IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 15, 2019, accepted May 14, 2019, date of publication May 20, 2019, date of current version July 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2917611

Predicting of Associations Between MicroRNA and Human Diseases Based on Multiple Similarities and Arbitrarily-Order Proximity Network Embedding

XIUJIN WU[1], WENHUA ZENG[1], YUXIU XU[1], BEIZHAN WANG[1], XIANGRONG LIU[2], FAN LIN[1,3], AND GIL ALTEROVITZ[3,4,5]

[1]Software School, Xiamen University, Xiamen 361005, China
[2]Department of Information Science and Engineering, Xiamen University, Xiamen 361005, China
[3]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA 02115, USA
[4]Harvard/MIT Division of Health Sciences and Technology, Harvard Medical School, Boston, MA 02115, USA
[5]Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Corresponding authors: Beizhan Wang (wangbz@xmu.edu.cn) and Xiangrong Liu (xrliu@xmu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61703196, and in part by the Natural Science Foundation of Fujian Province under Grant 2018J01549.

ABSTRACT Microbe-RNAs (miRNAs) play an important role and are associated with human diseases. However, considering the high cost and time-consuming biological experiments, using effective computational methods to discover the underlying association between the miRNAs and diseases would be valuable. This study presents a novel computational prediction model based on multiple-similarity and arbitrarily-order proximity network embedding. We obtain the Gaussian similarity from the disease–miRNA interaction matrix for the miRNA and disease. Then, considering the Gaussian similarity, disease semantic, phenotype similarity, and the miRNA functional similarity, we compute the miRNA–miRNA similarity matrix and the disease–disease similarity matrix. Most importantly, we improved the SVD matrix decomposition to extract the primary feature vector. We called it arbitrarily-order proximity network embedding method. By multiplying the feature vectors together, we calculate the final miRNA–disease association score matrix. According to the ranking scores, we can know which miRNA is mostly relevant to a disease. This process proved that our method achieved better prediction performance than other methods. In the experiment, after adding arbitrarily-order proximity network embedding to the inductive matrix completion method [1], the AUC of our method in leave-one-out cross-validation increased dramatically from 0.8034 to 0.92306. Meanwhile, the studies of three cases, namely, prostate neoplasms, breast neoplasms, and lung neoplasms, of the top 50 potential miRNAs predicted by our method were validated by the database of dbDEMC and mir2disease. This finding indicated that our method can effectively obtain the potential disease miRNA candidates. Comparison of our work with other algorithms reveals its reliable performance.

INDEX TERMS Bioinformatics, biomedical informatics, biological interactions, prediction methods.

I. INTRODUCTION

Microbe RNAs (miRNAs) in the human body have a substantial effect and are associated with various and miscellaneous human diseases [2]. MiRNAs are a set of single-stranded and short noncoding RNAs, including approximately 22–25 nucleotides [3]–[6]. When binding to the 3'-untranslated regions of the target mRNAs, miRNA affects the post-transcriptional and regulation level of gene expression [7]. miRNAs are also involved in many important biological processes, such as cell cycle control [8], growth [9], differentiation [10], development [11], aging, apoptosis [12], infection, and viral infection [13]. Humans have detected tens of thousands of various organism miRNAs

The associate editor coordinating the review of this manuscript and approving it for publication was Yonghong Peng.

that have reached to 28,645 (2588 for human) in the latest release of miRBase [JCMM]. Consequently [14]–[17], dysregulation and mutation of miRNAs can cause various complex human diseases, including neoplasms and cancer [18], because miRNAs have specific secondary structures and conserved sequences. The first miRNA lin-4 was found in the early 1990s by Lee et al. [8]. Heegaard et al. developed a method to measure the circulating levels of 30 miRNAs and found that the expression levels of miR-146b, miR-221, let-7a, miR-155, miR-17-5p, miR-27a, and miR-106a were greatly reduced in the serum of nonsmall cell lung neoplasm cases [19]. In addition, miRNA-17–92 cluster was found to be upregulated in polycystic kidney disease (PKD) and can be identified as a therapeutic target in PKD [20]. Increasing miRNAs have been discovered and were important to human disease. However, the biological experimental methods used to discover the association between miRNA and disease are expensive and time consuming. Therefore, numerous researchers are paying attention to the computational methods employed to predict potential miRNA–disease association to guide the biological experiments [21]–[29].

In the past few years, several computational models have been exploited to predict potential miRNA–disease association [30], [31]. Jiang et al. [32] have presented a hypergeometric distribution method by building a heterogeneous network based on the notion that functionally similar miRNAs tend to be associated with phenotypically similar diseases and vice versa [33]. However, this model only uses the information of the direct network neighbors of miRNAs, ignoring those indirectly linked to miRNAs. In addition, Xuan et al. [34] have developed a method named HDMP by considering the weighted k most similar neighbors of miRNAs, wherein the members in the same miRNA family or cluster were assigned with high weight. However, this model cannot be applied to new diseases without any known related miRNAs because it needs neighbors of miRNAs and its prediction accuracy is limited. Consequently, this model depends on the algorithm adopting local similarity measure [35], [36]. RWRMDA is the first global network-based method, which uses random walk method to infer miRNA–disease associations [25]. Subsequently, Chen et al. proposed another method called WBSMDA [27], which calculates a final score for potential miRNA–disease associations by integrating miRNA functional similarity, disease semantic similarity, known miRNA–disease associations, and Gaussian interaction profile kernel similarity of miRNAs and diseases. In addition, Xuan et al. [37] devised another computational model based on random walk on miRNA functional similarity network. They exploited the miRNA similarity, the disease similarity, the known miRNA–disease associations, the topology information of the bilayer network, and the information from different layers of network to predict disease miRNA candidates. In particular, this method is adoptable to predict potential miRNAs for diseases without known related miRNAs. Yu Qu et al. have proposed an approach named KATZLDA [38] to calculate the miRNA similarity and disease similarity, and then it integrated multiple data sources to construct a reliable heterogeneous network to predict miRNA–disease associations. Although the existing methods have made remarkable contributions, improvements are still needed [39].

All mentioned methods have their own strengths, and the existing methods can be categorized into five aspects: (i) neighborhood-based methods, such as HDMP [34] and CPTL [40]; (ii) random walk-based methods, such as RWRMDA [25], Shi's method [37], MIDP, and MIDPE [37]; (iii) machine learning-based methods, such as Xu's method [41] and RLSMDA [26]; (iv) path-based methods, such as KATZ [42] and PBMDA [43]; and (v) matrix completion, such as MCMDA [23] and IMCMDA [1].

Inspired by matrix completion-based and multiple-similarity approaches, we propose a model based on multiple-similarities and arbitrarily-order proximity network embedding (MSAOPNE) to predict miRNA–disease associations. First, we obtain the Gaussian similarity from the disease–miRNA interaction matrix for miRNA and disease. Second, we integrate the disease Gaussian similarity and semantic similarity together. Similarly, we integrate the miRNA Gaussian similarity and functional similarity together. Fourth, we extract the primary feature vectors by using arbitrarily-order proximity network embedding. Finally, we can calculate the final miRNA–disease association score matrix. According to the ranking scores, we can know which miRNA is mostly relevant to a disease.

In the experiment, we used two evaluation methods, namely, leave-one-out cross-validation (LOOCV) and five-fold cross-validation (five times CV), to verify the performance of our method. Our approach has achieved outstanding results in identifying potential miRNA–disease associations compared with existing methods. In the experiment, after adding arbitrarily-order proximity network embedding to the inductive matrix completion method [1], the area under curves of our method in global and local leave-one-out cross-validation increased by 11% achieves AUC(area under curve) of 0.91956 and 0.92306. For further verification, we used case studies to analyze the MSAOPNE performance. Experimental results show that the method has reliable performance in detecting new associations. We also found that some specific associations and corresponding miRNAs require further attention.

II. MATERIALS

This chapter will introduce the materials we use, consisting of four parts, namely, the initial miRNA–disease association network, Gaussian interaction profile kernel similarity, disease semantic and phenotype similarity, and miRNA function similarity.

A. HUMAN MIRNA-DISEASE ASSOCIATION NETWORK

This study used the known human miRNA–disease association data downloaded from HMDD V2.0 database [44] containing 5430 experimentally human miRNA–disease

associations between 383 diseases and 495 miRNAs. A disease–miRNA interaction matrix $A \in R^{nd \times nm}$ from known disease–miRNA associations was used, where nd and nm are the number of diseases and miRNA, respectively, and each row corresponds to a disease and each column represents a miRNA. If a disease d_i has an association with a miRNA m_j, then A_{ij} equals to 1, otherwise 0. We can define A matrix as follows:

$$\begin{cases} A_{ij} = 1 \\ A_{ij} = 0 \end{cases} \tag{1}$$

B. GAUSSIAN INTERACTION PROFILE KERNEL SIMILARITY

The Gaussian kernel, also known as the radial basis function, is a commonly used kernel function. Based on the assumption that functionally similar miRNAs show the same behavior interaction with similar diseases, we use $KD \in R^{383 \times 383}$ to define the Gaussian interaction kernel similarity matrix for diseases, and $KM \in R^{495 \times 495}$ is used to define the Gaussian interaction kernel similarity matrix for miRNAs. The interaction profile of miRNA i is the ith row vector of interaction association matrix $A \in R^{nd \times nm}$. We denoted vector $IP(d_i)$ and $IP(miRNA_i)$ to represent the i^{th} row vector and the j^{th} column vector. Then, the distance between any two row vectors is computed as the Gaussian interaction profile kernel of their corresponding diseases, similar to miRNAs. We can calculate them separately as follows:

$$KD = Gkl\left(d_i, d_j\right) = \exp\left(-\gamma_d \left\|IP(d_i) - IP(d_j)\right\|^2\right)$$

$$KM = Gkl\left(miRNA_i, miRNA_j\right) \tag{2}$$

$$= \exp\left(-\gamma_m \left\|IP(miRNA_i) - IP(miRNA_j)\right\|^2\right) \tag{3}$$

The adjustment coefficient kernel bandwidth γ_m and γ_d is computed as follows:

$$\gamma_m = \frac{\gamma_m^{'}}{\left[\frac{1}{nm}\sum_{i=1}^{nm}\left\|miRNA_i - miRNA_j\right\|^2\right]} \tag{4}$$

$$\gamma_d = \frac{\gamma_d^{'}}{\left[\frac{1}{nd}\sum_{i=1}^{nd}\left\|d_i - d_j\right\|^2\right]} \tag{5}$$

where $\gamma_m{}'$ and γ_d' are the original kernel bandwidth.

C. DISEASE SEMANTIC AND PHENOTYPE SIMILARITY MODEL

According to several computing models [27], [38], [45]–[47], we can construct all the diseases to a Directed Acyclic Graph (DAG). We can download these diseases based on the Medical Subject Headings descriptors from the National Library of Medicine (http://www.nlm.nih.gov/). We use the following computing models to finally obtain a weighted disease similarity network containing 146,689 similar associations among 383 diseases.

1) MODEL 1

The contribution values of disease d in DAG(D) [45] to the semantic value of disease D can be measured as

$$\begin{cases} D1_D(d) = 1 \text{ if } d = D \\ D1_D(d) = \max\left\{\varepsilon * D\left(d'\right) | d' \in \text{ children of } d\right\} \text{ if } d \neq D \end{cases} \tag{6}$$

where ε is the impact factor that if the children of D is far away, the impact factor is smaller than the nearest one. The semantic value of disease D is added in all the diseases in the DAG together, denoted as DV. Then, DV1 (D) $= \sum_{d \in T(D)} D2_D(d)$. The semantic similarity score between disease d_i and d_j can be defined as

$$SS1\left(d_i, d_j\right) = \frac{\sum_{t \in T(d_i) \cap T(d_j)}\left(D1_{d_i}(t) + D1_{d_j}(t)\right)}{DV_1\left(d_i\right) + DV_1(d_j)} \tag{7}$$

2) MODEL 2

Considering that if a specific disease was less, DAGs should contribute a high value to the semantic similarity of disease D. According to the model proposed by Xuan et al. [34], the semantic value of disease D can be measured

$$D2_D\,(\mathrm{d}) = -\log\left[\frac{the\ number\ of\ DAGs\ including\ t}{the\ number\ of\ diseases}\right] \tag{8}$$

Similar to model 1, the computed disease phenotype semantic similarity can be calculated as follows:

$$SS2\left(d_i, d_j\right) = \frac{\sum_{t \in T(d_i) \cap T(d_j)}\left(D2_{d_i}(t) + D2_{d_j}(t)\right)}{DV_2\left(d_i\right) + DV_2(d_j)} \tag{9}$$

3) INTEGRATING TWO MODELS TOGETHER

Consequently, we can combine these two models together and obtain the final disease semantic similarity value as

$$SS = \frac{SS1 + SS2}{2} \tag{10}$$

Obviously, if the value is large, two diseases are likely to be similar with each other. In addition, the similarity of two diseases is closely related to their semantic similarity and the phenotype similarity.

4) MIRNA FUNCTIONAL SIMILARITY MODEL

The members of miRNA family or cluster that have functional similarity are probably associated with similar diseases and vice versa. Wang et al. (2010) proposed a method to calculate the miRNA functional similarity. We can download the miRNA functional similarity data from http://www.cuilab.cn/files/images/cuilab/misim.zip. We denoted the matrix FS to represent the miRNA functional similarity. The element $FS(m_i, m_j)$ represents the similarity value between the miRNA m_i and the miRNA m_j.

III. METHODS

In this section we will introduce our whole method. First of all, we will introduce the method overview. Then you can see how to integrate all multiple similarity data together and

FIGURE 1. Main Structure of The MSAOPNE Method.

why we use the arbitrary-order proximity network embedding to extract the feature. Then finally introduce how to get the matrix scores.

A. METHOD OVERVIEW

As shown in Fig. 1, MSAOPNE consists of five steps. In Step 1, we use the known associations between the disease and miRNA to complete the initial interaction matrix. Then, MSAOPNE computes the Gaussian interaction profile kernel similarity for disease and miRNA from the known miRNA–disease interaction matrix. We inferred the miRNA–miRNA similarity matrix KM and disease–disease similarity matrix KD. In Step 2, considering the disease phenotype and semantic similarity matrix SS, we combine it with KD, which has been broken down by the Gaussian interaction profile kernel matrix that contains the known miRNA–disease interaction information as mentioned in the previous step. By multiplying it with the weight SSP, we added them together. The details are illustrated in Section 3.2. In Step 3, similar to the second step, we can multiply the weight FSP with the miRNA function similarity matrix FS and then add it with the Gaussian interaction profile kernel similarity matrix KM containing the known miRNA–disease interaction information. In Step 4, considering the contribution of the arbitrarily-order proximity neighbors, we use the arbitrarily-order proximity network feature embedding algorithm to extract the feature for miRNA and disease. Then, we can obtain the disease feature embedding matrix SDD and the miRNA similarity feature embedding matrix SMM. Finally, in Step 5, by multiplying the disease feature embedding vector SDD and the miRNA similarity feature embedding vector SMM, we obtain the final disease–miRNA similarity associate score matrix. According to the ranking scores, we can know which miRNA is mostly relevant to a disease.

B. INTEGRATION SIMILARITY FOR DISEASE AND MIRNA

To effectively compute the disease similarity, we have integrated the Gaussian similarity and the disease semantic and phenotype similarity together. We use the following formula to incorporate the information content of Gaussian kernel similarity matrix KD and disease semantic and phenotype similarity matrix SS, wherein SSP is the integrated weight.

$$\mathrm{SD} = \mathrm{SS} \odot \mathrm{SSP} + \mathrm{KD} \odot (1 - \mathrm{SSP}). \tag{11}$$

At the same time, we integrate the Gaussian kernel similarity matrix KM and the miRNA functional similarity matrix FS together, by the following formula, wherein FSP is the integrated weight.

$$\mathrm{SM} = \mathrm{FS} \odot \mathrm{FSP} + \mathrm{KM} \odot (1 - \mathrm{FSP}). \tag{12}$$

C. ARBITRARY-ORDER PROXIMITY NETWORK EMBEDDING

1) ARBITRARY-ORDER PROXIMITY DEFINITION

Probably, we have integrated all the similarities, including the Gaussian similarity, the disease semantic and phenotype similarity, and miRNA function similarity together. Then, both the disease–disease similarity matrix and the miRNA–miRNA similarity matrix are symmetric matrixes. Here, A indicates a symmetric adjacency matrix, where n is the order and $w_1 \ldots\ldots w_n$ are the weights. A^n refers to the nth order of the symmetric adjacency matrix A, which is a high-order proximity matrix. Then, we define the arbitrary-order proximity matrix function $\varphi(A)$ as

$$\mathrm{S} = \varphi(A) = w_1 A + w_2 A^2 + \cdots\cdots + w_n A^n \tag{13}$$

where $w_i = \beta^i$, $\beta \in (0, 1)$, and $\beta^i$is convergent. Thus, we can limit the weights in 0–1. Moreover, the weight is continuously decreasing, indicating that distant node means small the weight, guaranteeing that the function is convergent. Notably, we refer a proximity of order n as the weighted combination of all the orders from the 1st to the nth rather than the nth order alone. The matrix S is called the arbitrary-order proximity matrix.

2) EIGEN-DECOMPOSITION REWEIGHTING

Matrix A is a symmetric matrix. Matrix A^2 and matrix A^n are also symmetrical, similar to matrix S. According to the definition of an eigenpair, if matrix A has an eigenvalue λ and feature vector x, we can obtain the eigenpair of A^2, $[\lambda^2, x]$. Then, $[\varphi(\lambda), x]$ is an eigenpair of $\mathrm{S} = \varphi(\mathrm{A})$. According to the following definition of an eigenpair:

$$\mathrm{A}x = \lambda x \tag{14}$$

Then, we can easily obtain the following formula:

$$A^2 x = \mathrm{A}\lambda x = \lambda A x = \lambda^2 x \tag{15}$$

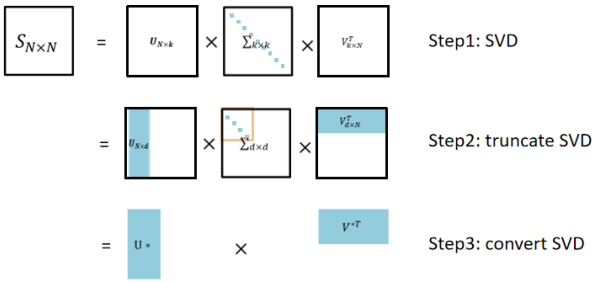

FIGURE 2. SVD Dimension Reduction.

By repeating the above process, we have:

$$\begin{aligned}\varphi(\lambda) &= (w_1\lambda + w_2\lambda^2 + \cdots + w_n\lambda^n) \\ &= \beta\lambda + \lambda^2\lambda^2 + \cdots + \lambda^n\lambda^n) = \frac{\beta\lambda}{1-\beta\lambda} \end{aligned} \quad (16)$$

3) SVD DIMENSION REDUCTION

How to get the feature embedding for the arbitrary-order proximity matrix S? The size of the disease–disease similarity matrix SD and the miRNA–miRNA similarity matrix SM is 383×383 and 495×495, respectively. Supposing that the size of the matrix is very large, it needs to be compressed and converted to two low-rank feature matrix U$*$ multiplied by V$*$. In addition, we can use SVD dimension reduction, as shown in Fig. 2.

From the SVD dimension reduction figure, we can see the following process:

$$\mathrm{S} = \mathrm{U}\sum V^T = (\mathrm{U}\sqrt{\Sigma})\left(\sqrt{\Sigma}V^T\right) = U^*V^{*T} \quad (17)$$

where $U^* = \mathrm{U}\sqrt{\Sigma}$, $V^{*T} = \sqrt{\sum}V^T$. The eigenvalue must be positive, forming the covariance matrix $\sum$. Thus, we will obtain the absolute value of each eigenvalue.

According to the Eckart–Young theorem: $\mathrm{S} = \sum \sigma_\mathrm{i}\mathrm{u}_\mathrm{i}\mathrm{v}_\mathrm{i}^\mathrm{T}$. If we select the top d eigenvalues to form the new covariance matrix $\sum$, the following matrix S^* must be closest to S.

$$S^* = \sum_{\mathrm{i}=1}^{\mathrm{d}} \sigma_\mathrm{i}\mathrm{u}_\mathrm{i}\mathrm{v}_\mathrm{i}^\mathrm{T}, (\sigma_\mathrm{d} > \sigma_{\mathrm{d}+1}) \quad (18)$$

Then, we have to minimize the following objective function:

$$\min_{U^*,\, V^*} \left\| S - U^*V^{*T} \right\|_F^2 \quad (19)$$

To prevent U^* and V^{*T} items from being too large, we added the regularization terms. Then, we can form them as follows:

$$\min_{U^*,\, V^*} \varphi = \left\| S - U^*V^{*T} \right\|_F^2 + \frac{\lambda_1}{2}\left\| U^* \right\|_F^2 + \frac{\lambda_2}{2}\left\| V^{*T} \right\|_F^2 \quad (20)$$

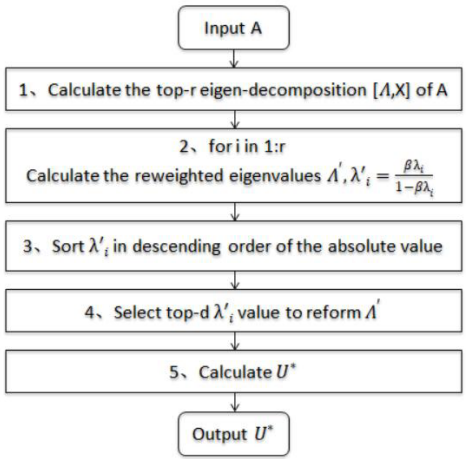

FIGURE 3. Main Process of The AOPNE Algorithm.

4) FEATURE EXTRACTING AND NETWORK EMBEDDING

The main process of AOPNE algorithm is shown in Figure 3. First, we will perform SVD eigen decomposition on matrix A. Then, we will rank the eigenvalues and select the top-r eigenvalues in decreasing order. Second, for the top-r eigenvalues of A matrix, we transform them with the following equation. Then, λ'_i is the eigenvalues of the arbitrary-order proximity matrix S. Third, we sort the eigenvalues λ'_i in descending order and then take the absolute value because SVD dimension reduction requires that covariance matrix $\sum$ must be positive. Fourth, we select the top-d eigenvalues λ'_i to reform the covariance matrix of S. Finally, we will calculate the main feature matrix U^* for the arbitrary order matrix S.

We will explain why we add an absolute value in step 3 and not change our results. We prove the theorem by showing that any eigenvalue except the top-r cannot be larger in absolute value than the d positive eigenvalues after $\varphi(\lambda_j)$. Assuming $|\lambda_i| \geq |\lambda_j|$ and $\lambda_i > 0$, we have

$$\begin{aligned}|\varphi(\lambda_i)| &= \left|\mathrm{w}_1\lambda_i + \ldots + \mathrm{w}_n\lambda_i^n\right| = \mathrm{w}_1|\lambda_i| + \ldots + \mathrm{w}_n|\lambda_i|^n \\ &\geq \mathrm{w}_1|\lambda_j| + \ldots + \mathrm{w}_n|\lambda_j|^n \geq \left|\mathrm{w}_1\lambda_j \right. \\ &\quad \left. + \ldots + \mathrm{w}_n\lambda_j{}^n\right| = \left|\varphi(\lambda_j)\right| \end{aligned} \quad (21)$$

If we enter the similarity matrix of the symmetric matrix, we will obtain its arbitrary order feature matrix. Using disease–disease similarity matrix SD instead of A, we will achieve the arbitrary order similarity matrix for SD to achieve the purpose of compressing the matrix and extracting the feature, similar to the miRNA–miRNA similarity matrix SM. Then, we will finally obtain the feature extract and network embedding matrix of SDD and SMM. Here, SDD was used to denote the feature embedding vector of disease–disease, and SMM was used to signify the feature embedding vector of miRNA–miRNA.

D. FINAL SCORE MATRIX

We will use the arbitrary-order proximity network embedding matrix SDD and SMM to calculate the prediction score

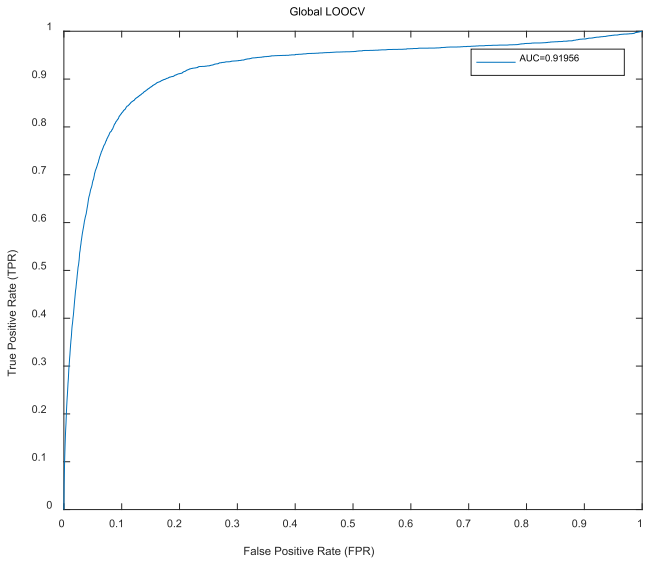

FIGURE 4. Global LOOCV.

between disease and miRNA by the following equation:

$$\text{Score Matrix} = \text{SDD} \odot \text{SMM} \tag{22}$$

The element Score $d_i * m_j$ is calculated to denote the predicted association possibility between disease d_i and miRNA m_j.

IV. EXPERIMENT

In the follow section, we will introduce the evaluation metric we used, and the good performance of our method, then three case studies of our method.

A. EVALUATION METRIC

The important criteria for evaluating the model are the receiver operating characteristic (ROC) curve and the area under curve (AUC) indicator. The ROC curve can reflect the classification effect of the classifier to a certain extent but is not intuitive enough. AUC intuitively reflects the classification ability of the ROC curve expression, which is defined as the area enclosed by the ROC curve and the coordinate axis. High AUC value means classification effect. AUC equals to 1 indicates that the model has perfect prediction performance. AUC equals to 0.5 indicates that the model only has random accuracy. The x-coordinate of ROC is a false positive rate (FPR, 1-specificity), and the y-coordinate is true positive rate (TPR, sensitivity). FPR is simply the possibility of predicting a positive sample, which was misclassified. TPR refers to the percentage of the positive test samples with higher ranks than the specific threshold. The computational formulae of FPR and TPR are as follows:

$$\text{specificity} : FPR = \frac{FP}{TN + FP} \tag{23}$$

$$\text{sensitivity} : TPR = \frac{TP}{TP + FN} \tag{24}$$

where TP denotes the number of true samples with scores higher than the specific threshold. Meanwhile, FN denotes the number of true samples with scores lower than the specific threshold. TN denotes the number of false samples with scores lower than the specific threshold. FP denotes the number of false samples with scores higher than the specific threshold.

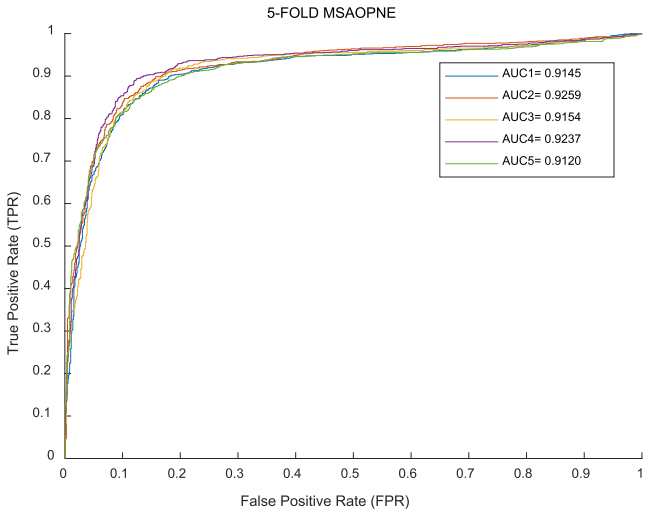

FIGURE 5. 5-FOLD MSAOPNE Performance.

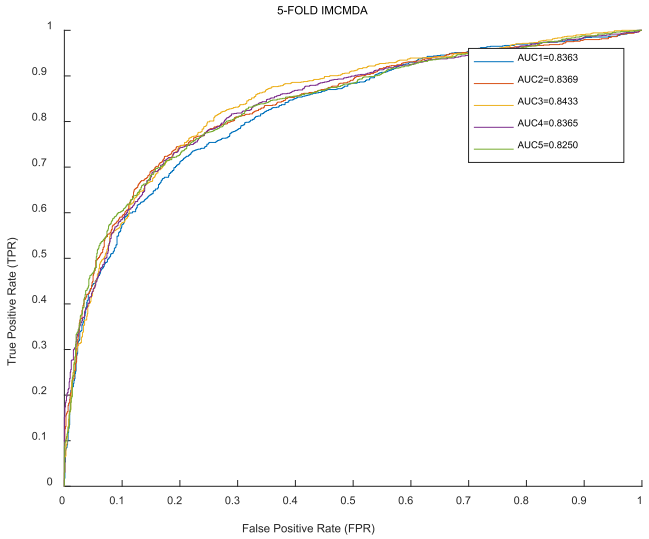

FIGURE 6. 5-FOLD IMCMDA Performance.

Another important criterion used for evaluating a model is whether it can predict potential associate miRNAs for new diseases. Some articles use this metric to evaluate. The motivation for this performance is to sort the probability of true associations in the prediction for a new disease. Moreover, we selected the TOP-N miRNAs with the highest probability of association, indicating that these miRNAs are most likely to have a relationship with the distinct disease. In the case studies, we use the TOP-N analysis method and select the largest N scores from the sorted score list. This method is currently widely used in disease–miRNA association prediction and some other recommendation systems.

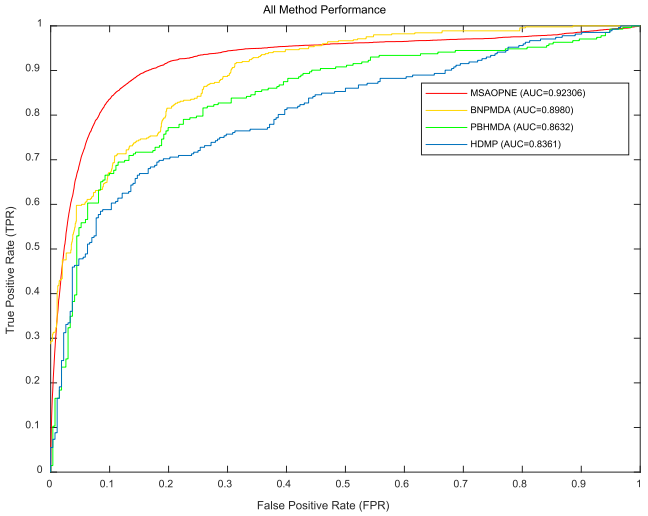

FIGURE 7. All Method Performance.

TABLE 1. Top 50 mirnas associated with prostated neoplasms predicted TY MSAOPNE.

miRNA	evidence	miRNA	evidence
hsa-mir-521	HMDD	hsa-mir-135b	HMDD
hsa-mir-1256	dbDEMC HMDD	hsa-mir-223	HMDD
hsa-mir-1296	HMDD	hsa-mir-32	dbDEMC HMDD
hsa-mir-616	dbDEMC HMDD	hsa-mir-330	HMDD
hsa-mir-647	dbDEMC HMDD	hsa-mir-96	HMDD
hsa-mir-193a	HMDD	hsa-mir-378a	HMDD
hsa-mir-642b	HMDD	hsa-mir-708	HMDD
hsa-mir-191	dbDEMC HMDD	hsa-mir-93	HMDD
hsa-mir-378b	HMDD	hsa-mir-320a	dbDEMC HMDD
hsa-mir-378c	dbDEMC HMDD	hsa-mir-642a	HMDD
hsa-mir-378d	dbDEMC HMDD	hsa-mir-151a	HMDD
hsa-mir-378e	HMDD	hsa-mir-452	dbDEMC HMDD
hsa-mir-378f	HMDD	hsa-mir-373	dbDEMC HMDD
hsa-mir-378g	HMDD	hsa-mir-133b	dbDEMC HMDD
hsa-mir-378h	HMDD	hsa-mir-488	dbDEMC HMDD
hsa-mir-378i	HMDD	hsa-mir-101	dbDEMC HMDD
hsa-mir-99b	dbDEMC HMDD	hsa-mir-146a	HMDD
hsa-mir-151b	dbDEMC HMDD	hsa-mir-449a	dbDEMC HMDD
hsa-mir-127	dbDEMC HMDD	hsa-mir-30c	dbDEMC HMDD
hsa-mir-182	dbDEMC HMDD	hsa-mir-15b	HMDD
hsa-mir-519d	HMDD	hsa-mir-185	dbDEMC HMDD
hsa-mir-106a	HMDD	hsa-mir-203	HMDD
hsa-mir-23b	dbDEMC HMDD	hsa-mir-301b	HMDD
hsa-mir-194	HMDD	hsa-mir-132	HMDD
hsa-mir-520c	HMDD	hsa-mir-146b	dbDEMC HMDD

The top 1-25 related miRNAs are shown in the first column, whereas the top 26-50 in the third column.

B. PERFORMANCE

We implement three types of cross validation, namely, Global LOOCV (Figure 4), and 5-FOLD cross-validation (Figure 5), to show the perfect performance of our method. In Global LOOCV, each known miRNA–disease association was left out in turn to be taken as test sample, and the other remaining known associations were treated as training samples. The AUC value of our approach has achieved 0.91956.

TABLE 2. Top 50 miRNAs associated with breast neoplasms predicted TY MSAOPNE.

miRNA	evidence	miRNA	evidence
hsa-mir-142	HMDD	hsa-mir-484	HMDD
hsa-mir-130b	dbDEMC HMDD	hsa-mir-28	dbDEMC
hsa-mir-150	dbDEMC HMDD	hsa-mir-32	dbDEMC HMDD
hsa-mir-92b	dbDEMC HMDD	hsa-mir-181d	dbDEMC
hsa-mir-15b	dbDEMC HMDD	hsa-mir-372	dbDEMC HMDD
hsa-mir-98	HMDD	hsa-mir-433	dbDEMC
hsa-mir-192	dbDEMC HMDD	hsa-mir-503	HMDD
hsa-mir-30e	HMDD	hsa-mir-337	dbDEMC
hsa-mir-130a	dbDEMC HMDD	hsa-mir-144	dbDEMC HMDD
hsa-mir-491	HMDD	hsa-mir-185	dbDEMC HMDD
hsa-mir-138	dbDEMC HMDD	hsa-mir-1271	dbDEMC
hsa-mir-99b	Other paper	hsa-mir-509	HMDD
hsa-mir-198	dbDEMC HMDD	hsa-mir-615	Other paper
hsa-mir-99a	dbDEMC HMDD	hsa-mir-376a	HMDD
hsa-mir-196b	dbDEMC HMDD	hsa-mir-432	dbDEMC
hsa-mir-106a	dbDEMC HMDD	hsa-mir-449b	dbDEMC
hsa-mir-134	dbDEMC HMDD	hsa-mir-208b	dbDEMC
hsa-mir-302e	dbDEMC	hsa-mir-485	HMDD
hsa-mir-302f	dbDEMC HMDD	hsa-mir-95	dbDEMC
hsa-mir-744	dbDEMC	hsa-mir-508	Other paper
hsa-mir-378a	HMDD	hsa-mir-655	dbDEMC
hsa-mir-494	HMDD	hsa-mir-518c	dbDEMC
hsa-mir-212	dbDEMC HMDD	hsa-mir-330	dbDEMC
hsa-mir-381	HMDD	hsa-mir-637	Other paper
hsa-mir-363	HMDD	hsa-mir-650	dbDEMC

The top 1-25 related miRNAs are shown in the first column, whereas the top 26-50 in the third column.

Different from Global LOOCV, the Local LOOCV only considered the ranking of the score generated by the test association among the candidate associations, which were merely related to the investigated disease. Then, our AUC value is also as high as 0.92306. To further prove our experimental results, we also made a 5-FOLD cross-validation. The 5-FOLD cross-validation, as the name implies, divides the data set into five parts, taking four of them as the training data and one part as the test data for experimentation. The result of 5-FOLD cross-validation AUC value and the accuracy have reached 0.9232 ∓ 0.0024 (Figure 5) and 0.9253 ∓ 0.002. Compared with the IMCMDA methods, if we ignore the arbitrarily-order proximity neighbor contribution in the network, we only can receive the mean AUC value and the mean accuracy of 0.8356 ∓ 0.0106 (Figure 5) and 0.8117 ∓ 0.002, respectively, as shown in the following figures. This finding also validates our hypothesis and reveals the importance of considering the arbitrarily-order proximity neighbor contribution.

We have compared our method with BNPMDA, PBHMDA, and HDMP based on the LOOCV framework. The known miRNA–disease association dataset used for this comparison was the same, i.e., the 5430 known associations between 495 miRNAs and 383 diseases in the HMDD V2.0 database. As for other input datasets required by these six methods, we either downloaded the corresponding data

TABLE 3. Top 50 miRNAs associated with lung neoplasms predicted TY MSAOPNE.

miRNA	evidence	miRNA	evidence
hsa-mir-708	dbDEMC	hsa-mir-34b	dbDEMC HMDD
hsa-mir-125b	HMDD	hsa-let-7b	dbDEMC HMDD
hsa-mir-1	HMDD	hsa-let-7c	HMDD
hsa-mir-155	HMDD	hsa-mir-17	dbDEMC HMDD
hsa-mir-21	HMDD	hsa-mir-106b	Other paper
hsa-mir-143	HMDD	hsa-mir-9	HMDD
hsa-mir-378a	HMDD	hsa-mir-199b	HMDD
hsa-mir-146a	dbDEMC	hsa-mir-142	HMDD
hsa-mir-145	HMDD	hsa-mir-15a	dbDEMC HMDD
hsa-mir-16	HMDD	hsa-mir-200c	dbDEMC HMDD
hsa-mir-146b	HMDD	hsa-mir-34a	HMDD
hsa-mir-222	HMDD	hsa-mir-196a	HMDD
hsa-mir-205	HMDD	hsa-mir-15b	Other paper
hsa-let-7d	HMDD	hsa-mir-133b	dbDEMC HMDD
hsa-let-7a	HMDD	hsa-mir-29c	dbDEMC HMDD
hsa-mir-138	HMDD	hsa-mir-132	dbDEMC HMDD
hsa-mir-125a	dbDEMC HMDD	hsa-mir-223	HMDD
hsa-mir-499a	HMDD	hsa-mir-151b	dbDEMC
hsa-mir-34c	dbDEMC HMDD	hsa-mir-133a	dbDEMC HMDD
hsa-mir-126	HMDD	hsa-mir-7	dbDEMC HMDD
hsa-mir-210	HMDD	hsa-let-7f	HMDD
hsa-mir-221	dbDEMC HMDD	hsa-mir-135a	dbDEMC HMDD
hsa-mir-193a	HMDD	hsa-mir-574	HMDD
hsa-mir-96	HMDD	hsa-mir-20b	dbDEMC
hsa-mir-451a	HMDD	hsa-mir-520b	dbDEMC

The top 1-25 related miRNAs are shown in the first column, whereas the top 26-50 in the third column.

from the supplementary files in the methods' literatures or collected the data from the sources specified in the literatures.

C. CASE STUDIES

To further verify the prediction accuracy of MSAOPNE, we performed case studies for three complex popular diseases of human beings, namely, prostate neoplasms, breast neoplasms, and lung neoplasms. We observed the number of miRNAs verified at the three diseases in the top 10, top 20, and even top 50.

Prostate neoplasm is one of the greatest threats to men's health worldwide. The miRNA expression levels can help treat patients suffering from prostate neoplasms. For example miR-488 inhibits androgen receptor expression in prostate carcinoma cells. These miRNAs were detected by MSAOPNE and are shown in the table below. In addition, we can see that the top 50 miRNAs in the correlation detection were all verified in the HMDD and dbDEMC database (Table 1).

Breast neoplasm is a common malignant tumor in women. Understanding the association between miRNA and breast neoplasm will help us detect, diagnose, and treat early breast neoplasms. MiRNA is widely involved in key indicators, such as breast neoplasm cell proliferation [48], invasion, and lymph node metastasis, because it plays an important role in the occurrence and development of breast neoplasms. For example, in some tumors, miR-142 regulates the properties of BCSCs at least in part by activating the WNT signaling pathway and miR-150 expression. These miRNAs were detected by MSAOPNE, and 46 out of the top 50 predicted breast neoplasm-related miRNAs were confirmed by HMDD and dbDEMC. Although 302e was not confirmed in the HMDD database, other members of the 302 series (302a, 302b, 302c, 302d, and 302f) were confirmed to be associated with breast neoplasms in the HMDD database. Probably, we can make a biological test about whether they are related based on the results of this score.

Lung neoplasm is one of the largest threats to men's health worldwide. The mortality rate of lung neoplasms is extremely high, and approximately 1.3 million people die every year due to the lung diseases worldwide. Neoplasms account for approximately one-third of all neoplasm deaths in the United States [49]. The detection of miRNA markers is of great significance for the early diagnosis of lung neoplasms [50]. Studying the characteristic miRNA expression profiles in tumor tissues may become an important means for early diagnosis, targeted therapy, and prognosis evaluation of tumors. The miRNAs detected by MSAOPNE and confirmed in the HMDD database are all illustrated in the following table. As a result, 48 of top 50 predicted miRNAs were confirmed by HMDD and dbDEMC.

V. CONCLUSION

In this study, we proposed MSAOPNE method for miRNA–disease association prediction. We have improved the previous methods, which are not suitable for the prediction of diseases without any known associated miRNAs. We have considered all similarity association and integrate all the similarity matrices together, including the Gaussian interactive kernel similarity, miRNA functional similarity, and two types of disease semantic similarity. In addition, we also considered the arbitrary-order proximity for disease–disease similarity matrix, and miRNA–miRNA similarity matrix. The accuracy rate is also improved by nearly 10 points than when the arbitrary order neighboring matrix was not considered previously. Based on the specificity of eigenpair to the symmetric matrix, we use the SVD matrix factorization to extract miRNA feature and disease feature. Finally, we multiply the disease feature matrix and the miRNA feature matrix to obtain the score matrix. In addition, we have implemented three cross-validations and several case studies on important human diseases. Moreover, MSAOPNE performed well in cross-validation and case studies.

The excellent performance of MSAOPNE is mainly attributed to the following important factors. First, the increasing numbers of disease–miRNA association data have been discovered these years due to the rapid development of the biological experiment technology. Several data are combined to predict the association between diseases and

miRNAs, resulting in an accurate model. Second, we use the known associations between the disease and miRNA to complete the initial interaction matrix and then integrate multiple-similarity network. We build a network, which integrated the miRNA functional similarity network, disease phenotype similarity network, and known miRNA–disease network. We also consider the Gaussian kernel similarity. Finally, most importantly, we take the advantage of the contribution of the arbitrarily-order proximity network embedding to obtain the potential relationships in miRNA matrix and disease matrix. Then, we use it to extract the feature, which ultimately determines our scores.

However, some limitations are still observed in this model. First, although known miRNA–disease association data have been more than before, they are still in small quantity for the prediction to obtain enough accurate results. Second, given that the calculation of the arbitrary degree of proximity is only for undirected graphs, our method is not valid for directed graphs. These shortages limit the application range of the MSAOPNE. These factors will be our future work directions.

ACKNOWLEDGEMENTS

The author would like to thank the instructors for giving him valuable advice, and the classmates for their help.

REFERENCES

[1] X. Chen, L. Wang, J. Qu, N. N. Guan, and J. Q. Li, "Predicting MiRNA—Disease association based on inductive matrix completion," *Bioinformatics*, vol. 34, no. 24, pp. 4256–4265, Dec. 2018.

[2] L. Wang, P. Y. Ping, L. N. Kuang, S. T. Ye, F. M. B. Lqbal, and T. R. Pei, "A novel approach based on bipartite network to predict human microbe-disease associations," *Current Bioinf.*, vol. 13, no. 2, pp. 141–148, Apr. 2018.

[3] V. J. N. Ambros, "The functions of animal MicroRNAs," *Nature*, vol. 431, nos. 7–6, pp. 350–355, Sep. 2004.

[4] G. Meister and T. Tuschl, "Mechanisms of gene silencing by double-stranded RNA," *Nature*, vol. 431, pp. 343–349, Sep. 2004.

[5] V. J. C. Ambros, "MicroRNAs: Tiny regulators with great potential," *Cell*, vol. 107, no. 7, pp. 823–826, Dec. 2001.

[6] D. P. Bartel, "MicroRNAs: Genomics, biogenesis, mechanism, and function," *Cell*, vol. 116, no. 2, pp. 281–297, Jan. 2004.

[7] D. P. Bartel, "MicroRNAs: Target recognition and regulatory functions," *Cell*, vol. 136, no. 2, pp. 215–233, Jan. 2009.

[8] R. C. Lee, R. L. Feinbaum, and V. Ambros, "The C. elegans heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*," *Cell*, vol. 75, no. 5, pp. 843–854, Dec. 1993.

[9] C. L. Jopling, Y. Minkyung, A. M. Lancaster, S. M. Lemon, and S. J. S. Peter, "Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA," *Science*, vol. 309, pp. 1577–1581, Sep. 2005.

[10] E. Amiska, "How MicroRNAs control cell division, differentiation and death," *Current Opinion Genet. Develop.*, vol. 15, no. 5, pp. 563–568, Oct. 2005.

[11] X. Karp and V. Ambros, "Developmental biology. Encountering MicroRNAs in cell fate signaling," *Science*, vol. 310, pp. 1288–1289, Nov. 2005.

[12] P. Xu, M. Guo, and B. A. Hay, "MicroRNAs and the regulation of cell death," *Trends Genet.*, vol. 20, no. 12, pp. 617–624, Dec. 2004.

[13] S. Bandyopadhyay, R. Mitra, U. Maulik, and M. Q. Zhang, "Development of the human cancer MicroRNA network," *Silence*, vol. 1, no. 1, p. 6, Dec. 2010.

[14] E. Londin *et al.*, "Analysis of 13 cell types reveals evidence for the expression of numerous novel primate-and tissue-specific MicroRNAs," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 10, pp. E1106–E1115, Mar. 2015.

[15] M. Yousef, W. Khalifa, E. Acar, and J. Allmer, "MicroRNA categorization using sequence motifs and k-mers," *BMC Bioinf.*, vol. 18, no. 1, p. 170, Dec. 2017.

[16] F. Grey, "Role of MicroRNAs in herpesvirus latency and persistence," *J. Gen. Virol.*, vol. 96, no. 4, pp. 739–751, Apr. 2015.

[17] L. Zhijun, L. Dapeng, W. Xinrui, L. Lisheng, and Z. Quan, "Cancer diagnosis through IsomiR expression with machine learning method," *Current Bioinf.*, vol. 13, no. 1, pp. 57–63, Feb. 2018.

[18] W. Tang, S. Wan, Z. Yang, A. E. Teschendorff, and Q. Zou, "Tumor origin detection with tissue-specific miRNA and DNA methylation markers," *Bioinformatics*, vol. 34, no. 3, pp. 398–406, Feb. 2018.

[19] N. H. Heegaard, A. J. Schetter, J. A. Welsh, M. Yoneda, E. D. Bowman, and C. C. Harris, "Circulating micro-RNA expression profiles in early stage nonsmall cell lung cancer," *Int. J. Cancer*, vol. 130, no. 6, pp. 1378–1386, Mar. 2012.

[20] P. Vishal *et al.*, "MIR-17 92 miRNA cluster promotes kidney cyst growth in polycystic kidney disease," *Proc. Nat. Acad. Sci.*, vol. 110, no. 26, pp. 10765–10770, Jun. 2013.

[21] X. Chen, Y. C. Clarence, X. Zhang, Z. H. You, Y. A. Huang, and G. Y. Yan, "HGIMDA: Heterogeneous graph inference for miRNA-disease association prediction," *Oncotarget*, vol. 7, no. 40, pp. 65257–65269, Oct. 2016.

[22] X. Zeng, X. Zhang, and Q. Zou, "Integrative approaches for predicting MicroRNA function and prioritizing disease-related MicroRNA using biological interaction networks," *Briefings Bioinf.*, vol. 17, no. 2, pp. 193–203, Jun. 2016.

[23] J. Q. Li, Z. H. Rong, X. Chen, G. Y. Yan, and Z. H. You, "MCMDA: Matrix completion for MiRNA-disease association prediction," *Oncotarget*, vol. 8, no. 13, pp. 21187–21199, Mar. 2017.

[24] X. Chen, Q. F. Wu, and G. Y. Yan, "RKNNMDA: Ranking-based KNN for MiRNA-disease association prediction," *Biology*, vol. 14, no. 7, pp. 952–962, Jul. 2017.

[25] Y. Huang *et al.*, "Regulatory long non-coding RNA and its functions," *J. Physiol. Biochem.*, vol. 68, no. 4, pp. 611–618, Dec. 2012.

[26] X. Chen and G. Y. Yan, "Semi-supervised learning for potential human MicroRNA-disease associations inference," *Sci. Rep.*, vol. 4, p. 5501, Jun. 2014.

[27] X. Chen *et al.*, "WBSMDA: Within and between score for MiRNA-disease association prediction," *Sci. Rep.*, vol. 6, no. 1, Feb. 2016, Art. no. 21106.

[28] L. Jiang, Y. Xiao, Y. Ding, J. Tang, and F. Guo, "FKL-Spa-LapRLS: An accurate method for identifying human MicroRNA-disease association," *Genomics*, vol. 19, no. 10, p. 911, Dec. 2019.

[29] L. Jiang, Y. Ding, J. Tang, and F. Guo, "MDA-SKF: Similarity kernel fusion for accurately discovering miRNA-disease association," *Frontiers Genet.*, vol. 9, p. 618, Dec. 2018.

[30] X. Zhang, Q. Zou, A. Rodriguez-Paton, and X. Zeng, "Meta-path methods for prioritizing candidate disease miRNAs," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 1, pp. 283–291, Jan. 2019.

[31] 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 Trans. Comput. Biol. Bioinf.*, vol. 14, no. 4, pp. 905–915, Aug. 2016.

[32] Q. Jiang *et al.*, "Prioritization of disease MicroRNAs through a human phenome-MicroRNAome network," *Syst. Biol.*, vol. 4, no. 1, p. S2, May 2010.

[33] X. Li, J. Xu, and Y. Li, "Prioritizing candidate disease MiRNAs by topological features in the miRNA-target dysregulated network," *Syst. Biol. Cancer Res. Drug Discovery*, vol. 10, no. 10, pp. 1857–1866, Oct. 2011.

[34] P. Xuan *et al.*, "Correction: Prediction of MicroRNAs associated with human diseases based on weighted k most similar neighbors," *PLoS ONE*, vol. 8, no. 8, Aug. 2013, Art. no. e70204.

[35] Y. Yang, W. Zhichen, and K. Wei, "Improving clustering of MicroRNA microarray data by incorporating functional similarity," *Current Bioinf.*, vol. 13, no. 1, pp. 34–41, Feb. 2018.

[36] Q. Zou, J. Li, L. Song, X. Zeng, and G. Wang, "Similarity computation strategies in the MicroRNA-disease network: A survey," *Briefings Funct. Genomics*, vol. 15, no. 1, pp. 55–64, Jul. 2016.

[37] P. Xuan *et al.*, "Prediction of potential disease-associated MicroRNAs based on random walk," *Bioinformatics*, vol. 31, no. 11, pp. 1805–1815, Jan. 2015.

[38] X. Chen, "KATZLDA: KATZ measure for the lncRNA-disease association prediction," *Sci. Rep.*, vol. 5, no. 1, Nov. 2015, Art. no. 16840,

[39] X. Zeng, L. Liu, L. Lü, and Q. Zou, "Prediction of potential disease-associated MicroRNAs using structural perturbation method," *Bioinformatics*, vol. 34, no. 14, pp. 2425–2432, 2018.

[40] J. Luo, P. Ding, C. Liang, B. Cao, and X. Chen, "Collective prediction of disease-associated MiRNAs based on transduction learning," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 6, pp. 1468–1475, Dec. 2017.
[41] J. Xu *et al.*, "Prioritizing candidate disease MiRNAs by topological features in the miRNA target-dysregulated network: Case study of prostate cancer," *Mol. Cancer Therapeutics*, vol. 10, no. 10, pp. 1857–1866, Oct. 2011.
[42] Q. Zou *et al.*, "Prediction of MicroRNA-disease associations based on social network analysis methods," *Biomed Res. Int.*, vol. 2015, pp. 1–9, 2015.
[43] Z. H. You *et al.*, "PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction," *PLoS Comput. Biol.*, vol. 13, no. 3, Mar. 2017, Art. no. e1005455.
[44] Y. Li *et al.*, "HMDD v2.0: A database for experimentally supported human MicroRNA and disease associations," *Nucleic Acids Res.*, vol. 42, no. 1, pp. D1070–D1074, Nov. 2013.
[45] X. Chen, C. C. Yan, C. Luo, W. Ji, Y. Zhang, and Q. Dai, "Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity," *Sci. Rep.*, vol. 5, Jun. 2015, Art. no. 11338.
[46] X. Chen, Y. A. Huang, X. S. Wang, Z. H. You, and K. C. Chan, "FMLNCSIM: Fuzzy measure-based lncRNA functional similarity calculation model," *Oncotarget*, vol. 7, no. 29, pp. 45948–45958, Jul. 2016.
[47] Y. A. Huang, X. Chen, Z. H. You, D. S. Huang, and K. C. Chan, "ILNCSIM: Improved lncRNA functional similarity calculation model," *Oncotarget*, vol. 7, no. 18, pp. 25902–25914, May 2016.
[48] W. C. Cho, "Molecular diagnostics for monitoring and predicting therapeutic effect in cancer," *Expert Rev. Mol. Diagnostics*, vol. 14, no. 7, pp. 9–12, Jan. 2011.
[49] M. Berwick and S. Schantz, "Chemoprevention of aerodigestive cancer," *Cancer Metastasis Rev.*, vol. 16, nos. 3–4, pp. 329–347, Sep. 1997.
[50] P. Wang *et al.*, "Early detection of lung cancer in serum by a panel of MicroRNA biomarkers," *Clin. Lung Cancer*, vol. 16, no. 4, pp. 313–319, Jul. 2015.

XIUJIN WU received the master's degree from Xiamen University, China, in 2008, where she is currently pursuing the Ph.D. degree with the Software School. Her researches focus on machine learning and bioinformatics.

WENHUA ZENG received the M.S. and Ph.D. degrees in automation profession from Zhejiang University, in 1986 and 1989, respectively. He is currently a Professor with the Software School, Xiamen University. His current research interests include machine learning, data mining, and evolutionary algorithms.

YUXIU XU received the bachelor's degree in computer science from Fujian Normal University, China, in 2017. She is currently pursuing the master's degree in software engineering with the Department of Xiamen University, China. She is interested in machine learning and bio-network data mining.

BEIZHAN WANG received the B.E., M.E., and Ph.D. degrees from Northwestern Polytechnical University, China, in 1987, 1997, and 2003, respectively. He is currently a Professor with the Software School, Xiamen University, China. His research interests include pattern recognition, machine learning, and data mining.

XIANGRONG LIU received the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China. He finished his postdoctoral training with the School of Electronics Engineering and Computer Science, Peking University. In 2009, he joined Xiamen University, and is currently a Professor with the Department of Computer Science. He has published over 40 journal and conference papers as well as one book chapters. His researches focus on data mining and bioinformatics.

FAN LIN received the M.S. and Ph.D. degrees from Xiamen University, in 2003 and 2013, respectively. He is currently an Associate Professor with the Software School, Xiamen University. His major research interests include online collaborative filter, machine learning, and recommender systems.

GIL ALTEROVITZ received the B.S. degree from Carnegie Mellon University and the S.M. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), and the Ph.D. degree in electrical and biomedical engineering from MIT. He is currently an Assistant Professor with the Harvard Medical School. His major research interests include Bayesian methods and network models.