



## OPEN ACCESS

EDITED BY  
Xuefeng Cui,  
Shandong University, China

REVIEWED BY  
Jin-Xing Liu,  
Qufu Normal University, China  
Wei Lan,  
Guangxi University, China

\*CORRESPONDENCE  
Minzhu Xie,  
xieminzhu@hunnu.edu.cn

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

RECEIVED 27 June 2022  
ACCEPTED 18 July 2022  
PUBLISHED 22 August 2022

CITATION  
Rao Y, Xie M and Wang H (2022), Predict  
potential miRNA-disease associations  
based on bounded nuclear  
norm regularization.  
*Front. Genet.* 13:978975.  
doi: 10.3389/fgene.2022.978975

COPYRIGHT  
© 2022 Rao, Xie and Wang. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which does  
not comply with these terms.

# Predict potential miRNA-disease associations based on bounded nuclear norm regularization

Yidong Rao, Minzhu Xie\* and Hao Wang

College of Information Science and Engineering, Hunan Normal University, Changsha, China

Increasing evidences show that the abnormal microRNA (miRNA) expression is related to a variety of complex human diseases. However, the current biological experiments to determine miRNA-disease associations are time consuming and expensive. Therefore, computational models to predict potential miRNA-disease associations are in urgent need. Though many miRNA-disease association prediction methods have been proposed, there is still a room to improve the prediction accuracy. In this paper, we propose a matrix completion model with bounded nuclear norm regularization to predict potential miRNA-disease associations, which is called BNNRMDA. BNNRMDA at first constructs a heterogeneous miRNA-disease network integrating the information of miRNA self-similarity, disease self-similarity, and the known miRNA-disease associations, which is represented by an adjacent matrix. Then, it models the miRNA-disease prediction as a relaxed matrix completion with error tolerance, value boundary and nuclear norm minimization. Finally it implements the alternating direction method to solve the matrix completion problem. BNNRMDA makes full use of available information of miRNAs and diseases, and can deal with the data containing noise. Compared with four state-of-the-art methods, the experimental results show BNNRMDA achieved the best performance in five-fold cross-validation and leave-one-out cross-validation. The case studies on two complex human diseases showed that 47 of the top 50 prediction results of BNNRMDA have been verified in the latest HMDD database.

## KEYWORDS

miRNA, disease, miRNA-disease associations, bounded nuclear norm regularization, matrix completion

## 1 Introduction

MicroRNA (miRNA) is a non-coding single-stranded RNA molecule of about 22 nt in length, which have been proved involved in gene regulation by binding to 3' UTRs of the target mRNAs. It plays a critical role in human cell differentiation, growth, and disease development. Accumulating evidence has shown that miRNAs are closely related to complex human diseases (Liu et al. (2010); Chen et al. (2019); Feng et al. (2012); Zhang et al. (2013)), and discovering miRNA-disease associations is of great significance for the prevention, diagnosis and treatment of human complex diseases. Recently, many miRNA-

disease associations have been confirmed and collected in different databases. For example, the HMDD v3.2 database (<http://www.cuilab.cn/hmdd>) contains 32281 confirmed associations between 850 diseases and 1102 miRNAs (Huang et al. (2019)). However, it is time-consuming and labor intensive for current biological experiments to determine miRNA-disease associations. Therefore, effective computational prediction models are in urgent need.

As so far, a number of computational miRNA-disease associations prediction models have been proposed [Chen et al. (2019); Chen and Zhang (2013); Jiang et al. (2013); Zeng et al. (2016b,a)], and all the models are based on the known miRNA-disease associations to predict the potential associations. For example, based on the known miRNA-disease associations and a miRNA-miRNA functional similarity network, Chen et al. [Chen X. et al. (2012)] developed a method RWRMDA, which used a global network similarity measurement and random walk with restart to predict potential miRNA-disease associations. Based on a miRNA-disease bilayer network constructed according to the above information, Xuan et al. [Xuan et al. (2015)] presented a method using random walk with restart to infer potential associations between miRNAs and diseases. By including an extra disease similarity network, Liao et al. [Liao et al. (2015)] proposed a diffusion-based method NDBM, which also used a global network similarity to predict miRNA-disease associations. Furthermore, Chen et al. (Chen et al. (2016)) integrated the information of miRNA functional similarity, disease semantic similarity, Gaussian interaction profile kernel similarity and the known miRNA-disease associations to build a heterogeneous network, and proposed a new prediction method HGIMDA. You et al. [You et al. (2017)] proposed a path-based prediction model PBMDA, which constructed a similar heterogeneous network and used a depth-first search algorithm to predict potential associations. Based on the heterogeneous network, Chen et al. [Chen et al. (2018c)] proposed a method BNPMDA, which adopted a bipartite network recommendation algorithm to infer potential associations between miRNAs and diseases.

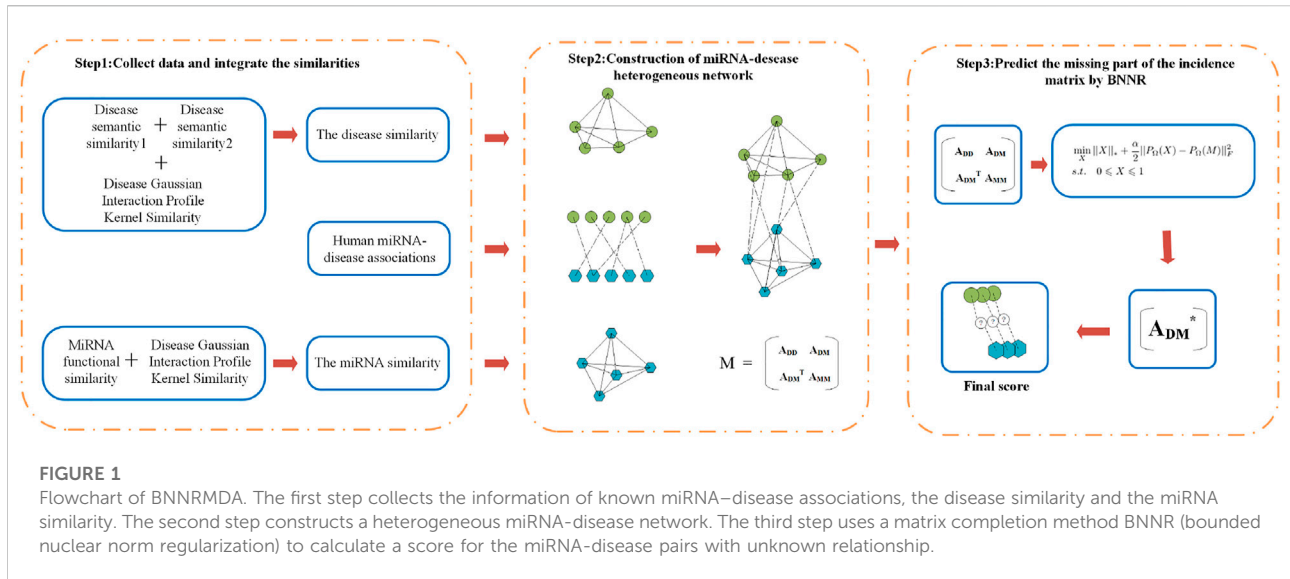
Recently, machine learning methods have been applied to miRNA-disease potential association prediction [Jiang et al. (2013); Chen and Yan (2014); Chen et al. (2018a); Zheng et al. (2019); Zeng et al. (2019); Liang et al. (2019); Li et al. (2020); Zhou et al. (2021)]. For example, Jiang et al. [Jiang et al. (2013)] used support vector machine (SVM) to predict miRNA-disease interaction. Chen et al. [Chen et al. (2018a)] employed a random forest algorithm to predict miRNA-disease associations and proposed a prediction model RFMDA. RFMDA can effectively distinguish related miRNA-disease pairs from unrelated miRNA-disease pairs. Zheng et al. [Zheng et al. (2019)] presented a prediction model MLMDA. MLMDA first used a deep auto-encoder neural network to extract features from the information of disease semantic similarity, Gaussian interaction profile kernel similarity, miRNA functional

similarity and miRNA sequences, and adopted a random forest classifier to predict potential associations between miRNAs and diseases based on the extracted features. Liang et al. [Liang et al. (2019)] proposed a method AMVML to infer disease-related miRNAs based on adaptive multi-view multi-label learning. Li et al. [Li et al. (2020)] proposed a miRNA-disease association prediction model NIMCGCN. NIMCGCN used graph convolutional networks to obtain the features of miRNA and disease, and then adopted a neural inductive matrix completion model to infer a new association matrix. Based on graph embedding and multiple meta-paths fusion, Zhang et al. [Zhang et al. (2020)] proposed a model M2GMDA to predict miRNA-disease associations. Based on a heterogeneous network integrating various known associations between miRNA, disease, protein, long non-coding RNA (lncRNA) and drugs, Li et al. [Li H. Y. et al. (2021)] proposed a miRNA-disease association prediction model DF-MDA. DF-MDA adopted a diffusion-based machine-learning method to extract features from the network, and a random forest classifier to predict miRNA-disease associations. Besides, other techniques such as structural deep network embedding [Gong et al. (2019)] and matrix decomposition [Chen et al. (2021); Li L. et al. (2021)] are also used in miRNA-disease association prediction.

To further improve the performance of miRNA-disease association prediction, we propose a novel Bounded Nuclear Norm Regularization based miRNA-disease association prediction model BNNRMDA. At first, BNNRMDA integrates the information of the disease semantic similarity, the miRNA functional similarity, the Gaussian interaction profile kernel similarity and the experimentally verified miRNA-disease associations to construct a heterogeneous miRNA-disease network. Since the number of verified miRNA-disease associations is very small than the total miRNA-disease pairs, the adjacent matrix of the network is sparse, BNNRMDA uses a bounded nuclear norm regularization method to complete the sparse matrix, and the element value of the completed matrix indicates the likelihood that the corresponding miRNA and disease are related. The experiments of leave-one-out cross-validation and five-fold cross-validation in a benchmark dataset showed that BNNRMDA is effective to predict potential miRNA-disease associations. In addition, case studies of colon neoplasms and lung neoplasms showed that the accuracy of BNNRMDA reached 94%.

## 2 Methods

Figure 1 gives a flowchart of BNNRMDA. The process of BNNRMDA consists of 3 steps. The first step collects the information of known miRNA-disease associations, the disease similarity and the miRNA similarity. The second step constructs a heterogeneous miRNA-disease network and obtain the corresponding adjacent matrix  $M$ . The third step uses a



matrix completion method to complete  $M$ , and predicts potential miRNA–disease associations based on the completed matrix.

## 2.1 Data and similarity measures

### 2.1.1 Validated human miRNA–disease associations

As most of similar works, the HMDD v2.0 database [Li et al. (2014)] was used as the benchmark dataset, which contains 5430 verified associations between 495 miRNAs and 383 diseases. For convenience, we used a  $n_d \times n_m$  binary matrix  $A_{DM}$  to store the validated associations from the database with  $n_d = 383$  and  $n_m = 495$ . If the  $i$ th disease has a known association with the  $j$ th miRNA, then the element  $A_{DM_{i,j}}$  is set to 1, otherwise it is 0.

### 2.1.2 MiRNA functional similarity

Based on the hypothesis that miRNAs with similar functions are more likely to be related to similar diseases, Wang et al. [Wang et al. (2010)] calculated the similarity between miRNAs based on the similarity of their associated disease DAGs. The miRNA functional similarity data was directly downloaded via the link <http://www.cuilab.cn/files/images/cuilab/misim.zip> provided by Wang et al. [Wang et al. (2010)]. We used a matrix  $FS$  to represent the data, where the element  $FS_{i,j}$  represents the functional similarity score between the  $i$ th miRNA and the  $j$ th miRNA.

### 2.1.3 Disease semantic similarity

We combined two disease semantic similarity measures to calculate the semantic similarity between two diseases. The first was introduced by Wang et al. (Wang et al. (2010)), which is based on the medical subject headings (MeSH) descriptors. The

MeSH descriptor of a disease is organized as a hierarchical directed acyclic graph (DAG) with each node is a disease term. For a disease  $D$ , let the DAG corresponding to its MeSH descriptor be  $DAG(D) = (T(D), E(D))$ .  $T(D)$  includes the node  $D$  and its ancestor nodes (more general disease term), and  $E(D)$  is the set of direct edges representing the parent-child relationship between the disease terms.

The semantic contribution of a disease term  $t$  to  $D$  in  $DAG(D)$  is defined as Eq. 1.

$$C_{1D}(t) = \begin{cases} 1 & \text{if } t = D \\ \max\{\theta C_{1D}(t') \mid t' \text{ is a child of } t\} & \text{if } t \neq D \end{cases} \quad (1)$$

where  $\theta$  is the semantic contribution factor, and is set 0.5 as suggested in Wang et al. (2010).

The semantic value of disease  $D$  is calculated by Eq. 2, and the Wang’s similarity between diseases  $d_i$  and  $d_j$  is defined as Eq. 3.

$$V_1(D) = \sum_{d \in T(D)} C_{1D}(d). \quad (2)$$

$$SS_1(d_i, d_j) = \frac{\sum_{t \in T(d_i) \cap T(d_j)} (C_{1d_i}(t) + C_{1d_j}(t))}{V_1(d_i) + V_1(d_j)}. \quad (3)$$

The second disease similarity measure was introduced by Xuan et al. (Xuan et al. (2013)), and it is defined as Eq. 4.

$$SS_2(d_i, d_j) = \frac{2 \sum_{t \in T(d_i) \cap T(d_j)} IC(t)}{\sum_{t \in T(d_i)} IC(t) + \sum_{t \in T(d_j)} IC(t)}. \quad (4)$$

$IC(t)$  is the information content of the likelihood of  $t$  occurring as a node in a disease DAG, and can be calculated as follows.

$$IC(t) = -\log \left[ \frac{\text{the number of DAGs containing } t}{\text{the total number of DAGs}} \right]. \quad (5)$$

Finally, we average the above two similarity measures of  $d_i$  and  $d_j$  and obtain the combined disease semantic similarity.

$$SS(d_i, d_j) = \frac{SS_1(d_i, d_j) + SS_2(d_i, d_j)}{2}. \tag{6}$$

### 2.1.4 Gaussian interaction profile kernel similarity

Based on similar diseases may be related to miRNAs with similar functions, Gaussian interaction profile kernel (GIPK) similarity has been widely used to calculate miRNA similarity and disease similarity. Let  $K(d_i)$  be the vector containing elements at the  $i$ th row of the binary miRNA-disease association matrix  $A_{DM}$ , and  $K(m_j)$  be the vector containing elements at the  $j$ th column of  $A_{DM}$ .  $K(d_i)$  and  $K(m_j)$  represent the interaction profiles of disease  $d_i$  and miRNA  $m_j$  respectively.

The equations to calculate the disease GIPK similarity and the miRNA GIPK similarity are as follows.

$$GKSD(d_i, d_j) = \exp(-\rho_d \|K(d_i) - K(d_j)\|^2) \tag{7}$$

$$GKSM(m_i, m_j) = \exp(-\rho_m \|K(m_i) - K(m_j)\|^2) \tag{8}$$

The coefficients  $\rho_d$  and  $\rho_m$  are defined in the following equations.

$$\rho_d = \frac{n_d}{\sum_{i=1}^{n_d} \|K(d_i)\|^2} \tag{9}$$

$$\rho_m = \frac{n_m}{\sum_{i=1}^{n_m} \|K(m_i)\|^2} \tag{10}$$

## 2.2 Similarity integration and heterogeneous network construction

Since some diseases do not have any MeSH descriptor, we cannot calculate the semantic similarity between these diseases and others. In the case, we use GIPK similarity to replace the semantic similarity. Similarly, when the functional similarity between two miRNAs is missing, the corresponding GIPK similarity is used instead. Finally we obtain a disease similarity matrix  $A_{DD}$  and a miRNA similarity matrix  $A_{MM}$  as follows.

$$A_{DD_{i,j}} = \begin{cases} SS(d(i), d(j)) & \text{if } SS(d(i), d(j)) \neq 0, \\ GKSD(d(i), d(j)) & \text{otherwise.} \end{cases} \tag{11}$$

$$A_{MM_{i,j}} = \begin{cases} FS(m(i), m(j)) & \text{if } FS(m(i), m(j)) \neq 0, \\ GKSM(m(i), m(j)) & \text{otherwise.} \end{cases} \tag{12}$$

We integrate the information of disease similarity, miRNA similarity, the known miRNA-disease associations into a heterogeneous miRNA-disease network. The heterogeneous miRNA-disease network is encoded into a  $(n_d + n_m) \times (n_d + n_m)$  matrix  $M$  as follows.

$$M = \begin{bmatrix} A_{DD} & A_{DM} \\ A_{DM}^T & A_{MM} \end{bmatrix}.$$

## 2.3 Matrix completion with bounded nuclear norm regularization

Since the verified miRNA-disease associations are much less than the total miRNA-disease pairs,  $A_{DM}$  is very sparse (most elements are 0). The miRNA-disease association prediction problem can be model as the matrix completion problem of  $M$ . If  $M_{i,j}$  corresponds to a known miRNA-disease association, indicates a miRNA similarity or a disease similarity, it called a known entry. Let  $\Omega = \{(i, j) | M_{i,j} \text{ is a known entry}\}$ . The goal of our miRNA-disease association prediction is to find appropriate values for the unknown entries of  $M$  as the final miRNA-disease association prediction scores.

The matrix completion problem of  $M$  is generally formulated as find a matrix  $M^*$  such that the projections of  $M^*$  and  $M$  onto  $\Omega$  are equal and the rank of  $M^*$  is minimized, and the formulation is as follows:

$$\begin{aligned} & \min_{M^*} \text{rank}(M^*) \\ \text{s.t. } & \mathcal{P}_\Omega(M^*) = \mathcal{P}_\Omega(M), \end{aligned}$$

where  $\mathcal{P}_\Omega(\cdot)$  is the projection function such that

$$\mathcal{P}_\Omega(X)_{ij} = \begin{cases} X_{ij}, & (i, j) \in \Omega \\ 0, & (i, j) \notin \Omega \end{cases}.$$

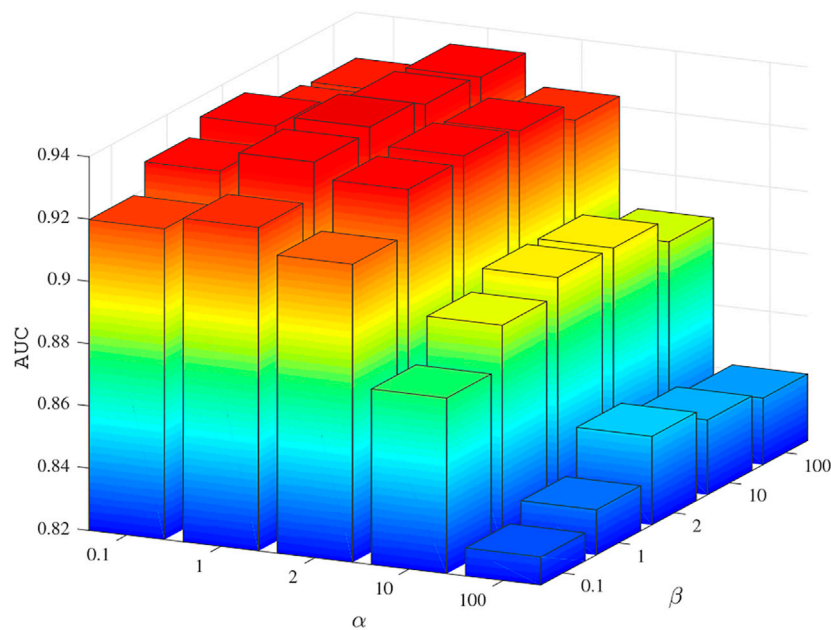
However the rank minimization matrix completion problem is NP-hard. Inspired by Yang et al. (2019), we model the miRNA-disease association prediction as a relaxed matrix completion with error tolerance, value boundary and nuclear norm minimization, which is called the BNNR (bounded nuclear norm regularization) model (Yang et al. (2019)) and is formulated as follows.

$$\begin{aligned} & \min_X \|X\|_* + \frac{\alpha}{2} \|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(M)\|_F^2 \\ \text{s.t. } & 0 \leq X \leq 1, \end{aligned} \tag{13}$$

where  $\|X\|_*$  is the nuclear norm of  $X$ , i.e. the sum of all singular values of  $X$ .  $\|\cdot\|_F$  is the Frobenius norm and  $\alpha$  is the parameter that balances the nuclear norm and error term. The BNNR model is a convex optimization problem, and many effective algorithms such as AMM (alternating direction method) (Chen C. H. et al. (2012)) could be used solve it.

To use AMM (Chen C. H. et al. (2012)) to solve the BNNR model, we introduce an auxiliary matrix  $H$ , and the BNNR model is equivalent to the following model.

$$\begin{aligned} & \min_X \|X\|_* + \frac{\alpha}{2} \|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(M)\|_F^2 \\ \text{s.t. } & X = H, \quad 0 \leq H \leq 1. \end{aligned} \tag{14}$$



**FIGURE 2**  
The AUC values using different  $\alpha$  and  $\beta$  values in five fold CV experiments on the training dataset.

Therefore, the extended Lagrange function is:

$$\mathcal{L}(H, X, Y, \alpha, \beta) = \|X\|_* + \frac{\alpha}{2} \|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(M)\|_F^2 + \text{Tr}(Y^T(X - H)) + \frac{\beta}{2} \|X - H\|_F^2, \quad (15)$$

where  $\beta > 0$  is the penalty parameter and  $Y$  is the Lagrange multiplier. The model (14) could be solved by an iterative process, whose details could be found in (Yang et al. (2019)).

After a series of iterations, a convergent  $H$  would be finally obtained, which is denoted by  $H^*$ . Let

$$M^* = \begin{bmatrix} A_{DD}^* & A_{DM}^* \\ A_{DM}^{*T} & A_{MM}^* \end{bmatrix} = H^*. \quad (16)$$

The predicted miRNA-disease associations are found from matrix  $A_{DM}^*$ .

## 3 Results

### 3.1 Parameter setting

The values of parameters  $\alpha$  and  $\beta$  were determined by 5-fold cross-validation experiments on the benchmark dataset. The values were chosen from 0.1, 1, 2, 10, 100, and the AUC results are shown in Figure 2. The experimental results show that when  $\alpha = 1$  and  $\beta = 2$ , BNNRMDA achieved the best performance. Therefore, in the following experiments, we set  $\alpha = 1$  and  $\beta = 2$ .

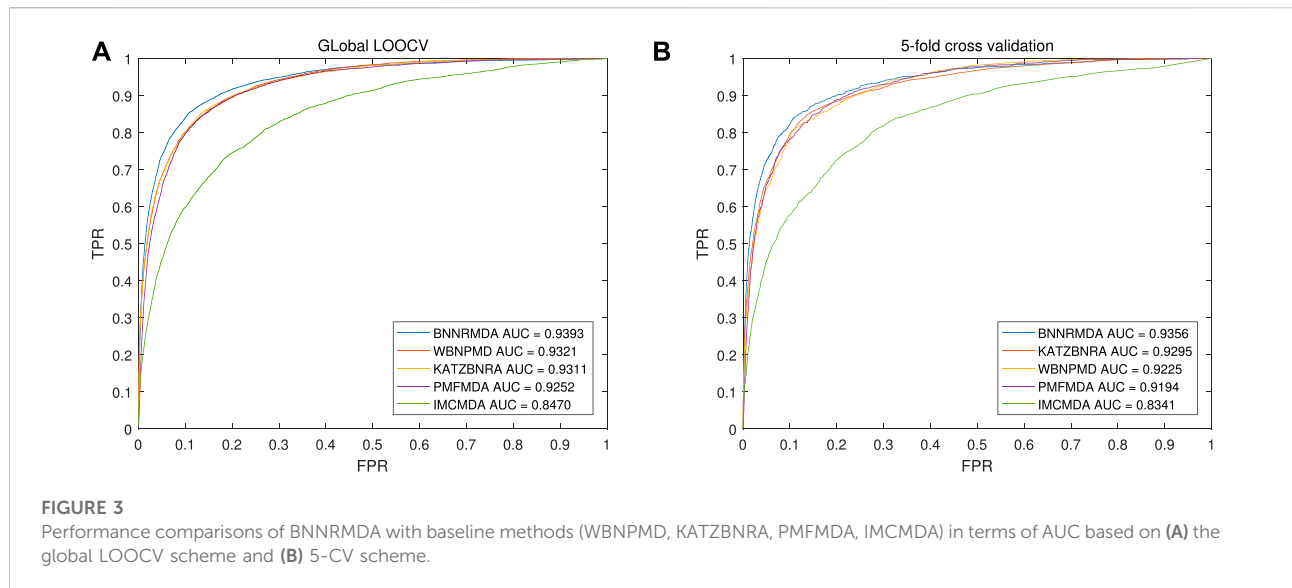
### 3.2 Performance evaluation

We compared BNNRMDA with four state-of-the-art methods IMCMDA (Chen et al. (2018b)), KATZBNRA (Li et al. (2019)), PMFMDA (Xu et al. (2019)) and WBNPMD (Xie et al. (2019)) using global leave-one-out cross-validation (LOOCV) and 5-fold cross-validation (5-fold CV).

The benchmark dataset was from the HMDD v2.0 database, which contains 5430 known miRNA-disease associations. Under the global LOOCV framework, each known association is selected out for testing, the others are used as the training set, and all unknown miRNA-disease associations will be used as candidate associations. After BNNRMDA calculates all associated prediction scores, the rank of each test sample will be obtained by comparing with the candidate samples. Higher the rank of the test sample, more effective our model is. We changed the threshold to calculate the true positive rate (TPR) and false positive rate (FPR) and drew the ROC curve. The area under the ROC curve (AUC) was calculated to compare the performance. The higher the AUC value, the better the performance of the model. The experimental results of the global LOOCV is shown in Figure 3 (a). The AUC values of BNNRMDA, IMCMDA, KATZBNRA, PMFMDA, and WBNPMD are 0.9393, 0.8470, 0.9311, 0.9252, 0.9321 respectively.

In the 5-CV framework, all known miRNA-disease association pairs will be randomly divided into five parts; in each experiment, one part is tested, and the other four parts are used as a training set. Similar to LOOCV, the AUC values are used to compare the performance of these models. Figure 3 (b)





**TABLE 1** The top 50 potential miRNAs associated with colon neoplasms.

(2 pt) miRNA	Evidence	miRNA	Evidence
hsa-mir-155	dbDEMCM;HMDD	hsa-mir-31	dbDEMCM;HMDD
hsa-mir-21	dbDEMCM;HMDD	hsa-mir-146b	dbDEMCM
hsa-mir-146a	dbDEMCM;HMDD	hsa-mir-141	dbDEMCM;HMDD
hsa-mir-20a	dbDEMCM;HMDD	hsa-mir-199a	unconfirmed
hsa-mir-16	dbDEMCM	hsa-mir-24	dbDEMCM;HMDD
hsa-mir-125b	dbDEMCM;HMDD	hsa-let-7a	dbDEMCM;HMDD
hsa-mir-15b	dbDEMCM;HMDD	hsa-mir-150	dbDEMCM;HMDD
hsa-mir-29b	dbDEMCM;HMDD	hsa-mir-200b	dbDEMCM;HMDD
hsa-mir-143	dbDEMCM;HMDD	hsa-mir-7	dbDEMCM
hsa-mir-101	HMDD	hsa-mir-9	dbDEMCM
hsa-mir-19b	dbDEMCM	hsa-mir-148a	dbDEMCM;HMDD
hsa-mir-34a	dbDEMCM;HMDD	hsa-let-7c	dbDEMCM;HMDD
hsa-mir-29a	dbDEMCM;HMDD	hsa-mir-221	dbDEMCM;HMDD
hsa-mir-106b	dbDEMCM;HMDD	hsa-mir-23a	dbDEMCM;HMDD
hsa-mir-19a	dbDEMCM;HMDD	hsa-mir-107	dbDEMCM;HMDD
hsa-mir-196a	dbDEMCM;HMDD	hsa-mir-133b	dbDEMCM;HMDD
hsa-mir-125a	dbDEMCM;HMDD	hsa-mir-34c	unconfirmed
hsa-mir-1	dbDEMCM;HMDD	hsa-mir-25	dbDEMCM;HMDD
hsa-mir-15a	dbDEMCM;HMDD	hsa-mir-30c	dbDEMCM;HMDD
hsa-mir-223	dbDEMCM;HMDD	hsa-mir-29c	dbDEMCM
hsa-mir-214	dbDEMCM	hsa-let-7b	dbDEMCM;HMDD
hsa-mir-133a	dbDEMCM;HMDD5	hsa-mir-26a	unconfirmed
hsa-mir-132	dbDEMCM;HMDD	hsa-mir-203	dbDEMCM;HMDD
hsa-mir-18a	dbDEMCM;HMDD	hsa-let-7i	dbDEMCM;HMDD
hsa-mir-92a	dbDEMCM;HMDD	hsa-mir-222	dbDEMCM;HMDD

**TABLE 2** The top 50 potential miRNAs associated with lung neoplasms.

(2 pt) miRNA	Evidence	miRNA	Evidence
hsa-mir-106b	dbDEMCM	hsa-mir-429	dbDEMCM
hsa-mir-20b	dbDEMCM	hsa-mir-296	unconfirmed
hsa-mir-130a	dbDEMCM;HMDD	hsa-mir-129	dbDEMCM;HMDD
hsa-mir-16	dbDEMCM;HMDD	hsa-mir-708	dbDEMCM
hsa-mir-23b	dbDEMCM	hsa-mir-211	dbDEMCM
hsa-mir-342	dbDEMCM;HMDD	hsa-mir-196b	dbDEMCM;HMDD
hsa-mir-15a	dbDEMCM;HMDD	hsa-mir-302c	dbDEMCM
hsa-mir-378a	unconfirmed	hsa-mir-302b	dbDEMCM
hsa-mir-195	dbDEMCM;HMDD	hsa-mir-328	dbDEMCM;HMDD
hsa-mir-15b	dbDEMCM	hsa-mir-99b	dbDEMCM
hsa-mir-122	dbDEMCM;HMDD	hsa-mir-149	dbDEMCM;HMDD
hsa-mir-193b	dbDEMCM	hsa-mir-423	HMDD
hsa-mir-424	dbDEMCM	hsa-mir-152	dbDEMCM;HMDD
hsa-mir-144	dbDEMCM;HMDD	hsa-mir-449b	dbDEMCM
hsa-mir-92b	dbDEMCM	hsa-mir-194	dbDEMCM;HMDD
hsa-mir-130b	dbDEMCM;HMDD	hsa-mir-208a	HMDD
hsa-mir-204	dbDEMCM	hsa-mir-302a	dbDEMCM
hsa-mir-451a	dbDEMCM;HMDD	hsa-mir-491	dbDEMCM
hsa-mir-99a	dbDEMCM;HMDD	hsa-mir-452	dbDEMCM
hsa-mir-449a	dbDEMCM;HMDD	hsa-mir-373	dbDEMCM;HMDD
hsa-mir-10a	dbDEMCM;HMDD	hsa-mir-625	dbDEMCM
hsa-mir-141	dbDEMCM;HMDD	hsa-mir-181d	dbDEMCM
hsa-mir-139	dbDEMCM;HMDD	hsa-mir-367	dbDEMCM
hsa-mir-151a	unconfirmed	hsa-mir-520a	dbDEMCM
hsa-mir-28	dbDEMCM	hsa-mir-520d	dbDEMCM

shows the results of the 5CV experiment, and BNNRMDA achieved the best AUC of 0.9356. The AUC values of IMCMDA, KATZBNRA, PMFMDA, and WBNPMD are 0.8341, 0.9295, 0.9194, 0.9225 respectively.

### 3.3 Case studies

In order to further verify the effect of the BNNRMDA model, we conducted case studies on two human diseases colon neoplasms and lung neoplasms. These diseases pose a great threat to human life. For example, lung neoplasms are one of the common neoplasms in the human body. In recent years, a large number of colon neoplasms cases have died, posing a major threat to human life (DeSantis et al. (2019)). For colon neoplasms, after removing all known related miRNA-disease pairs, we rank the final prediction results of the miRNA related to them. We use two miRNA-disease association databases for verification, namely the dbDEMC (Yang et al. (2017)) database and the HMDD (Huang et al. (2019)) database. As can be clearly seen in Table 1, 47 of the top 50 prediction results have been confirmed to be related in dbDEMC2.0 and HMDD v3.2. Similarly, the results of the top 50 miRNAs predicted for lung neoplasms are shown in Table 2. Among them, 47 of the top 50 can be confirmed in dbDEMC2.0 and HMDD v3.2.

## 4 Conclusion

We proposed a new miRNA-disease association prediction model BNNRMDA. BNNRMDA constructs a miRNA-disease heterogeneous network by integrating miRNA similarity network, disease similarity network and miRNA-disease known association network, and formulates the miRNA-disease association prediction problem as a relaxed matrix completion with error tolerance, value boundary and nuclear norm minimization (BNNR), and at last uses alternating direction method (AMM) to obtain an optimal solution. The global leave-one-out cross-validation experiments and the five-fold cross-validation framework experiments on the benchmark dataset show that BNNRMDA performs better than four state-of-the-art methods. In addition, the case studies on two complex human diseases also illustrate the reliability of BNNRMDA. BNNRMDA can be an effective tool to identify potential miRNA-disease associations. There are some factors that will affect the final prediction results of BNNRMDA. First of all, the materials we used include experimentally verified miRNA-

disease associations, miRNA functional similarities, and disease semantic similarities. These data may contain noises and outliers, and appropriate preprocess of the data might enhance the prediction accuracy of BNNRMDA. The choice of parameters  $\alpha$  and  $\beta$  has a certain impact on the prediction performance, and how to choose appropriate parameters based on some statistical characteristics of data is challenging.

## Data Availability Statement

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.cuilab.cn/files/images/cuilab/misim.zip>.

## Author contributions

YY and MX conceived the study and the conceptual design of the work. YY and HW collected the data and implemented the algorithm BNNRMDA. YY tested the algorithms and drafted the manuscript. MX and HW polished the manuscript. All authors have read and approved the manuscript.

## Funding

This work has been supported by the National Natural Science Foundation of China (Grant Nos. 62172028 and 61772197).

## Conflict of Interest

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

## Publisher's Note

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

## References

Chen, C. H., He, B. S., and Yuan, X. M. (2012a). Matrix completion via an alternating direction method. *IMA J. Numer. Analysis* 32, 227–245. doi:10.1093/imanum/drq039

Chen, H., and Zhang, Z. (2013). Similarity-based methods for potential human microRNA-disease association prediction. *BMC Med. Genomics* 6, 12. doi:10.1186/1755-8794-6-12

- Chen, X., Liu, M. X., and Yan, G. Y. (2012b). Rwrmda: Predicting novel human microRNA-disease associations. *Mol. Biosyst.* 8, 2792–2798. doi:10.1039/c2mb25180a
- Chen, X., Sun, L. G., and Zhao, Y. (2021). Ncmcmda: miRNA-disease association prediction through neighborhood constraint matrix completion. *Brief. Bioinform.* 22, 485–496. doi:10.1093/bib/bbz159
- Chen, X., Wang, C. C., Yin, J., and You, Z. H. (2018a). Novel human miRNA-disease association inference based on random forest. *Mol. Ther. Nucleic Acids* 13, 568–579. doi:10.1016/j.omtn.2018.10.005
- Chen, X., Wang, L., Qu, J., Guan, N. N., and Li, J. Q. (2018b). Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* 34, 4256–4265. doi:10.1093/bioinformatics/bty503
- Chen, X., Xie, D., Wang, L., Zhao, Q., You, Z. H., and Liu, H. (2018c). Bnpmda: Bipartite network projection for miRNA-disease association prediction. *Bioinformatics* 34, 3178–3186. doi:10.1093/bioinformatics/bty333
- Chen, X., Xie, D., Zhao, Q., and You, Z. H. (2019). MicroRNAs and complex diseases: From experimental results to computational models. *Brief. Bioinform.* 20, 515–539. doi:10.1093/bib/bbx130
- Chen, X., Yan, C. C., Zhang, X., You, Z. H., Huang, Y. A., and Yan, G. Y. (2016). Hgmda: Heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget* 7, 65257–65269. doi:10.18632/oncotarget.11251
- Chen, X., and Yan, G. Y. (2014). Semi-supervised learning for potential human microRNA-disease associations inference. *Sci. Rep.* 4, 5501. doi:10.1038/srep05501
- DeSantis, C. E., Miller, K. D., Goding Sauer, A., Jemal, A., and Siegel, R. L. (2019). Cancer statistics for african Americans, 2019. *Ca. Cancer J. Clin.* 69, 211–233. doi:10.3322/caac.21555
- Feng, B., Dong, T. T., Wang, L. L., Zhou, H. M., Zhao, H. C., Dong, F., et al. (2012). Colorectal cancer migration and invasion initiated by microRNA-106a. *PLoS One* 7, e43452. doi:10.1371/journal.pone.0043452
- Gong, Y., Niu, Y., Zhang, W., and Li, X. (2019). A network embedding-based multiple information integration method for the miRNA-disease association prediction. *BMC Bioinforma.* 20, 468. doi:10.1186/s12859-019-3063-3
- Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., et al. (2019). HMDD v3.0: A database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.* 47, D1013–D1017. doi:10.1093/nar/gky1010
- Jiang, Q., Wang, G., Jin, S., Yu, L., and Wang, Y. (2013). Predicting human microRNA-disease associations based on support vector machine. *Int. J. Data Min. Bioinform.* 8, 282–293. doi:10.1504/ijdm.2013.056078
- Li, H. Y., You, Z. H., Wang, L., Yan, X., and Li, Z. W. (2021a). DF-MDA: An effective diffusion-based computational model for predicting miRNA-disease association. *Mol. Ther.* 29, 1501–1511. doi:10.1016/j.ymt.2021.01.003
- Li, J., Zhang, S., Liu, T., Ning, C. X., Zhang, Z. X., and Zhou, W. (2020). Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction. *Bioinformatics* 36, 2538–2546. doi:10.1093/bioinformatics/btz965
- Li, L., Gao, Z., Wang, Y. T., Zhang, M. W., Ni, J. C., Zheng, C. H., et al. (2021b). Scmfmda: Predicting microRNA-disease associations based on similarity constrained matrix factorization. *PLoS Comput. Biol.* 17, e1009165. doi:10.1371/journal.pcbi.1009165
- Li, S., Xie, M., and Liu, X. (2019). A novel approach based on bipartite network recommendation and KATZ model to predict potential micro-disease associations. *Front. Genet.* 10, 1147. doi:10.3389/fgene.2019.01147
- Li, Y., Qiu, C., Tu, J., Geng, B., Yang, J., Jiang, T., et al. (2014). HMDD v2.0: A database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* 42, D1070–D1074. doi:10.1093/nar/gkt1023
- Liang, C., Yu, S., and Luo, J. (2019). Adaptive multi-view multi-label learning for identifying disease-associated candidate miRNAs. *PLoS Comput. Biol.* 15, e1006931. doi:10.1371/journal.pcbi.1006931
- Liao, B., Ding, S., Chen, H., Li, Z., and Cai, L. (2015). Identifying human microRNA-disease associations by a new diffusion-based method. *J. Bioinform. Comput. Biol.* 13, 1550014. doi:10.1142/S0219720015500146
- Liu, W., Zabornyk, O., Wang, H., Shiao, Y. H., Nickerson, M. L., Khalil, S., et al. (2010). Mir-23b targets proline oxidase, a novel tumor suppressor protein in renal cancer. *Oncogene* 29, 4914–4924. doi:10.1038/nc.2010.237
- Wang, D., Wang, J., Lu, M., Song, F., and Cui, Q. (2010). Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* 26, 1644–1650. doi:10.1093/bioinformatics/btq241
- Xie, G., Fan, Z., Sun, Y., Wu, C., and Ma, L. (2019). Wbnpmd: Weighted bipartite network projection for microRNA-disease association prediction. *J. Transl. Med.* 17, 322. doi:10.1186/s12967-019-2063-4
- Xu, J., Cai, L., Liao, B., Zhu, W., Wang, P., Meng, Y., et al. (2019). Identifying potential miRNAs-disease associations with probability matrix factorization. *Front. Genet.* 10, 1234. doi:10.3389/fgene.2019.01234
- Xuan, P., Han, K., Guo, M., Guo, Y., Li, J., Ding, J., et al. (2013). Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLoS One* 8, e70204. doi:10.1371/journal.pone.0070204
- Xuan, P., Han, K., Guo, Y., Li, J., Li, X., Zhong, Y., et al. (2015). Prediction of potential disease-associated microRNAs based on random walk. *Bioinformatics* 31, 1805–1815. doi:10.1093/bioinformatics/btv039
- Yang, M. Y., Luo, H. M., Li, Y. H., and Wang, J. X. (2019). Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics* 35, I455–I463. doi:10.1093/bioinformatics/btz331
- Yang, Z., Wu, L., Wang, A., Tang, W., Zhao, Y., Zhao, H., et al. (2017). dbDEMOC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res.* 45, D812–D818. doi:10.1093/nar/gkw1079
- You, Z. H., Huang, Z. A., Zhu, Z., Yan, G. Y., Li, Z. W., Wen, Z., et al. (2017). Pbmmda: A novel and effective path-based computational model for miRNA-disease association prediction. *PLoS Comput. Biol.* 13, e1005455. doi:10.1371/journal.pcbi.1005455
- Zeng, X., Ding, N., Rodríguez-Patón, A., Lin, Z., and Ju, Y. (2016a). Prediction of microRNA-disease associations by matrix completion. *Curr. Proteomics* 13, 151–157. doi:10.2174/157016461302160514005711
- Zeng, X. X., Wang, W., Deng, G. S., Bing, J. X., and Zou, Q. (2019). Prediction of potential disease-associated microRNAs by using neural networks. *Mol. Ther. Nucleic Acids* 16, 566–575. doi:10.1016/j.omtn.2019.04.010
- Zeng, X., Zhang, X., and Zou, Q. (2016b). Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Brief. Bioinform.* 17, 193–203. doi:10.1093/bib/bbv033
- Zhang, L., Li, X., Ke, Z., Huang, L., Liang, Y., Wu, J., et al. (2013). Mir-99a may serve as a potential oncogene in pediatric myeloid leukemia. *Cancer Cell. Int.* 13, 110. doi:10.1186/1475-2867-13-110
- Zhang, L., Liu, B., Li, Z., Zhu, X., Liang, Z., and An, J. (2020). Predicting miRNA-disease associations by multiple meta-paths fusion graph embedding model. *BMC Bioinforma.* 21, 470. doi:10.1186/s12859-020-03765-2
- Zheng, K., You, Z. H., Wang, L., Zhou, Y., Li, L. P., and Li, Z. W. (2019). Mlmda: A machine learning approach to predict and validate microRNA-disease associations by integrating of heterogeneous information sources. *J. Transl. Med.* 17, 260. doi:10.1186/s12967-019-2009-x
- Zhou, F., Yin, M. M., Jiao, C. N., Zhao, J. X., Zheng, C. H., and Liu, J. X. (2021). Predicting miRNA-disease associations through deep autoencoder with multiple kernel learning. *IEEE Trans. Neural Netw. Learn. Syst.* 2021, 1–10. doi:10.1109/TNNLS.2021.3129772