Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 20(6): 10659–10674.
DOI: 10.3934/mbe.2023471
Received: 05 October 2022
Revised: 15 March 2023
Accepted: 16 March 2023
Published: 13 April 2023

Research article

IMC-MDA: Prediction of miRNA-disease association based on induction matrix completion

Zejun Li[1], Yuxiang Zhang[2], Yuting Bai[3], Xiaohui Xie[1] and Lijun Zeng[1,*]

[1] School of Computer and Information Science, Hunan Institute of Technology, Hengyang 412002, China

[2] School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, Henan, 450001, China

[3] College of Information Science and Engineering, Hunan University, Changsha 410082, Hunan, China

* **Correspondence:** Email: zenglijun@hnit.edu.cn.

Abstract: To comprehend the etiology and pathogenesis of many illnesses, it is essential to identify disease-associated microRNAs (miRNAs). However, there are a number of challenges with current computational approaches, such as the lack of ”negative samples”, that is, confirmed irrelevant miRNA-disease pairs, and the poor performance in terms of predicting miRNAs related with “isolated diseases”, i.e. illnesses with no known associated miRNAs, which presents the need for novel computational methods. In this study, for the purpose of predicting the connection between disease and miRNA, an inductive matrix completion model was designed, referred to as IMC-MDA. In the model of IMC-MDA, for each miRNA-disease pair, the predicted marks are calculated by combining the known miRNA-disease connection with the integrated disease similarities and miRNA similarities. Based on LOOCV, IMC-MDA had an AUC of 0.8034, which shows better performance than previous methods. Furthermore, experiments have validated the prediction of disease-related miRNAs for three major human diseases: colon cancer, kidney cancer, and lung cancer.

Keywords: miRNA-disease; miRNAs; disease; matrix completion; directed acyclic graphs

1. Introduction

Small non-coding RNAs known as microRNAs (miRNAs) have a length of 20 to 25 nucleotides. miRNAs are extremely important regulatory RNAs that govern genes in a unique and irreplaceable manner [1–4]. However, numerous studies have revealed that target mRNAs may also be positively regulated by miRNAs. Additionally, there is evidence that aberrant miRNAs are linked to a variety of

human illnesses, including cancer [5, 6]. miRNAs can affect human diseases through interactions with other miRNAs, interactions between miRNAs and proteins, interactions between miRNAs and long non-coding RNAs [7, 8], and interactions between miRNA and environmental factors [9, 10]. There have been many reports on miRNA-disease research in recent years. The miR2Disease and Human miRNA Related Disease Database (HMDD) are two sweeping databases that were structured by Li et al. [11] and Jiang et al. [12] by collating experimental data that have been published to support human miRNAs and disease association. With the cause of researching the manifestation of aberrant miRNAs in various cancer diseases, Yang et al. [13] developed an miRNA database (dbDEMC) for miRNAs that are differently expressed in human cancers. As a result, the identification of illness-related miRNAs (also known as disease miRNAs) aids in understanding the molecular basis of disease as well as in preventing, diagnosing and treating diseases.

Lately, numerous academics have advocated the anticipation of disease-associated miRNAs. Experiments are costly and take plenty of time when trying to confirm miRNAs that are related with disease [14–16]. Many researchers support the evolution of puissant computational techniques to make large-scale predictions of new human miRNA-illness relationships [17–20]. At the same time, a large amount of research has produced a large amount of data on illnesses and miRNAs, which provided a firm basis for the development of computational methods. The main intention of this calculation is to forecast how disease and miRNA will interact. The primary problem of miRNA-disease association inference is the link prediction problem of heterogeneous networks [21]. Some scholars have developed a calculation method for measuring the similarity of miRNA and disease. The key similarity calculation techniques and their work for the future were summarized by Zou et al. [22]. The most popular of these computational techniques are network prediction and machine learning.

Among them, machine learning algorithms have gained widespread adoption in the field of bioinformatics. For instance, they are utilized in biological sequence prediction [23–25], circRNA-disease interaction prediction [26, 27], and ncRNA-protein interaction prediction [28, 29]. These applications have significantly advanced the study of disease-miRNA association prediction [30, 31]. Jiang et al. [32] raised a model called Naive Bayes that prioritizes disease-associated miRNAs via the integration of genomic data and they suggested a method for classification a support vector machine to distinguish the connection of negative diseases and positive miRNAs. A forecast approach that functionally improves the miRNA target imbalance network was introduced by Xu et al. [33]. Zeng et al. [34] raised two multi-path approaches for forecasting disease-related genes on the basis of gene-disease heterogeneous networks, which were utilized to forecast the connections between diseases and miRNA. In order to further distinguish the connections between diseases and miRNA, Xiao et al. [35] developed and implemented a positive matrix factorization strategy for graph regularization and achieved successful results. Unfortunately, because non-positive samples of miRNA and disease connections are difficult to obtain, many machine learning algorithms encounter bottlenecks, resulting in less than ideal prediction results [36]. In order to prioritize the detection of miRNA illness connections without using non-positive samples, the regularized least squares method was created by Chen et al. [37] for miRNA disease association. A semi-supervised classification approach termed RLSMDA predicts the association of isolated diseases.

At present, more researchers are starting to forecast the relationship between disease and miRNA by using a web-based approach [38–41]. These approaches rank the predictions, with higher rankings indicating a stronger likelihood of their relationship. Web-based approaches frequently start with the

premise that, in general, the greater the similarity of miRNAs, the more illnesses with which they are associated. Focusing this hypothesis, Gu et al. [42] suggested a network projection consistency algorithm to forecast disease-related miRNAs, fully utilizing the correlation of disease-miRNAs and the miRNAs functional similarity. A model between the miRNA-disease association predictive score and it was established by Chen et al. [43], showing that the WBSMDA can predict the association with illness in the absence of any known associated miRNA. And Liu et al. [44] integrated multiple data information to compute the similarity between miRNA and disease, and they found heterogeneous networks at the bottom of the known connections between disease and miRNA. Good results were obtained by restarting the random walk to forecast the relationship between disease and miRNAs in the network [45]. Chen et al. [46] constructed miRNA-miRNA functional network by using a generic network similarity metric, and presented a miRNA-disease association (RWRMDA) Random Walk with Restart for underlying disease-miRNA connections prediction. Unluckily, miRNA-related information is ignored when random walks are made to specific diseases. And it is impossible to forecast novel miRNAs of any unknown miRNA (isolated disease). An algorithm (referred to as HDMP) was put up by Xuan et al. [47] to forecast the connections of disease and miRNA. With the purpose of predicting potential candidate miRNAs for a certain illness, the HDMP binds to the properties and functional similarity of miRNAs. It ignores the topology formed between the neighbors and only takes into account the k neighbors in the candidate set that are most similar. A novel random walk-based prediction method was put out by Xuan et al. [48] by using various topological orderings and the properties of nodes. Recently, Chen and Zhang [49] suggested a network-based consistency-based reasoning (NetCBI) approach by using global network measurements to forecast the associations of potential disease-miRNA. NetCBI is a heterogeneous network that combines known association networks and disease-miRNA similarity networks to construct global associations that predict disease-miRNA associations. As compared to RWRMDA, NetCBI is able to forecast the disease-miRNA correlation in isolated disease, but its cross-validation performance is less effective.

As stated above, there are certain limitations to the current calculating method for forecasting disease-miRNA connection. First, some approaches based on machine learning face the challenge of collecting negative samples. Second, a number of methods are unable to forecast shielded disease-related miRNAs. Lastly, despite the fact that some approaches, like NetCBI, are able to forecast separated diseases, their cross-validation performance is subpar [49]. For the sake of addressing these challenging issues, we propose a method using the induction matrix to predict the miRNA-disease association (IMC-MDA). IMC-MDA first discovers the potential by combining the similarity network of illness semantic the miRNAs functional; then, the algorithm is finished through the matrix. The IMC-MDA algorithm has obvious advantages over other approaches.

2. Materials and methods

2.1. Data preprocessing

The association between miRNAs and disease was mainly from a human miRNA-disease association database (HMDD V2.0), which contains 5430 associated data sets, which were validated by biological wet experiments. First, we want to remove the association of the wrong name. The name of the disease was compared to the name in the International Medicine Database (http://www.nlm.nih.gov/), and the miRNA name is based on the name in miRBase 18.0 database. For example, hsa-let-7b was

not found in the miRBase 18.0. Therefore, we will delete the data. Second, there are many useless data points in the associated data provided in HMDD V2.0. For example, data associated with hsa-let-7a and neoplasms indicate that hsa-let-7a in miRNA is associated with ”tumor.” However, ”tumor” has many types, such as ”breast tumor”. Therefore, you need to delete this association. After processing, 5325 effective correlations were discovered, comprising 383 diseases and 495 miRNAs. Additionally, 383 diseases and 495 miRNAs were included in the 5325 effective connections that were found after processing.

2.2. Measuring miRNA similarity and disease similarity

2.2.1. Disease similarity

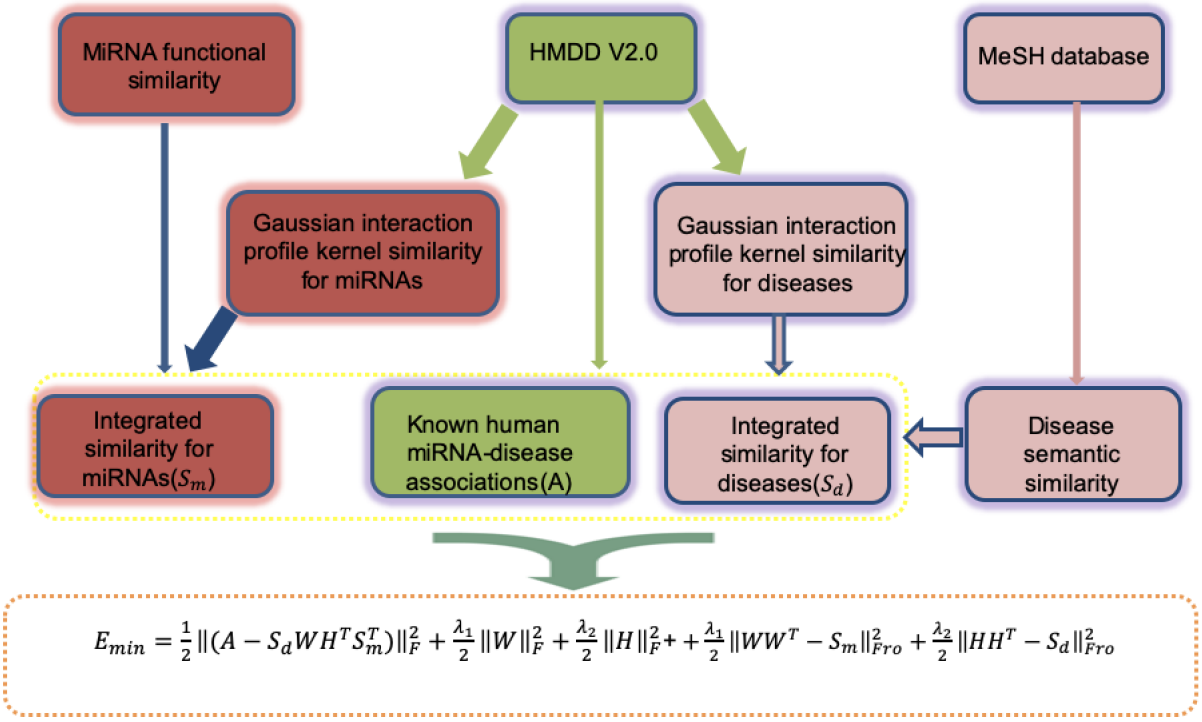

Figure 1. Workflow of IMC-MDA for discovering potential disease-miRNA associations.

In this study, we computed the similarities between illness pairs by using hierarchical directed acyclic graphs (DAGs) [50]. And the expression for a disease d's DAG map is $DAG_d = (d, T_d, E_d)$, where T_d denotes the group of disease d ancestor nodes and E_d denotes the associated link. Defining $D_d(t)$ as the disease t's semantic contribution in DAG_d to disease d is shown in Eq 2.1, where Δ is the semantic contribution element, assuming t is the t'parent node, penalized, with a range of [0,1] In this work, we applied the value of 0.5.

$$D_d(t) = \begin{cases} 1, & \text{if } t = d \\ \max\{\Delta * D_d(t') | t' \in \text{children of } t\}, & \text{if } t \neq d \end{cases} \tag{2.1}$$

Finally, semantic similarity is based on the DAG graph. Based on the hypothesis, it was assumed that the more diseases shared by the DAG maps of the two diseases, the greater the similarity and vice versa. Equation 2.2 considers the connection relationship between the ancestral nodes of the disease to compute the likeness between diseases $d1$ and $d2$.

$$D(d1, d2) = \frac{\sum_{t \in T_{d1} \cap T_{d2}} \left(D_{d1}(t) + D_{d2}(t)\right)}{\sum_{t \in T_{d1}} D_{d1}(t) + \sum_{t \in T_{d2}} D_{d2}(t)} \tag{2.2}$$

2.2.2. MiRNA functional similarity

There have been numerous studies showing that if diseases are more similar, then their associated miRNAs will be more similar, and the converse is also the conclusion. Therefore, Wang et al. [50] evaluated the likelihood between miRNAs based on known correlation data. Particularly, for two miRNAs m_a and m_b, let $DT_a = \{d_{a1}, d_{a2}, ..., d_{ak}\}$ and $DT_b = \{d_{b1}, d_{b2}, ..., d_{bl}\}$ denote lots of diseases related to m_a and m_b, respectively. Then, the likelihood between m_a and m_b is computed as follows:

$$R(m_a, m_b) = \frac{\sum_{i=1}^{k} S\left(d_{ai}, DT_b\right) + \sum_{j=1}^{l} S\left(d_{bj}, DT_a\right)}{k + l} \tag{2.3}$$

where l and k stand for, respectively, the number of illnesses in DT_b and DT_a. The similarity of miRNAs is, according to the definition, a number between zero and one.

2.2.3. Gaussian interaction profile kernel similarity of diseases and miRNAs

Similar illness-associated miRNAs ought to have more in common functionally, and vice versa. Then we computed the similarity of miRNA and disease by computing the nuclear similarity of the Gaussian interaction distribution. In order to indicate whether each miRNA has a known link to the illness d, we first use the vector $VP(d_i)$. Then, the interaction spectrum is used to compute the Gaussian interaction kernel similarity of the diseases d_j and d_i, as follows:

$$KD\left(d_i, d_j\right) = \exp\left(-\gamma_d \left\|VP\left(d_i\right) - VP\left(d_j\right)\right\|^2\right) \tag{2.4}$$

where the kernel bandwidth adjustment coefficient is γ_d, and the coefficient γ'_d needs to be updated by taking the evenness of the associations with the miRNAs for total illnesses and dividing it by the new bandwidth coefficient γ_d.

$$\gamma_d = \frac{\gamma'_d}{\frac{1}{nd}\sum_{i=1}^{nd} \|VP(d_i)\|^2} \tag{2.5}$$

Equally, it is possible to derive the kernel similarity of the Gaussian interaction distribution for miRNA according to Eqs 2.4 and 2.5.

$$KM\left(r_i, r_j\right) = \exp\left(-\gamma_m \left\|VP\left(r_i\right) - VP\left(r_j\right)\right\|^2\right) \tag{2.6}$$

$$\gamma_m = \frac{\gamma'_m}{\frac{1}{nm}\sum_{i=1}^{nm} \|VP(r_i)\|^2} \tag{2.7}$$

2.2.4. Integrated similarity for miRNAs and diseases

We constructed a matrix of illness similarity by using the semantic similarity, simultaneously, it also shares similarities in its chemical structure. Therefore, we only took the disease's semantic similarity into account, which will cause the similarity matrix's sparsity. In our work, we introduce Gaussian kernel similarity and solve sparsity by integrating Gaussian kernel similarity and semantic similarity. Therefore, the similarity matrix of disease d_i and d_j is constructed as follows:

$$S_d(d_i, d_j) = \begin{cases} D(d_i, d_j), & d_i \text{ and } d_j \text{ has semantic similarity} \\ KD(d_i, d_j), & \text{otherwise} \end{cases} \tag{2.8}$$

Equally, miRNA similarity is re-defined as follows:

$$S_m(r_i, r_j) = \begin{cases} R(r_i, r_j), & r_i \text{ and } r_j \text{ has functional similarity} \\ KM(r_i, r_j), & \text{otherwise} \end{cases} \tag{2.9}$$

2.3. Predictive disease-miRNA association based on induction matrix

We developed a new approach that is based on the induction matrix completion algorithm (IMC-MDA) to more precisely forecast the unknown relationship between diseases and miRNAs, and it entails two steps (Figure 1). First, on the basis of the converged data sources, the similarity of miRNA to disease was computed, and for a more thorough similarity, the similarities of the calculations were integrated. Second, we used our proposed matrix completion algorithm framework to infer potential associations.

2.3.1. Construction of disease-miRNA double-layer network

We constructed a bilayer network for disease-miRNA by integrating miRNA function and disease-similar networks, miRNAs and Gaussian nuclear similarity of disease with known miRNAs and diseases.

Suppose $D = \{D(i, j)_{i=1,j=1}^{n,n}\}$ is the illness similarity network matrix, $R = \{R(i, j)_{i=1,j=1}^{m,m}\}$ is the miRNA similarity network matrix and $A = \{A(i, j)_{i=1,j=1}^{m,n}\}$ is the disease-miRNA interaction network, where m and n stand for the actual amounts of miRNAs and disease respectively. Figure 2 provides a straightforward illustration of a heterogeneous network. The heterogeneous network's adjacency matrix can be expressed as below:

$$H = \begin{bmatrix} D & A \\ A^T & R \end{bmatrix} \tag{2.10}$$

A^T represents the transpose of matrix A.

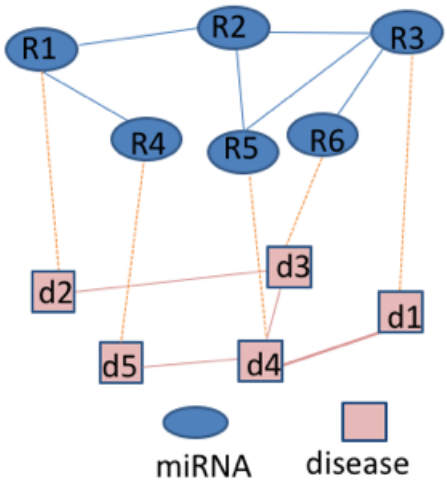

Figure 2. Heterogeneous network associated with miRNA and disease.

2.4. IMC-MDA

Currently, an efficient technology for predicting missing values in data has been extensively utilized, known as Matrix Completion. Its purpose is to find a suitable matrix to achieve an optimal approximation of the original matrix. We introduce the IMC-MDA model, which is a new induction matrix-bedded model for the prediction of the connections of disease-miRNA. On the basis of established connections, disease and miRNA similarity, the IMC-MDA model was put into practice. Here, we chose the matrix of disease similarity $S_d = R^{nd\times nd}$ and the matrix of miRNA similarity $S_m = R^{nm\times nm}$ as the characteristic matrix of the disease nd and the nm of the miRNA, and $Sm(j)$ and $Sd(i)$ denote the feature vector of the miRNA $m(j)$ disease $d(i)$ respectively. The IMC's major thought is making use of a known entry from the disease-miRNA correlation matrix A to renew the matrix $Z = R^{nd\times nm}$ in the shape $Z = WH^T$, where $H \in R^{nm\times r}$ and $W \in R^{nd\times r}$, and r is the wanted level (rank(W)), rank(H)) equal to min. The inductive matrix completion algorithm's convergence speed will be impacted by the coefficient r; however, the influence on the outcome is minimal. The parameter $Score(d(i), m(j))$ is computed to represent the expected probability of connections in disease $d(i)$ and miRNA $m(j)$. Solving the following optimization problems can yield the matrices W and H:

$$E_{min} = \frac{1}{2}\left\|A - S_d W H^T S_m^T\right\|_F^2 + \frac{\lambda_1}{2}\|W\|_F^2 + \frac{\lambda_2}{2}\|H\|_F^2 + \frac{\lambda_1}{2}\left\|WW^T - S_m\right\|_{Fro}^2 + \frac{\lambda_2}{2}\left\|HH^T - S_d\right\|_{Fro}^2 \quad (2.11)$$

where λ_1 and λ_2 are regularization parameters which balance the tracking norm constraints and the observed losses of the entries. In the experiment, we set $\lambda_1 = \lambda_2 = 1$ and the matrix's Frobenius norm is $\|\cdot\|_F$. $\frac{\lambda_1}{2}\|W\|_F^2$ and $\frac{\lambda_2}{2}\|H\|_F^2$ prevent over fitting problems, and we are able to employ the method proposed by Jain and Dhillon to address the minimum issues [51]. W and H are first created as random dense matrices. And the iterative equation is then used to update W and H. When the convergence requirement is met, the iterative procedure ought to come to an end. Generally, the convergence criteria is set to 10^6. The iterative equation in the flowchart shown in Figure 1 provides the stage of the detailed algorithm to address the minimum issue. The forecasted marks of miRNA $m(j)$ and disease $d(i)$ are able to be computed by using H and W as follows:

$$Score(d(i), m(j)) = S_d(i)WH^T S_m^T(j) \tag{2.12}$$

If $newd(i)$ is a novel disease for which no known miRNAs are connected, we can calculate all miRNA entries $newd(i)$ as long as we have the disease's eigenvector $Score(newd(i), j)$.

3. Results and discussion

3.1. Performance evaluation index

In order to more fairly evaluate the predictive accuracy of IMC-MDA, we implemented a cross-validation experimental framework for known miRNA-disease connections: for each illness $d(i)$, we selected every known miRNA-disease pairs (miRNA-disease pairs $(m(j) - d(i))$ used as an example) as the test samples, and the remaining pairs were treated as training samples. At first, the known miRNA-disease pair $(m(j) - d(i))$ was intentionally transformed into an unproven miRNA-disease pair. MiRNA-disease pair $d(i)$ that has not been verified is regarded as a candidate sample, and then the forecasted scores of the miRNA-disease pair $(m(j) - d(i))$ are ranked against it. The model can be deemed to successfully predict the miRNA-disease pair $(m(j) - d(i))$ if the level of the test miRNA-disease pair $(m(j) - d(i))$ is greater than a specified threshold. On the basis of the LOOCV framework, the IMC-MDA method is compared to the RLSMDA, MCMDA, RWRMDA, HDMP, Maxflow and MiRAI in this paper.

3.2. Performance on predicting miRNA-disease associations

We contrasted our method with a few basic methods in order to confirm its effectiveness. The details of the comparison algorithm are provided below. Maxflow [52]: It uses miRNA, disease similarity and the association network of disease and miRNAs. Subsequently, a mapped miRNAome-phenome network map was created by further combining the three networks (the metrics we utilized for comparison were $\alpha = 0.1, \gamma = 100, \beta = 0.6, \sigma = 10, \eta = 6$). MCMDA: MCMDA introduces a matrix completion algorithm for known disease and miRNA association matrices to predict unknown associations. RLSMDA: This method combines two disease space and miRNAs space training classifiers based on the regularized least squares algorithm (coefficients were set to $\omega = 0.9, \eta_d = \eta_m = 1$). HDMP: In order to establish a more trustworthy correlation score for unlabeled miRNAs, the k nearest neighbors and miRNA functional similarity for each miRNA were connected. Additionally, the HDMP weights miRNAs differently depending on miRNA families or clusters (the factors that we compared are $\alpha = 4, \beta = 4, k = 20$). RWRMDA: In order to predict probable disease miRNAs, Chen added random walks to the miRNA functional similarity network (parameter setting is $r = 0.2$).

Utilizing the consequences of LOOCV plots the receiver operating characteristic (ROC) curve. The true positive rate (TPR) and false positive rate (FPR) are plotted on the ROC graph's Y and X axes, respectively. Figure 3 demonstrates the ROC curve on the ground of LOOCV. The assessment metric for the model can be determined by the area under the curve (AUC). Thus, LOOCV, RWRMDA, IMC-MDA, MCMDA, RLSMDA, Maxflow and HDMP respectively obtained AUC values of 0.7891, 0.8372, 0.7718, 0.6953, 0.7774 and 0.7702. Therefore, compared to the previous model, it is easier to see an enhancement in forecasting the correlation between miRNA-disease and IMC-MDA.

In particular, the paired t-test was used to further study how the algorithms differed in their capacity

for reasoning. The outcomes of LOOCV were subjected to a paired t-test. We can see how IMC-MDA differs significantly from the previous models (MCMDA, HDMP, MRLSMDA, RWRMDA, Maxflow), as the P values were 2.14E-13, 5.27E-26, 2.7E-87, 9.31E-13 and 8.8E-20.

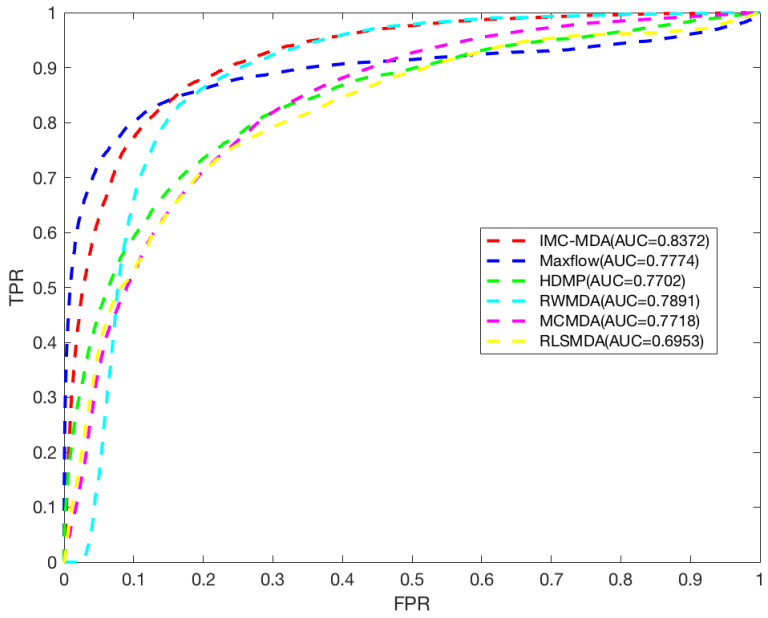

Figure 3. Comparison of IMC-MDA with five best performers for miRNA-disease associations.

3.3. Predicting novel disease-related miRNAs

On the other hand, we performed simulation tests on individual diseases with the same cross-validation approach for the purpose of assessing the effectiveness of IMC-MDA the new diseases in the absence of any known linked miRNAs. Different from cross-validation tests, all connections related to the test disease were removed during training. This operation ensures that the prediction-related candidates only use information about the remaining diseases as well as the disease and miRNA similarity information. We applied all eliminated *d*-related miRNAs as the non-negative test samples.

For the purpose of obtaining an impartial comparison, we conducted a study on six prevalent diseases that are linked to at least 80 proven connections. The main performance evaluation metric was the area under the precise recall curve (AUPR). Because MCMDA, HDMP, and RWRMDA do not predict new disease associations, we only compared two methods. The results are shown in Table 1. The mean AUPR for IMC-MDA, RLSMDA and Maxflow were 0.6353, 0.5573 and 0.5589, respectively, for the eight test diseases. IMC-MDA performed best for most of these diseases, with an average AUPR higher than those of other algorithms by 0.078 and 0.076.

3.4. Case studies

The effectiveness of IMC-MDA predictions for novel miRNA diseases was demonstrated through the use of three different kinds of case studies. They all displayed positive results. Three prevalent hu-

man diseases were considered in the first case study (colon tumors, kidney tumors, lung tumors). Three databases—dbDEMC, Phenomir and miR2Disease—were used to analyze the predicted miRNAs for these diseases. Case studies help us confirm the effectiveness of IMC-MDA even further. Then, we noted that the validated miRNAs' number was linked to the three illnesses in the top 10 and top 50, respectively, based on these two databases.

Table 1. Results of LRMCMDA and other approaches on predicting new diseases whose connections had been eliminated.

Disease name	AUPR		
	IMC-MDA	**Maxflow**	**RLSMDA**
Breast neoplasms	0.6795	0.6689	0.6749
Colorectal neoplasm	0.6327	0.5412	0.5415
Glioblastoma	0.5580	0.4537	0.4012
Heart failure	0.5728	0.5327	0.5310
Melanoma	0.7028	0.6357	0.6740
Prostatic neoplasms	0.6458	0.5023	0.5208
Stomach neoplasms	0.6351	0.6001	0.6021
Urinary bladder neoplasms	0.6309	0.5238	0.5255

In the gastrointestinal tract, the most universal malignant tumor at the moment is a colon tumor. By 2018, there were an estimated 97,220 colon tumors in the USA, of which roughly 50,630 resulted in death. However, plenty of miRNAs related to colon tumors have also been validated by some biological studies recently. For instance, when the basic expression level of colonic epithelial cells in our normal humans, the miR-106a's expression in colon tumors was lower. In colon cancer cells, it has also been proved that MiR-145 is able to down-regulate the IRS-1 protein. Hence, targeting the IRS-1-30-untranslated region prevents the development of colon carcinoma cells. IMC-MDA was used in this case study to forecast possible colon tumor-associated miRNAs. The results demonstrated that dbDEMC and miR2Disease included 10 of the top 10 forecasted colon tumor-related miRNAs (see Table 2).

According to research, approximately 3% of adult malignancies are kidney tumors, which are also one of the most pervasive types of malignant tumors of the human genitourinary system. Each year, there are more than 250,000 new instances of renal tumors that are diagnosed. Certain miRNAs may be helpful in treating kidney tumors. For instance, miR-141 is expressed at a considerably lower level in kidney tumor cells than in healthy human kidney cells. By implementing IMC-MDA, we demonstrated potential kidney tumor-associated miRNAs. As a consequence, dbDEMC or miR2Disease confirmed nine of the top 10 candidates for kidney tumor-associated miRNAs (see Table 2).

Lung cancer is one of the world's most threatened human life cancers. Its cancer incidence and mortality rate are among the highest for all cancers worldwide. Moreover, there is a clear upward trend in many countries every year. Among them, the male mortality rate for lung cancer ranks first among all cancers, and the female rank is also very high, ranking second. Previous studies have revealed that lung cancer is firmly correlated with many miRNAs. Moreover, by means of biological experiments, more than 120 relevant miRNAs have been identified. Through the IMC-MDA algorithm for those miRNAs that are unknown but may be related to lung cancer, it can be clearly seen that PhenomiR

and dbDEMC have identified all top 10 candidates (see Table 2); the results also fully confirm the effectiveness of the IMC-MDA algorithm.

Table 2. Top 10 potential miRNA candidates detected by IMC-MDA based on the databases for the three selected diseases.

Cancer	No. of miRNAs confirmed by the databases	Top 10 ranked predictions		
		Rank	miRNAs	Evidences
Conlon Neoplasms	**10**	1	hsa-mir-19b	miR2Disease
		2	hsa-mir-211b	miR2Disease
		3	hsa-mir-18a	miR2Disease
		4	hsa-mir-155	miR2Disease
		5	hsa-mir-34a	miR2Disease
		6	hsa-mir-223	miR2Disease
		7	hsa-mir-7e	dbDEMC
		8	hsa-mir-7d	dbDEMC
		9	hsa-mir-34b	dbDEMC
		10	hsa-mir-143	dbDEMC
Kidney Neoplasms	**9**	1	hsa-mir-23a	dbDEMC
		2	hsa-mir-7a	miR2Disease
		3	hsa-mir-145	dbDEMC
		4	hsa-mir-19b	miR2Disease
		5	hsa-mir-20a	miR2Disease
		6	hsa-mir-17	dbDEMC
		7	hsa-mir-200b	dbDEMC
		8	hsa-mir-7d	Unconfirmed
		9	hsa-mir-126	dbDEMC
		10	hsa-mir-19a	miR2Disease
Lung Neoplasms	**10**	1	hsa-mir-15a	PhenomiR
		2	hsa-mir-16	dbDEMC 2.0
		3	hsa-mir-429	dbDEMC 2.0
		4	hsa-mir-451a	dbDEMC 2.0
		5	hsa-mir-383	dbDEMC 2.0
		6	hsa-mir-449a	dbDEMC 2.0
		7	hsa-mir-141	dbDEMC 2.0
		8	hsa-mir-193b	dbDEMC 2.0
		9	hsa-mir-302d	dbDEMC 2.0
		10	hsa-mir-106b	PhenomiR

4. Conclusions

A growing body of evidence suggests that miRNAs are crucial in the emergence of illnesses, particularly cancers. Identifying disease-related miRNAs helps us to comprehend how diseases are triggered

as well as how to treat them.The use of network-based models and machine learning for forecasting miRNA-disease connections is widespread, but localizations still remain.The difficulty of obtaining negative training samples needs to be addressed, and the prediction accuracy has to be enhanced. In order to address these issues, we have proposed an inductive matrix completion model which was founded on the prediction of MiRNA-disease association (IMC-MDA). The IMC-MDA model combines disease similarities and miRNA similarities with known disease-miRNA connections to obtain a projected score for each miRNA-disease combination. We can clearly see that the IMC-MDA algorithm has generated more accurate prediction results than the five most sophisticated approaches.

IMC-MDA, however, has certain drawbacks. First, the similarity measure in IMC-MDA might not be the best one. We define miRNA similarity via disease and utilize disease semantic similarity. However, there exist some similarities. For instance, shared causative genes can characterize the similarity of the diseases. This approach will eventually be strengthened by merging other miRNA similarities and disease sources. Second, it is important to note that IMC-MDA only utilizes miRNAs and disease information. Since miRNAs and diseases are linked to other molecular components, such as proteins, miRNA and long non-coding RNA, to further enhance the prediction efficiency, it would be intriguing to incorporate this additional information.

Acknowledgments

This study was supported by the Postdoctoral Science Foundation of China (No. 2020M672487), the National Nature Science Foundation of China (Grant No. 62172158), the Project of Hunan Institute of Technology (No. HQ20004), the Hunan Natural Science Foundation (No. 2021JJ40120) and the Excellent Youth Project of Hunan Provincial Education Department (No. 21B0802).

Conflict of interest

The authors assert that no conflict of interest exists.

References

1. G. Meister, T. Tuschl, Mechanisms of gene silencing by double-stranded RNA, emphNature, **431** (2004), 343–349. https://doi.org/10.1038/nature02873
2. S. M. Hammond, An overview of microRNAs, *Adv. Drug Deliv. Rev.*, **87** (2015), 3–14. https://doi.org/10.1016/j.addr.2015.05.001
3. S. Rajasekaran, D. Pattarayan, P. Rajaguru, P. S. Gandhi, R. K. Thimmulappa, MicroRNA Regulation of Acute Lung Injury and Acute Respiratory Distress Syndrome, *J. Cell. Physiol.*, **231** (2016), 2097–2106. https://doi.org/10.1002/jcp.25316
4. Y. Meng, C. Lu, M. Jin, J. Xu, X. Zeng, J. Yang, A weighted bilinear neural collaborative filtering approach for drug repositioning, *Brief. Bioinformatics*, **2** (2022), bbab581. https://doi.org/10.1093/bib/bbab581
5. Y. W. Kong, D. Ferland-McCollough, T. J. Jackson, M. Bushell, microRNAs in cancer management, *Lancet Oncol.*, **13** (2012), e249–e258. https://doi.org/10.1016/S1470-2045(12)70073-6

6. M. Chen, Y. Zhang, A. Li, Z. Li, W. Liu, Z. Chen, Bipartite heterogeneous network method based on co-neighbor for MiRNA-disease association prediction, *Front. Genet.*, **10** (2019), 385. https://doi.org/10.3389/fgene.2019.00385
7. L. Cai, M. Gao, X. Ren, X. Fu, J. Xu, P. Wang, et al., MILNP: Plant lncRNA-miRNA Interaction Prediction Based on Improved Linear Neighborhood Similarity and Label Propagation, *Front. Plant Sci.*, **7** (2017), page 637. https://doi.org/10.3389/fpls.2022.861886
8. L. Zhuo, S. Pan, J. Li, X. Fu Predicting miRNA-lncRNA interactions on plant datasets based on bipartite network embedding method, **207** (2022), 97–102. https://doi.org/10.1016/j.ymeth.2022.09.002
9. L. Peng, Y. Tu, L. Huang, Y. Li, X. Fu, X. Chen, DAESTB: inferring associations of small molecule–miRNA via a scalable tree boosting model based on deep autoencoder, *Briefings in Bioinformatics*, **23** (2022), bbac478. https://doi.org/10.1093/bib/bbac478
10. J. Wei, L. Zhuo, Z. Zhou, X. Lian, X. Fu, X. Yao, GCFMCL: predicting miRNA-drug sensitivity using graph collaborative filtering and multi-view contrastive learning, *Briefings in Bioinformatics*, **24** (2023), bbad247. https://doi.org/10.1093/bib/bbad247
11. Y. Li, C. Liang, K. Wong, J. Luo, Z. Zhang, Mirsynergy: detecting synergistic miRNA regulatory modules by overlapping neighbourhood expansion, *Bioinformatics*, **30** (2014), 2627–2635. https://doi.org/10.1093/bioinformatics/btu373
12. Q. Jiang, Y. Wang, Y. Hao, L. Juan, M. Teng, X. Zhang, et al., miR2Disease: a manually curated database for microRNA deregulation in human disease, *Nucleic Acids Res.*, **37** (2009), D98–D104. https://doi.org/10.1093/nar/gkn714
13. Z. Yang, F. Ren, C. Liu, S. He, G. Sun, Q. Gao, et al., dbDEMC: a database of differentially expressed miRNAs in human cancers, *BMC Genom.*, **11** (2010), 1–8. https://doi.org/10.1186/1471-2164-11-S4-S5
14. Q. Jiang, G. Wang, T. Zhang, Y. Wang, Predicting human microrna-disease associations based on support vector machine, *2010 IEEE Int. Confer. Bioinformatics Biomed.*, (2010), 467–472. https://doi.org/10.1109/BIBM.2010.5706611
15. P. Wang, W. Zhu, B. Liao, L. Cai, L. Peng, J. Yang, Predicting influenza antigenicity by matrix completion with antigen and antiserum similarity, *Front. Microbiol.*, **9** (2018), 2500. https://doi.org/10.3389/fmicb.2018.02500
16. L. Shen, F. Liu, L. Huang, G. Liu, L. Zhou, L. Peng, VDA-RWLRLS: An anti-SARS-CoV-2 drug prioritizing framework combining an unbalanced bi-random walk and Laplacian regularized least squares, *Comput. Biol. Med.*, **140** (2022), 105–119. https://doi.org/10.1016/j.compbiomed.2021.105119
17. L. Cai, C. Lu, J. Xu, Y. Meng, P. Wang, X. Fu, et al., Drug repositioning based on the heterogeneous information fusion graph convolutional network, *Brief. Bioinformatics*, **22** (2021), bbab319. https://doi.org/10.1093/bib/bbab319
18. Y. Chen, X. Fu, Z. Li, L. Peng, L. Zhuo, Prediction of lncRNA–protein interactions via the multiple information integration, *Front. Bioeng. Biotechnol.*, **9** (2021), 647113. https://doi.org/10.3389/fbioe.2021.647113

19. J. Wei, L. Zhuo, S. Pan, X. Lian, X. Yao, X. Fu, Headtailtransfer: An efficient sampling method to improve the performance of graph neural network method in predicting sparse ncRNA–protein interactions, *Comput. Biol. Med.*, **157** (2023), 106783. https://doi.org/10.1016/j.compbiomed.2023.106783

20. L. Zhuo, B. Song, Y. Liu, Z. Li, X. Fu, Predicting ncRNA–protein interactions based on dual graph convolutional network and pairwise learning, *Brief. Bioinformatics*, **23** (2022), bbac339. https://doi.org/10.1093/bib/bbac339

21. X. Zhang, X. Zeng, Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks, *Bio-inspired Comput. Model. Algorithms*, (2019), 75–105. https://doi.org/10.1142/9789813143180_0003

22. Q. Zou, J. Li, L. Song, X. Zeng, G. Wang, Similarity computation strategies in the microRNA-disease network: a survey, *Brief Funct. Genomics*, **15** (2016), 55–64. https://doi.org/10.1093/bfgp/elv024

23. L. Cai, X. Ren, X. Fu, L. Peng, M. Gao, X. Zeng, iEnhancer-XG: interpretable sequence-based enhancers and their strength predictor, *Bioinformatics*, **37** (2021), 1060–1067. https://doi.org/10.1093/bioinformatics/btaa914

24. X. Fu, L. Cai, X. Zeng, Q. Zou, StackCPPred: a stacking and pairwise energy content-based prediction of cell-penetrating peptides and their uptake efficiency, *Bioinformatics*, **36** (2020), 3028–3034. https://doi.org/10.1093/bioinformatics/btaa131

25. X. Fu, L. Ke, L. Cai, X. Chen, X. Ren, M. Gao, Improved prediction of cell-penetrating peptides via effective orchestrating amino acid composition feature representation, *IEEE Access*, **7** (2019), 163547–163555. https://doi.org/10.1109/ACCESS.2019.2952738

26. W. Liu, T. Tang, X. Lu, X. Fu, Y. Yang, L. Peng, MPCLCDA: predicting circRNA–disease associations by using automatically selected meta-path and contrastive learning, *Brief. Bioinformatics*, **24** (2023), bbad227. https://doi.org/10.1093/bib/bbad227

27. L. Peng, C. Yang, Y. Chen, W. Liu, Predicting CircRNA-disease associations via feature convolution learning with heterogeneous graph attention network, *IEEE J. Biomed. Health. Inform.*, **27** (2023), 3072–3082. https://doi.org/10.1109/JBHI.2023.3260863

28. T. Wang, W. Wang, X. Jiang, J. Mao, L. Zhuo, M. Liu, et al., ML-NPI: predicting interactions between noncoding RNA and protein based on meta-learning in a large-scale dynamic graph, *J. Chem. Inf. Model.*, **64** (2023), 2912–2920. https://doi.org/10.1021/acs.jcim.3c01238

29. Z. Zhou, Z. Du, J. Wei, L. Zhuo, S. Pan, X. Fu, et al., MHAM-NPI: Predicting ncRNA-protein interactions based on multi-head attention mechanism, *Comput. Biol. Med.*, **163** (2023), 107143. https://doi.org/10.1016/j.compbiomed.2023.107143

30. Q. Liao, X. Fu, L. Zhuo, H. Chen, An efficient model for predicting human diseases through miRNA based on multiple-types of contrastive learning, *Front. Microbiol.*, **14** (2023), 1325001. https://doi.org/10.3389/fmicb.2023.1325001

31. W. Liu, H. Lin, L. Huang, L. Peng, T. Tang, Q. Zhao, et al., Identification of miRNA–disease associations via deep forest ensemble learning based on autoencoder, *Brief. Bioinformatics*, **23** (2022), bbac104. https://doi.org/10.1093/bib/bbac104

32. Q. Jiang, G. Wang, Y. Wang, An approach for prioritizing disease-related microRNAs based on genomic data integration, *2010 3rd Int. Confer. Biomed. Eng. Inform.*, **6** (2010), 2270–2274. https://doi.org/10.1109/BMEI.2010.5639313
33. J. Xu, C. Li, J. Lv, Y. Li, Y. Xiao, T. Shao, et al., Prioritizing Candidate Disease miRNAs by Topological Features in the miRNA Target–Dysregulated Network: Case Study of Prostate Cancer, *Mol. Cancer Ther.*, **10** (2011), 1857–1866. https://doi.org/10.1158/1535-7163.MCT-11-0055
34. X. Zeng, Y. Liao, Y. Liu, Q. Zou, Prediction and validation of disease genes using HeteSim Scores, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **14** (2016), 687–695. https://doi.org/10.1109/TCBB.2016.2520947
35. Q. Xiao, J. Luo, C. Liang, J. Cai, P. Ding, A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations, *Bioinformatics*, **34** (2018), 239–248. https://doi.org/10.1093/bioinformatics/btx545
36. J. Xu, L. Cai, B. Liao, W. Zhu, P. Wang, Y. Meng, et al., Identifying potential mirnas–disease associations with probability matrix factorization, *Front. Genet.*, **10** (2019), 1234. https://doi.org/10.3389/fgene.2019.01234
37. X. Chen, G. Yan, Semi-supervised learning for potential human microRNA-disease associations inference, *Sci. Rep.*, **4** (2014), 1–10. https://doi.org/10.1038/srep05501
38. W. Liu, X. Sun, L. Yang, K. Li, Y. Yang, X. Fu, NSCGRN: a network structure control method for gene regulatory network inference, *Brief. Bioinformatics*, **23** (2022), bbac156. https://doi.org/10.1093/bib/bbac156
39. Q. Qu, X. Chen, B. Ning, X. Zhang, H. Nie, L. Zeng, et al., Prediction of miRNA-disease associations by neural network-based deep matrix factorization, *Methods*, **212** (2023), 1–9. https://doi.org/10.1016/j.ymeth.2023.02.003
40. W. Liu, Y. Yang, X. Lu, X. Fu, R. Sun, L. Yang, et al., NSRGRN: a network structure refinement method for gene regulatory network inference, *Brief. Bioinformatics*, **24** (2023), bbad129. https://doi.org/10.1093/bib/bbad129
41. L. Peng, C. Yang, L. Huang, X. Chen, X. Fu, W. Liu, RNMFLP: predicting circRNA–disease associations based on robust nonnegative matrix factorization and label propagation, *Brief. Bioinformatics*, **24** (2023), bbac155. https://doi.org/10.1093/bib/bbad155
42. C. Gu, B. Liao, X. Li, K. Li, Network consistency projection for human miRNA-disease associations inference, *Sci. Rep.*, **6** (2016), 1–10. https://doi.org/10.1038/srep36054
43. X. Chen, C. C. Yan, X. Zhang, Z. You, L. Deng, Y. Liu, et al., WBSMDA: within and between score for MiRNA-disease association prediction, *Sci. Rep.*, **6** (2016), 1–9. https://doi.org/10.1038/srep21106
44. Y. Liu, X. Zeng, Z. He, Q. Zou, Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **14** (2016), 905–915. https://doi.org/10.1109/TCBB.2016.2550432
45. A. Li, Y. Deng, Y. Tan, M. Chen, A novel mirna-disease association prediction model using dual random walk with restart and space projection federated method, *PLoS One*, **6** (2021), e0252971. https://doi.org/10.1371/journal.pone.0252971

46. X. Chen, M. Liu, G. Yan, RWRMDA: predicting novel human microRNA–disease associations, *Mol. BioSyst.*, **8** (2012), 2792–2798. https://doi.org/10.1039/c2mb25180a

47. P. Xuan, K. Han, M. Guo, Y. Guo, J. Li, J. Ding, et al., Prediction of microRNAs Associated with Human Diseases Based on Weighted k Most Similar Neighbors, *PloS One*, **8** (2013), e70204. https://doi.org/10.1371/journal.pone.0070204

48. P. Xuan, C. Sun, T. Zhang, Y. Ye, T. Shen, Y. Dong, Gradient boosting decision tree-based method for predicting interactions between target genes and drugs, *Front. Genet.*, **10** (2019), 459. https://doi.org/10.3389/fgene.2019.00459

49. H. Chen, Z. Zhang, Similarity-based methods for potential human microRNA-disease association prediction, *BMC Med. Genom.*, **6** (2013), 1–9. https://doi.org/10.1186/1755-8794-6-12

50. D. Wang, J. Wang, M. Lu, F. Song, Q. Cui, Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases, *Bioinformatics*, **26** (2010), 1644–1650. https://doi.org/10.1093/bioinformatics/btq241

51. P. Jain, I. S. Dhillon, Provable inductive matrix completion, *arXiv preprint*, (2013), `arXiv:1306.0626`.

52. D. Wang, J. Wang, M. Lu, F. Song, Q. Cui, H. Yu, et al., Large-scale prediction of microRNA-disease associations by combinatorial prioritization algorithm, *Sci. Rep.*, **7** (2017), 1–15. https://doi.org/10.1038/srep43792

AIMS Press