rufo Rodríguez, Concepción Muñoz, Fernando García, Félix Bonilla, et al. Deregulated expression of miR-106a predicts survival in human colon cancer patients. *Genes, Chromosomes and Cancer*, 47(9):794–802, 2008.
- [55] Hiroki Ogata-Kawata, Masashi Izumiya, Daisuke Kurioka, Yoshi-taka Honma, Yasuhide Yamada, Koh Furuta, Toshiaki Gunji, Hideki Ohta, Hiroyuki Okamoto, Hikaru Sonoda, et al. Circulating exosomal microRNAs as biomarkers of colon cancer. *PLoS one*, 9(4):e92921, 2014.
- [56] Xiaolan Zhu, Yuefeng Li, Huiling Shen, Hao Li, Lulu Long, Lulu Hui, and Wenlin Xu. mir-137 inhibits the proliferation of lung cancer cells by targeting cdc42 and cdk6. *FEBS letters*, 587(1):73–81, 2013.
- [57] Yong Dai, Weiguo Sui, Huijuan Lan, Qiang Yan, He Huang, and YuanShuai Huang. Comprehensive analysis of microRNA expression patterns in renal biopsies of lupus nephritis patients. *Rheumatology international*, 29(7):749–754, 2009.



Yuping Sun received the B.E. and Ph.D. degrees in computer science from the South China University of Technology in 2009 and 2015. From 2013 to 2015, he was a visiting Ph.D. student in the Department of Mathematics, National University of Singapore. He was a Post-Doctoral Research Fellow with the School of Automation Science and Engineering, South China University of Technology in 2015–2018. Since 2018, he has been with Guangdong University of Technology, where he became an associate professor in 2021. His research interests include pattern recognition, machine learning, bioinformatics and computer vision.



Guobo Xie received the M.E. and Ph.D. degrees from the Guangdong University of Technology, Guangzhou, in 2004 and 2012 respectively. Since 2004, he has been with the Guangdong University of Technology, where he became a Full Professor in 2014. His current research interests include bioinformatics, cloud computing and big data.



Zhiyi Lin received the M.S. degree in computer science and technology from the Wuhan University of Technology, Wuhan, China, in 2006, and the Ph.D. degree in computer software and theory from the State Key Laboratory of Software Engineering, Wuhan University, Wuhan, in 2009. Currently, he is a teacher at School of Computer Science, Guangdong University of Technology, Guangzhou, China. His research interests include evolutionary computation, pattern recognition, and bioinformatics.



Huijie Zhang is a graduate student in the School of Computer Science at Guangdong University of Technology. His research interests are bioinformatics and graph neural networks.



Juntao Fang is an undergraduate student in the School of Computer Science at Guangdong University of Technology. His research interests are bioinformatics and neural networks.



Guosheng Gu received the M.S. Degree in Applied Mathematics from South China Normal University, Guangzhou, China, in 2004 and the Ph.D. Degree in Computer Application Technology from South China University of Technology, Guangzhou, China, in 2007. Currently, he is a teacher at School of Computers, Guangdong University of Technology, Guangzhou, China. His research interests include multimedia information security and image processing.