



MSFSP: A Novel miRNA–Disease Association Prediction Model by Federating Multiple-Similarities Fusion and Space Projection

Yi Zhang^{1†}, Min Chen^{2*†}, Xiaohui Cheng¹ and Hanyan Wei^{3*}

¹ School of Information Science and Engineering, Guilin University of Technology, Guilin, China, ² School of Computer Science and Technology, Hunan Institute of Technology, Hengyang, China, ³ School of Pharmacy, Guilin Medical University, Guilin, China

OPEN ACCESS

Edited by:

Xianwen Ren,
Peking University, China

Reviewed by:

Xing Chen,
China University of Mining and
Technology, China
Nana Guan,
Guizhou University of Finance and
Economics, China

*Correspondence:

Min Chen
chenmin@hnit.edu.cn
Hanyan Wei
why@glmc.edu.cn

† These authors share first authorship

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Genetics

Received: 28 January 2020

Accepted: 27 March 2020

Published: 30 April 2020

Citation:

Zhang Y, Chen M, Cheng X and Wei H
(2020) MSFSP: A Novel
miRNA–Disease Association
Prediction Model by Federating
Multiple-Similarities Fusion and Space
Projection. *Front. Genet.* 11:389.
doi: 10.3389/fgene.2020.00389

Growing evidences have indicated that microRNAs (miRNAs) play a significant role relating to many important bioprocesses; their mutations and disorders will cause the occurrence of various complex diseases. The prediction of miRNAs associated with underlying diseases *via* computational approaches is beneficial to identify biomarkers and discover specific medicine, which can greatly reduce the cost of diagnosis, cure, prognosis, and prevention of human diseases. However, how to further achieve a more reliable prediction of potential miRNA–disease associations with effective integration of different biological data is a challenge for researchers. In this study, we proposed a computational model by using a federated method of combined multiple-similarities fusion and space projection (MSFSP). MSFSP firstly fused the integrated disease similarity (composed of disease semantic similarity, disease functional similarity, and disease Hamming similarity) with the integrated miRNA similarity (composed of miRNA functional similarity, miRNA sequence similarity, and miRNA Hamming similarity). Secondly, it constructed the weighted network of miRNA–disease associations from the experimentally verified Boolean network of miRNA–disease associations by using similarity networks. Finally, it calculated the prediction results by weighting miRNA space projection scores and the disease space projection scores. Leave-one-out cross-validation demonstrated that MSFSP has the distinguished predictive accuracy with area under the receiver operating characteristics curve (AUC) of 0.9613 better than that of five other existing models. In case studies, the predictive ability of MSFSP was further confirmed as 96 and 98% of the top 50 predictions for prostatic neoplasms and lung neoplasms were successfully validated by experimental evidences and supporting experimental evidences were also found for 100% of the top 50 predictions for isolated diseases.

Keywords: disease similarity, miRNA similarity, multiple-similarities fusion, space projection, computational prediction model

INTRODUCTION

The microRNAs (miRNAs) widely found in eukaryotes are those non-coding RNAs of about 20–25 nucleotides (Iorio et al., 2005). Life processes such as cell growth (Fernando et al., 2012; Zhu et al., 2016), differentiation (Miska, 2005), proliferation (Cheng et al., 2005), aging (Xu et al., 2004), signal transduction (Carthew and Sontheimer, 2009), etc. have been found to be associated with miRNAs. Increasing evidences continually confirm that complex diseases in humans including cancers, Alzheimer, diabetes, and lymphoma are closely related to miRNAs. In addition, some former researches proved that miRNAs can be considered as tumor genes or tumor suppressor genes. Therefore, inferring novel miRNA–disease associations have clinical significance for various human diseases due to miRNAs' potential roles in diagnosis biomarkers and treatment targets. Massive associations have been obtained *via* traditional biotic experiments and stored in some public databases. The traditional bio-experimental methods have high precision, but whose process is complex and time-consuming (Liang et al., 2019). Predicting and ranking potential miRNA–disease associations effectively and rapidly *via* computational identification methods are extremely vital to speed up the bio-experimental validation processes as well as reduce the blindness and time consumption of bio-experiments (Chen et al., 2015c, 2019c; Zeng et al., 2016b; Peng et al., 2017a, 2020).

On the basic assumption that functionally related miRNAs tend to be associated with phenotypically similar diseases and *vice versa* (Lu et al., 2008; Bandyopadhyay et al., 2010; Wang et al., 2010), various computational identification methods have been proposed continuously (Chen et al., 2017d, 2018c; Chen and Qu, 2018). Jiang et al. (2010a) proposed a miRNA–disease association prediction model that first used the hypergeometric distribution and constructed the functionally related miRNA network through the number of shared target genes to uncover the associations between miRNAs and diseases, but it needs to integrate other bioinformatics sources to improve model performance. Jiang et al. (2010b) proposed an approach that prioritized disease-related miRNAs based on integrating genomic data. Li et al. (2011) proposed a computational framework with which to prioritize human cancer-related miRNAs; it used the functional consistency score of miRNA–target genes and cancer-related genes to measure the associations between cancer and miRNAs. Xu et al. (2014) systematically prioritized disease-specific miRNAs by using the known disease genes and context-dependent miRNA–target interactions derived from the expression data of a matched miRNA–miRNA pair. Lack of excellent predictive performance of the above-mentioned methods may be attributed to the high false positive rate of the target genes.

Li J. et al. (2014) utilized recommendation systems to predict the associations between environmental factors, miRNAs, and diseases, but these cannot predict isolated diseases (without any known associated miRNAs) and new miRNAs (without any known associated diseases). Zhang Y. et al. (2019) used bipartite network projection (LSGSP) with known associations to reconstruct the family information, miRNA similarity network,

and disease similarity network for predicting the potential miRNA–disease associations. Although LSGSP does not need negative samples, it cannot achieve good performance only with limited number of known associations. Chen et al. (2018h) proposed a bipartite recommendation algorithm to predict miRNA–disease associations (BNPMDA) that improved the prediction accuracy distinctly with the utilization of bias ratings. Chen et al. (2018b) proposed a novel information diffusion method based on network consistency (IDNC) for uncovering disease-related miRNAs. Despite not needing negative samples and simple algorithm design, too many parameters in different databases make IDNC take a long time to find the optimal values.

In recent years, some researchers have attempted to use the topological similarity of graph to predict a miRNA–disease association (Nalluri et al., 2015; Chen et al., 2016b, 2017c, 2018e; Sun et al., 2016; You et al., 2017; Zeng et al., 2018). Chen et al. (2017b) proposed the super-disease and miRNA concepts to design a novel computational model with which to infer miRNA–disease associations. Bipartite heterogeneous network method based on co-neighbor (Chen et al., 2019a), ELLPMDA of ensemble learning and link prediction (Chen et al., 2018j), and label propagation model with linear neighborhood (Li et al., 2018) were used for various types of miRNA–disease association prediction, but those did not figure out the easy way for parameter optimization. Random walk on heterogeneous network (Chen et al., 2012, 2016a, 2018a; Xuan et al., 2015; Liu et al., 2017; Luo and Xiao, 2017; Mugunga et al., 2017; Peng et al., 2018) used for inferring miRNA–disease associations has achieved excellent prediction results with global attributes, but all of their results were partial to such miRNAs that have more known associations with diseases.

Inspired by the successful application of machine learning methods in the field of bioinformatics, many researchers used supervised machine learning methods to predict a miRNA–disease association (Chen et al., 2015a,b, 2017a, 2018d,f, 2019a,b,d; Luo et al., 2017a; Xuan et al., 2018, 2019b; Wang C.-C. et al., 2019; Wang L. et al., 2019; Zhang L. et al., 2019; Zhao et al., 2019), but which need negative samples for training. Because it is hard to obtain the experimentally verified less-known miRNA–disease associations and negative samples, some semi-supervised learning approaches (such as regularized least squares) with remarkable prediction results were proposed (Chen and Huang, 2017; Chen et al., 2017c, 2018k; Peng et al., 2017b; Xu et al., 2019). Chen and Huang (2017) used Laplacian regularized sparse subspace learning for miRNA–disease association prediction (LRSSLMDA); it projected diverse statistical feature profiles into a common subspace and selected important diverse features with a L1-norm constraint. Jiang et al. (2018) proposed a novel similarity kernel fusion (SKF) method that integrated multiple-similarity kernels to construct the accurate network similarity on which to utilize Laplacian regularized least squares for potential associations inference. It can avoid to lose the initial information during the process and can eliminate some noises in integrated similarity kernels. Luo et al. (2017b) presented a semi-supervised method with Kronecker regularized least squares to predict the potential (or missing) miRNA–disease associations. However, the above

semi-supervised solutions without the need for negative samples still have the limitation in initial values setting and optimal parameters of iteration selecting. Zeng et al. (2016a), Li et al. (2017), Chen et al. (2018g), Xiao et al. (2018), Xuan et al. (2019a), Xuan et al. (2019c), and Peng et al. (2017b) utilized the matrix completion to infer the potential miRNA–disease associations. Chen et al. (2018i) uncovered the potential miRNA–disease associations through integrating low-rank matrix decomposition and the sparse learning method. Qu et al. (2019) utilized matrix decomposition and label propagation to infer potential miRNA–disease associations. Tang et al. (2019) made full use of the miRNA functional similarity, the disease semantic similarity, and a dual Laplacian regularization term to work for the matrix completion of miRNA–disease associations. Chen et al. (2020) proposed a new computational model (NCMCMMDA) that innovatively integrated neighborhood constraint with matrix completion to find out the absent miRNA–disease associations. Even though all of the above methods only needed experimentally validated miRNA–disease associations to make prediction with a good prediction effect, the optimal parameters selection still cannot be solved very well. Although computational prediction models have attracted a lot of interests in recent years and many distinguished research progresses have been achieved, the identification of the potential associations between miRNAs and diseases still remains to involve a large number of unclear and incomplete works that need to be further improved:

- (1) The prediction accuracy still needs to be enhanced further;
- (2) Isolated diseases and new miRNAs cannot be handled directly;
- (3) Similarity construction processes are not accurate enough.

Around the above limitations, a global prediction method (MSFSP) that combined with multiple-similarities fusion (MSF) attribute was proposed, which could predict the associations between all diseases (including isolated diseases) and miRNAs (including new miRNAs) without needing negative samples. MSFSP mainly consisted of the following steps:

- (1) Reconstructed disease similarity network (fused by disease semantic similarity, disease functional similarity, and disease Hamming similarity) and miRNA similarity network (fused by miRNA functional similarity, miRNA sequence similarity, and miRNA Hamming similarity);
- (2) Reconstructed miRNA–disease network *via* integrated disease similarity, miRNA similarity, and verified Boolean network of miRNA–disease associations;
- (3) Obtained the final prediction scores of miRNA–disease associations by using the space projections of reconstructed miRNA–disease network on similarities spaces.

MATERIALS AND METHODS

Known MiRNA–Disease Associations

The experimentally verified miRNA–disease associations downloaded from HMDD v2.0 (Li Y. et al., 2014) with pre-treatment were composed of 495 processed miRNAs (formed a collection of miRNAs $M = \{m_1, m_2, \dots, m_i, \dots, m_{n_m}\}$,

$n_m = 495$), 383 diseases (formed a collection of diseases $D = \{d_1, d_2, \dots, d_j, \dots, d_{n_d}\}$, $n_d = 383$), and 5,430 known miRNA–disease associations (formed a matrix $\mathbf{MD}^{n_m \times n_d}$). The element value of $\mathbf{MD}(i, j)$ in $\mathbf{MD}^{n_m \times n_d}$ is set to 1 if the miRNA node m_i ($i = 1, 2, \dots, n_m$) is associated with the disease node d_j ($j = 1, 2, \dots, n_d$); otherwise, it is set to 0.

Disease Semantic Similarity and Disease Functional Similarity

According to the description in Wang et al. (2010), disease similarities based on semantic information were denoted by matrix $\mathbf{DD}_{ss}^{n_d \times n_d}$; it can be calculated *via* utilizing the arborescence attribute of disease in the MeSH database (Lowe and Barnett, 1994) where every disease node was marked in directed acyclic graph. Two diseases have more similar phenotypes when they associate with the same genes, based on which many researchers used the disease–gene associations to calculate disease functional similarity (Luo et al., 2017b; Jiang et al., 2018). As described in detail in Jiang et al. (2018), disease functional similarities were denoted by the matrix $\mathbf{DD}_{fs}^{n_d \times n_d}$.

MiRNA Functional Similarity and MiRNA Sequence Similarity

miRNA–miRNA functional similarities were downloaded from Wang et al. (2010), and the pairwise miRNA functional similarities were denoted by the matrix $\mathbf{MM}_{fs}^{n_m \times n_m}$. The miRNA sequence similarities obtained from the miRBase database (Kozomara and Griffiths-Jones, 2013) were denoted by the matrix $\mathbf{MM}_{ss}^{n_m \times n_m}$.

Hamming Similarity

Hamming similarity for vectors is a function that measures the number of equal components, divided by the length of vectors (Charikar, 2002). It is known that diseases with similar phenotypes are often related to similar miRNAs. Thereby, we defined disease Hamming similarity (denoted by the matrix $\mathbf{DD}_{hs}^{n_d \times n_d}$), whose element value is shown as follows:

$$\mathbf{DD}_{hs}(i, j) = 1 - \frac{\sum_{k=1}^{n_m} \text{IdSim}(\mathbf{MD}(k, i), \mathbf{MD}(k, j))}{n_m} \quad (1)$$

$$\text{IdSim}(\mathbf{MD}(k, i), \mathbf{MD}(k, j)) = \begin{cases} 1, & \text{if } \mathbf{MD}(k, i) \neq \mathbf{MD}(k, j) \\ 0, & \text{if } \mathbf{MD}(k, i) = \mathbf{MD}(k, j) \end{cases} \quad (2)$$

where $\mathbf{DD}_{hs}(i, j)$ represents the Hamming similarity between disease node d_i and d_j .

Similarly, we used \mathbf{MD}^T that denoted the transposed matrix of \mathbf{MD} to define miRNA Hamming similarity (denoted by matrix $\mathbf{MM}_{hs}^{n_m \times n_m}$). The corresponding element value in $\mathbf{MM}_{hs}^{n_m \times n_m}$ is shown as follows:

$$\mathbf{MM}_{hs}(i, j) = 1 - \frac{\sum_{k=1}^{n_d} \text{IdSim}(\mathbf{MD}^T(k, i), \mathbf{MD}^T(k, j))}{n_d} \quad (3)$$

$$\text{IdSim}(\mathbf{MD}^T(k, i), \mathbf{MD}^T(k, j)) = \begin{cases} 1, & \text{if } \mathbf{MD}^T(k, i) \neq \mathbf{MD}^T(k, j) \\ 0, & \text{if } \mathbf{MD}^T(k, i) = \mathbf{MD}^T(k, j) \end{cases} \quad (4)$$

where $MM_{hs}(i, j)$ represents the Hamming similarity between miRNA nodes m_i and m_j .

Multiple-Similarities Fusion

In this section, we used similarity kernel fusion (Wang et al., 2014; Jiang et al., 2018, 2019) to integrate three miRNA similarities (miRNA functional similarities $MM_{fs}^{n_m \times n_m}$, miRNA sequence similarities $MM_{ss}^{n_m \times n_m}$, and miRNA Hamming similarities $MM_{hs}^{n_m \times n_m}$) into one matrix $MM_{is^*}^{n_m \times n_m}$ that represented integrated miRNA similarities and three disease

similarities (disease functional similarities $DD_{fs}^{n_d \times n_d}$, disease semantic similarities $DD_{ss}^{n_d \times n_d}$, and disease Hamming similarities $DD_{hs}^{n_d \times n_d}$) into one matrix $DD_{is^*}^{n_d \times n_d}$ that represented integrated disease similarities. Details on the integration are in the following discussion.

Firstly, using similar methods mentioned in Jiang et al. (2018, 2019), the corresponding sparse matrices for three miRNA similarities denoted by $MM_{sfs}^{n_m \times n_m}$, $MM_{sss}^{n_m \times n_m}$, and $MM_{shs}^{n_m \times n_m}$, respectively, were constructed and the corresponding sparse matrices for three disease similarities were denoted by $DD_{sfs}^{n_d \times n_d}$,

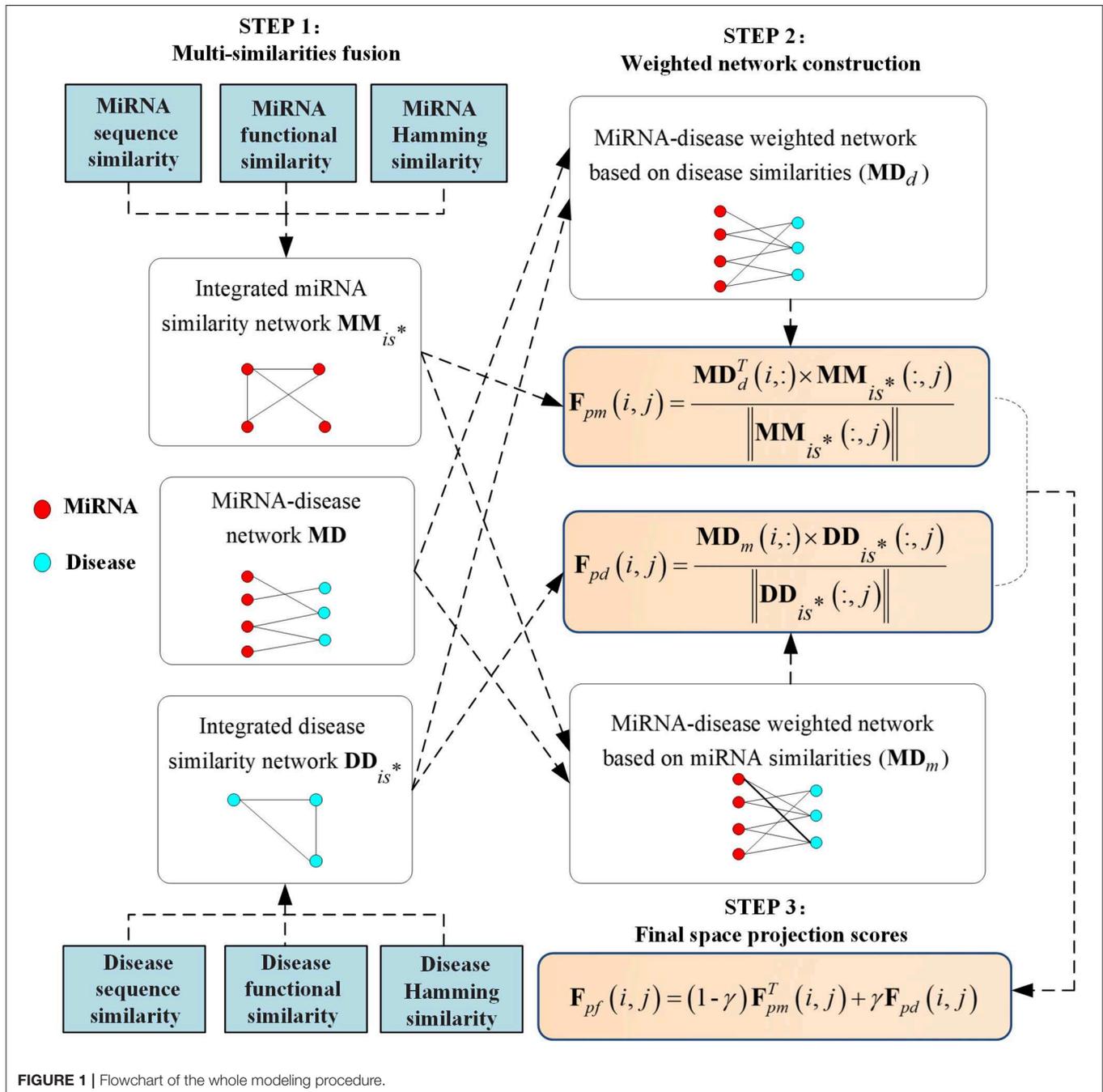


FIGURE 1 | Flowchart of the whole modeling procedure.

$DD_{sss}^{n_d \times n_d}$, and $DD_{shs}^{n_d \times n_d}$, respectively.

$$MM_{sfs}(i, j) = \begin{cases} 0, & \text{if } m_j \notin N_{m_i} \\ \frac{MM_{fs}(i, j)}{\sum_{m_k \in N_{m_i}} MM_{fs}(i, k)}, & \text{if } m_j \in N_{m_i} \end{cases} \quad (5)$$

where N_{m_i} represents the collection of all neighbors of miRNA node m_i , including m_i in the corresponding three miRNA similarities matrices (MM_{sfs} , MM_{sss} , and MM_{shs}), and the number of N_{m_i} was set to 36.

Similarly, $MM_{sss}(i, j)$, $MM_{shs}(i, j)$, $DD_{sfs}(i, j)$, $DD_{sss}(i, j)$, and $DD_{shs}(i, j)$ were constructed by using the above representation.

Secondly, the integrated normalized matrices and sparse matrices are as follows:

$$\begin{aligned} (\overline{MM}_{fs})^{t+1} = & \delta(MM_{sfs} \times \frac{(\overline{MM}_{ss})^t + (\overline{MM}_{hs})^t}{2} \times MM_{sfs}^T) \\ & + (1 - \delta) \frac{(\overline{MM}_{ss})^0 + (\overline{MM}_{hs})^0}{2} \end{aligned} \quad (6)$$

where \overline{MM}_{ss} and \overline{MM}_{hs} are the normalizations for MM_{ss} and MM_{hs} , respectively. MM_{sfs}^T denotes the transposed matrix of MM_{sfs} ; $(\overline{MM}_{ss})^t$ and $(\overline{MM}_{hs})^t$ are the t^{th} iteration results of \overline{MM}_{ss} and \overline{MM}_{hs} , respectively. t was set to 10 and δ was set to 0.1, which are similar as those defined in Jiang et al. (2018, 2019). $(\overline{MM}_{ss})^0$ and $(\overline{MM}_{hs})^0$ represented the initial status of \overline{MM}_{ss} and \overline{MM}_{hs} , respectively, with the detailed calculation shown as follows:

$$(\overline{MM}_{ss}(i, j))^0 = \frac{MM_{ss}(i, j)}{MM_{fs}(i, j) + MM_{ss}(i, j) + MM_{hs}(i, j)} \quad (7)$$

$$(\overline{MM}_{hs}(i, j))^0 = \frac{MM_{hs}(i, j)}{MM_{fs}(i, j) + MM_{ss}(i, j) + MM_{hs}(i, j)} \quad (8)$$

Furthermore, similar representations of $(\overline{MM}_{ss})^{t+1}$ and $(\overline{MM}_{hs})^{t+1}$ could be obtained as that of $(\overline{MM}_{fs})^{t+1}$. After $t + 1$ iterations, the temporarily integrated miRNA similarity denoted by matrix MM_{is} was calculated as follows:

$$MM_{is} = \frac{(\overline{MM}_{fs})^{t+1} + (\overline{MM}_{ss})^{t+1} + (\overline{MM}_{hs})^{t+1}}{3} \quad (9)$$

Thirdly, a weighted matrix $W_m^{n_m \times n_m}$ for eliminating noises during the calculation process was constructed, as mentioned in Jiang et al. (2019). Then, the finally integrated miRNA similarity denoted by matrix $MM_{is}^{n_m \times n_m}$ was obtained via taking a dot product:

$$W_m(i, j) = \begin{cases} 1, & \text{if } m_i \in N_{m_j} \text{ and } m_j \in N_{m_i} \\ 0, & \text{if } m_i \notin N_{m_j} \text{ and } m_j \notin N_{m_i} \\ 0.5, & \text{otherwise} \end{cases} \quad (10)$$

$$MM_{is}^* = MM_{is} \circ W_m \quad (11)$$

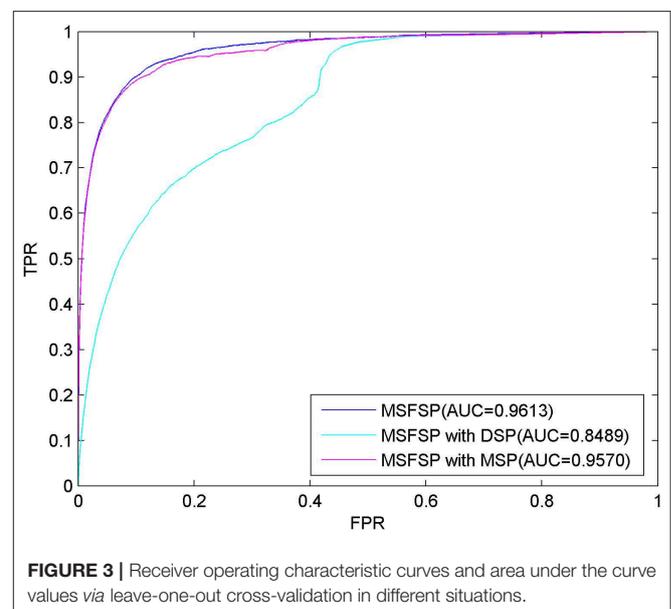
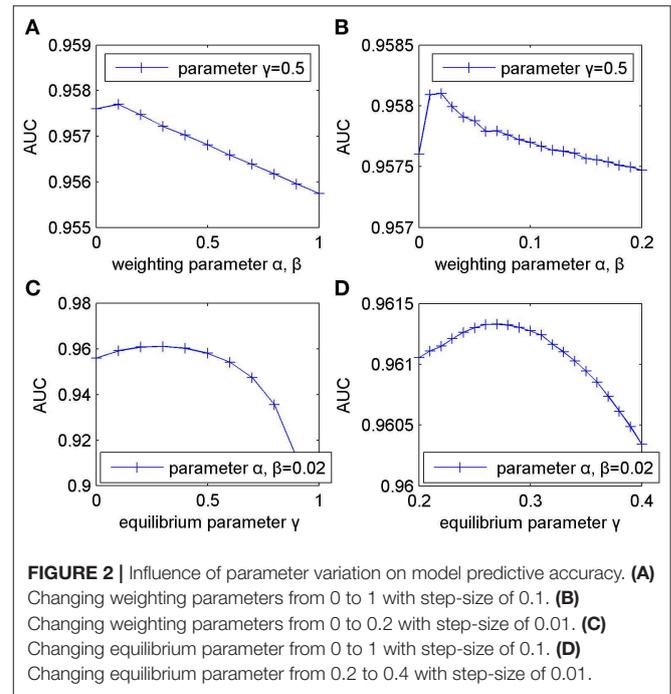
The finally integrated disease similarity matrix $DD_{is}^{n_d \times n_d}$, the weighted matrix $W_d^{n_d \times n_d}$, and the temporarily integrated disease

similarity matrix $DD_{is}^{n_d \times n_d}$ can be calculated by a similar calculation process as that of MM_{is}^* :

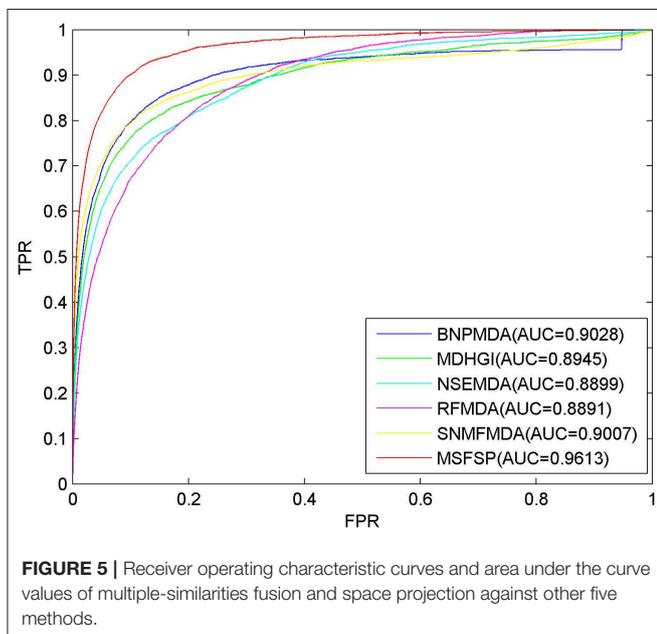
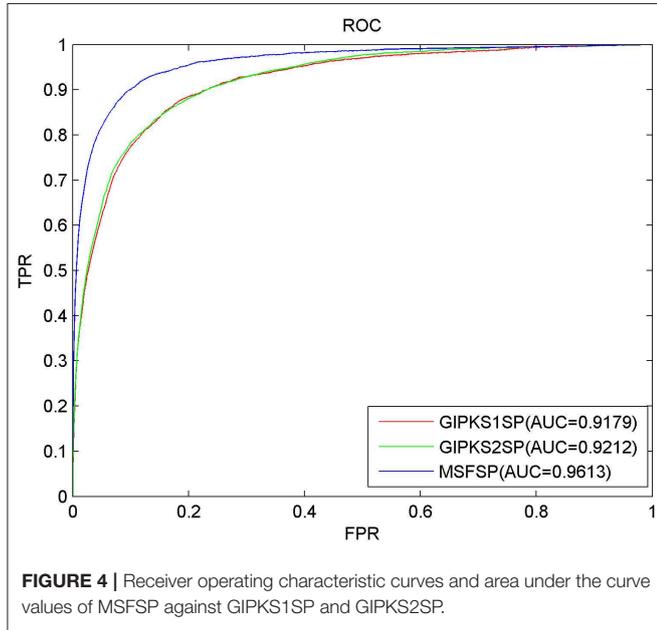
$$DD_{is}^* = DD_{is} \circ W_d \quad (12)$$

Weighted Network Construction

On account of the hypothesis that miRNAs with similar functions are often related to the diseases with similar phenotypes, many methods for miRNA–disease association prediction were proposed. Though the network MD of known experimentally



verified miRNA–disease association plays a very important role in these prediction methods, network **MD** is only a Boolean network which can indicate if the miRNA–disease association exists or not, without any information of the extent of association. Therefore, in order to enhance the predictive validity, we used **MD** and similarities between miRNAs (diseases) to accurately construct a weighted network with which to uncover potential miRNA–disease associations.



Weighted Network Construction Based on MiRNA Similarities

The contribution value of the other miRNA node $m_k (k \neq i)$ to m_i (denoted by C_{m_k}) was defined as follows:

$$C_{m_k} = MM_{is^*}(i, k) \times MD(k, j) \tag{13}$$

where $MM_{is^*}(i, k)$ is the finally integrated miRNA similarity between m_i and m_k , and $MD(k, j)$ represents the Boolean value of the association between m_k and d_j .

If there is an association between m_k and d_j , the more similar m_k and m_i are, the higher the contribution value of m_k to the weight between m_i and d_j . Based on the discussion above, the miRNA–disease weighted network based on miRNA similarities (denoted by $MD_m^{n_m \times n_d}$) was defined as follows:

$$MD_m(i, j) = MD(i, j) + \alpha \sum_{k=1, k \neq i}^{n_m} C_{m_k} \tag{14}$$

where $MD_m(i, j)$ is the weight between miRNA node m_i and disease node d_j , the equilibrium parameter being $\alpha \in [0, 1]$.

Weighted Network Construction Based on Disease Similarities

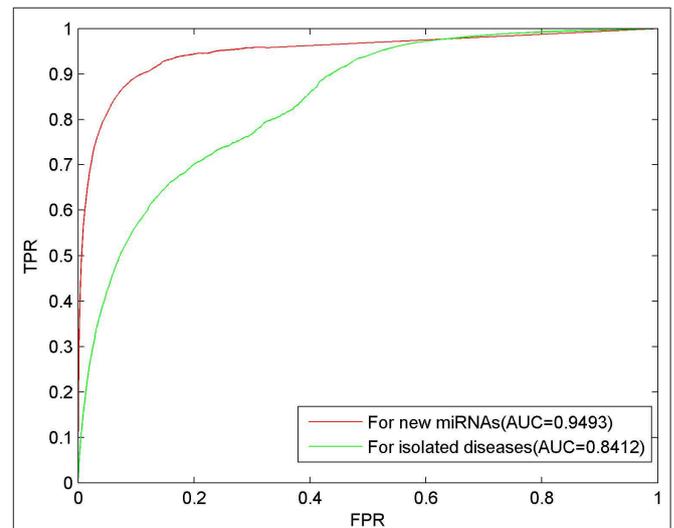
Similarly, the contribution value of the other disease nodes $d_k (k \neq i)$ to d_i (denoted by C_{d_k}) was defined as follows:

$$C_{d_k} = MD(i, k) \times DD_{is^*}(k, j) \tag{15}$$

miRNA–disease weighted network based on disease similarities (denoted by $MD_d^{n_m \times n_d}$) was defined as follows:

$$MD_d(i, j) = MD(i, j) + \beta \sum_{k=1, k \neq j}^{n_d} C_{d_k} \tag{16}$$

where equilibrium parameter $\beta \in [0, 1]$.



Space Projection Scores Based on Similarities

To enhance the predictive accuracy further, we integrated $MD_d^{n_m \times n_d}$ and $MM_{is^*}^{n_m \times n_m}$ to construct miRNA space projection scores denoted by matrix $F_{pm}^{n_d \times n_m}$, shown as follows:

$$F_{pm}(i, j) = \frac{MD_d^T(i, :) \times MM_{is^*}(:, j)}{\|MM_{is^*}(:, j)\|} \quad (17)$$

where MD_d^T is the transposed matrix of MD_d , and $\|MM_{is^*}(:, j)\|$ is the norm of vector $MM_{is^*}(:, j)$.

Similarly, we integrated $MD_m^{n_m \times n_d}$ and $DD_{is^*}^{n_d \times n_d}$ to construct disease space projection scores denoted by matrix $F_{pd}^{n_m \times n_d}$, shown as follows:

$$F_{pd}(i, j) = \frac{MD_m(i, :) \times DD_{is^*}(:, j)}{\|DD_{is^*}(:, j)\|} \quad (18)$$

where $\|DD_{is^*}(:, j)\|$ is the norm of vector $DD_{is^*}(:, j)$.

Finally, we integrated $F_{pm}(i, j)$ and $F_{pd}(i, j)$ to obtain the final prediction score $F_{pf}(i, j)$, shown as follows:

$$F_{pf}(i, j) = (1 - \gamma)F_{pm}^T(i, j) + \gamma F_{pd}(i, j) \quad (19)$$

where F_{pm}^T is the transposed of matrix of F_{pm} , and the equilibrium parameter $\gamma \in [0, 1]$ represents the importance degree of $F_{pm}(i, j)$ and $F_{pd}(i, j)$.

Therefore, we will integrate disease similarities, miRNA similarities, and weighted networks to obtain the final prediction scores $F_{pf}^{n_m \times n_d}$, whose higher value means a higher probability that miRNA m_i associates with disease d_j . The detailed calculation steps of F_{pf} are shown in **Figure 1** for clarity.

RESULTS

Influence of Parameter Selection on Performance

This section mainly discussed the influences of different types of parameters (weighting parameter α, β and equilibrium parameter γ) on the predictive performance of MSFSP. For simplicity, we set α and β to be of the same value.

Firstly, we fixed γ to 0.5 and changed α and β from 0 to 1 with a step-size of 0.1. After performing LOOCV, the results showed that AUC reached an optimal value of 0.9577 when α and β were set to 0.1. Then, the AUC values decreased gradually when α and β increased from 0.1 to 1, which caused the corresponding curve to decline linearly (shown in **Figure 2**). Therefore, α and β should range from 0 to 0.2 to get the optimal value.

TABLE 1 | Top 50 lung neoplasm-related miRNAs.

Rank	MiRNA name	Database of evidence	Rank	MiRNA name	Database of evidence
1	hsa-mir-16	HMDD, dbDEMCC, miR2Disease	26	hsa-mir-668	dbDEMCC
2	hsa-mir-106b	HMDD, dbDEMCC	27	hsa-mir-208a	HMDD
3	hsa-mir-15a	HMDD, dbDEMCC	28	hsa-mir-708	dbDEMCC
4	hsa-mir-141	HMDD, dbDEMCC, miR2Disease	29	hsa-mir-663b	dbDEMCC
5	hsa-mir-15b	dbDEMCC	30	hsa-mir-196b	HMDD, dbDEMCC
6	hsa-mir-194	HMDD, dbDEMCC	31	hsa-mir-328	HMDD, dbDEMCC
7	hsa-mir-130a	HMDD, dbDEMCC, miR2Disease	32	hsa-mir-342	HMDD, dbDEMCC
8	hsa-mir-151a	dbDEMCC	33	hsa-mir-149	HMDD, dbDEMCC
9	hsa-mir-429	dbDEMCC, miR2Disease	34	hsa-mir-1236	dbDEMCC
10	hsa-mir-99a	HMDD, dbDEMCC, miR2Disease	35	hsa-mir-320a	dbDEMCC
11	hsa-mir-122	HMDD, dbDEMCC	36	hsa-mir-370	dbDEMCC
12	hsa-mir-195	HMDD, dbDEMCC, miR2Disease	37	hsa-mir-181d	dbDEMCC
13	hsa-mir-20b	dbDEMCC	38	hsa-mir-144	HMDD, dbDEMCC
14	hsa-mir-193b	dbDEMCC	39	hsa-mir-302b	dbDEMCC
15	hsa-mir-378a	dbDEMCC	40	hsa-mir-363	dbDEMCC
16	hsa-mir-129	HMDD, dbDEMCC	41	hsa-mir-424	dbDEMCC
17	hsa-mir-153	HMDD, dbDEMCC	42	hsa-mir-130b	HMDD, dbDEMCC
18	hsa-mir-451a	HMDD, dbDEMCC	43	hsa-mir-373	HMDD, dbDEMCC
19	hsa-mir-10a	HMDD, dbDEMCC	44	hsa-mir-204	dbDEMCC, miR2Disease
20	hsa-mir-28	dbDEMCC	45	hsa-mir-211	dbDEMCC
21	hsa-mir-92b	dbDEMCC	46	hsa-mir-139	HMDD, dbDEMCC, miR2Disease
22	hsa-mir-625	dbDEMCC	47	hsa-mir-367	dbDEMCC
23	hsa-mir-152	HMDD, dbDEMCC	48	hsa-mir-384	Unconfirmed
24	hsa-mir-296	dbDEMCC	49	hsa-mir-148b	HMDD, dbDEMCC
25	hsa-mir-23b	dbDEMCC	50	hsa-mir-423	HMDD, dbDEMCC, miR2Disease

Next, in order to get more accurate weighting parameters, we fixed γ to 0.5 again and changed α and β from 0 to 0.2 with a step-size of 0.01. The corresponding changing curve is shown in **Figure 2**, where the optimal AUC of 0.9581 was obtained when α and β were both 0.02.

Then, based on $\alpha = \beta = 0.02$, we evaluated the influence of γ on MSFSP in a similar way as detailed above. We increased γ from 0 to 1 with a step-size of 0.1 to obtain the corresponding results shown in **Figure 2**, where the optimal, suboptimal, and third-best value of AUC were obtained when γ was 0.3, 0.2, and 0.4, respectively. However, AUC decreased when γ increased from 0.4. Therefore, γ should range from 0.2 to 0.4 to get the optimal value. We increased γ from 0.2 to 0.4 with a step-size of 0.01 to get more accurate parameter values with α and β fixed to 0.02. The changing curve in **Figure 2** shows the optimal value of 0.9613 when γ was 0.27.

In conclusion, our parameter selections were $\alpha = \beta = 0.02$ and $\gamma = 0.27$.

Comparison of Predictive Performance Under Different Situations

We performed LOOCV to evaluate the predictive performance of MSFSP under the following different situations: (1) with all relevant information (MSFSP with all), (2) only with miRNA space projection (MSFSP with MSP), and (3) only with disease

space projection (MSFSP with DSP). The ROC curves for the above different situations are shown in **Figure 3**, where the AUC value of MSFSP with all was 0.9613, the AUC value of MSFSP with MSP was 0.9570, and the AUC value of MSFSP with DSP was 0.8489. Therefore, MSFSP showed reliable predictive performance for inferring miRNA–disease associations effectively.

Comparison of Predictive Performance With Different Integrated Similarity Constructions

Concerning the limitations of the sparsity and incompleteness existing in disease semantic similarity and miRNA functional similarity, we used MSF in MSFSP to construct the integrated disease similarity and the integrated miRNA similarity with which to solve these limitations. Some other researchers integrated disease semantic similarity (miRNA functional similarity) with Gaussian interaction profile kernel similarity to construct the integrated diseases similarity (the integrated miRNA similarity) with which to solve the same limitations (Chen et al., 2016c, 2017d; Chen and Huang, 2017; Zhao et al., 2018, 2019). In order to compare which of the two ways wherein integrated similarities were constructed has better predictive result, we compared MSF used in MSFSP with Gaussian interaction profile kernel similarity used in Chen and Huang

TABLE 2 | Top 50 prostatic neoplasm-related miRNAs.

Rank	MiRNA name	Database of evidence	Rank	MiRNA name	Database of evidence
1	hsa-mir-29c	HMDD, dbDEMCC	26	hsa-mir-1229	dbDEMCC
2	hsa-mir-10b	dbDEMCC, miR2Disease	27	hsa-mir-944	dbDEMCC
3	hsa-mir-429	HMDD	28	hsa-mir-1227	HMDD, dbDEMCC
4	hsa-mir-19a	HMDD, dbDEMCC	29	hsa-mir-451a	dbDEMCC
5	hsa-mir-155	HMDD, dbDEMCC	30	hsa-mir-139	HMDD, dbDEMCC
6	hsa-mir-181a	HMDD, dbDEMCC, miR2Disease	31	hsa-mir-625	dbDEMCC
7	hsa-mir-210	HMDD, dbDEMCC, miR2Disease	32	hsa-mir-150	HMDD, dbDEMCC
8	hsa-mir-199b	HMDD, dbDEMCC, miR2Disease	33	hsa-mir-128	HMDD, dbDEMCC
9	hsa-mir-19b	HMDD, dbDEMCC, miR2Disease	34	hsa-mir-370	HMDD, dbDEMCC, miR2Disease
10	hsa-mir-18a	HMDD, dbDEMCC	35	hsa-mir-18b	dbDEMCC
11	hsa-mir-142	dbDEMCC	36	hsa-mir-28	dbDEMCC
12	hsa-mir-9	HMDD, dbDEMCC	37	hsa-mir-135a	HMDD, dbDEMCC
13	hsa-mir-192	HMDD, dbDEMCC	38	hsa-mir-10a	HMDD, dbDEMCC, miR2Disease
14	hsa-mir-125a	dbDEMCC, miR2Disease	39	hsa-mir-149	HMDD, dbDEMCC, miR2Disease
15	hsa-let-7f	dbDEMCC, miR2Disease	40	hsa-mir-140	dbDEMCC
16	hsa-mir-24	HMDD, dbDEMCC, miR2Disease	41	hsa-mir-20b	HMDD, dbDEMCC
17	hsa-let-7i	dbDEMCC	42	hsa-mir-302b	dbDEMCC
18	hsa-let-7e	dbDEMCC	43	hsa-mir-328	dbDEMCC
19	hsa-let-7g	HMDD, dbDEMCC, miR2Disease	44	hsa-mir-30e	dbDEMCC
20	hsa-mir-7	HMDD, dbDEMCC	45	hsa-mir-103b	dbDEMCC
21	hsa-mir-196a	HMDD, dbDEMCC	46	hsa-mir-633	Unconfirmed
22	hsa-mir-206	HMDD, dbDEMCC	47	hsa-mir-300	Unconfirmed
23	hsa-mir-138	HMDD	48	hsa-mir-30b	dbDEMCC, miR2Disease
24	hsa-mir-30a	HMDD, dbDEMCC, miR2Disease	49	hsa-mir-497	HMDD, dbDEMCC, miR2Disease
25	hsa-mir-103a	dbDEMCC	50	hsa-mir-663b	dbDEMCC

(2017) and Zhao et al. (2019). We used the Gaussian interaction profile kernel similarity coming from Chen and Huang (2017) to replace MSF in MSFSP, and we got a new prediction model called GIPKS1SP as one object to be compared. Similarly, we used Gaussian interaction profile kernel similarity coming from Zhao et al. (2019) to replace MSF in MSFSP, and we got another new prediction model called GIPKS2SP as another object to be compared. After performing LOOCV, the AUCs of GIPKS1SP, GIPKS2SP, and MSFSP were 0.9179, 0.9212, and 0.9613, respectively (shown in **Figure 4**). The more reliable predictive performance obtained *via* MSFSP proved that MSF is better than the Gaussian interaction profile kernel similarity to construct the integrated similarities.

Comparison to Other Methods

To our knowledge, BNPMDA (Chen et al., 2018h), MDHGI (Chen et al., 2018i), NSEMDA (Wang C.-C. et al., 2019), RFMDA (Chen et al., 2018f), and SNMFMDA (Zhao et al., 2018) are the most advanced prediction methods in inferring miRNA–disease associations so far. Due to the fact that the databases used by these five methods are similar with that of MSFSP, we compared MSFSP with these five methods on the predictive performance. The LOOCV results in **Figure 5** show that the AUC values of BNPMDA, MDHGI, NSEMDA, RFMDA, SNMFMDA, and MSFSP were 0.9028, 0.8945,

0.8899, 0.8891, 0.9007, and 0.9613, respectively. MSFSP achieved the superior prediction effect, at 6.09, 6.94, 7.42, 7.51, and 6.30% higher than BNPMDA, MDHGI, NSEMDA, RFMDA, and SNMFMDA, respectively.

Prediction of New MiRNAs and Isolated Diseases

With the continuously developing miRNA recognition technology, more and more miRNAs are being discovered, but whose associations with diseases are unknown. The prediction for isolated diseases and new miRNAs will definitely accelerate the scientists' understanding of the molecular mechanisms of diseases as well as how diseases occur. Therefore, the prediction for isolated diseases and new miRNAs has become a hot research topic in recent years.

For each miRNA, we removed all related associations with diseases to simulate the new miRNA. For each disease, we removed all related associations with miRNAs to simulate the isolated disease. Through LOOCV, the prediction results shown as AUC of 0.9493 and 0.8412, respectively, were obtained, where the ROC curve demonstrated the excellent predictive performance of MSFSP on inferring new miRNA-related diseases, as well as isolated diseases related with miRNAs (as can be seen in **Figure 6**).

TABLE 3 | Top 50 isolated disease-related miRNAs (prostatic neoplasm as a case).

Rank	MiRNA name	Database of evidence	Rank	MiRNA name	Database of evidence
1	hsa-mir-125b	HMDD, dbDEMC, miR2Disease	26	hsa-let-7a	HMDD, dbDEMC, miR2Disease
2	hsa-mir-21	HMDD, dbDEMC, miR2Disease	27	hsa-mir-92a	HMDD
3	hsa-mir-145	HMDD, dbDEMC, miR2Disease	28	hsa-mir-143	HMDD, miR2Disease
4	hsa-mir-99a	HMDD, dbDEMC, miR2Disease	29	hsa-mir-133b	HMDD, dbDEMC
5	hsa-mir-200c	HMDD, dbDEMC	30	hsa-mir-18a	HMDD, dbDEMC
6	hsa-mir-155	HMDD, dbDEMC	31	hsa-mir-146b	HMDD, dbDEMC
7	hsa-mir-141	HMDD, dbDEMC, miR2Disease	32	hsa-let-7g	HMDD, dbDEMC, miR2Disease
8	hsa-mir-200a	HMDD, dbDEMC	33	hsa-mir-218	HMDD, dbDEMC, miR2Disease
9	hsa-mir-183	HMDD, dbDEMC, miR2Disease	34	hsa-let-7c	HMDD, dbDEMC, miR2Disease
10	hsa-mir-100	HMDD, dbDEMC, miR2Disease	35	hsa-let-7i	dbDEMC
11	hsa-mir-9	dbDEMC	36	hsa-let-7f	dbDEMC, miR2Disease
12	hsa-mir-199a	HMDD, dbDEMC, miR2Disease	37	hsa-let-7d	HMDD, dbDEMC, miR2Disease
13	hsa-mir-34c	HMDD, dbDEMC	38	hsa-mir-7	HMDD, dbDEMC
14	hsa-mir-126	HMDD, dbDEMC, miR2Disease	39	hsa-mir-203	HMDD, dbDEMC
15	hsa-mir-29c	HMDD, dbDEMC	40	hsa-mir-1	HMDD, dbDEMC
16	hsa-mir-20a	HMDD, dbDEMC, miR2Disease	41	hsa-mir-574	HMDD, dbDEMC
17	hsa-mir-17	HMDD, dbDEMC, miR2Disease	42	hsa-let-7e	dbDEMC
18	hsa-mir-19a	HMDD, dbDEMC	43	hsa-mir-34b	HMDD, dbDEMC
19	hsa-mir-146a	HMDD, dbDEMC, miR2Disease	44	hsa-mir-101	HMDD, dbDEMC, miR2Disease
20	hsa-mir-200b	HMDD, dbDEMC	45	hsa-mir-19b	HMDD, dbDEMC, miR2Disease
21	hsa-mir-27a	HMDD, dbDEMC, miR2Disease	46	hsa-mir-10b	dbDEMC, miR2Disease
22	hsa-mir-34a	HMDD, dbDEMC, miR2Disease	47	hsa-mir-375	HMDD, dbDEMC, miR2Disease
23	hsa-let-7b	HMDD, dbDEMC, miR2Disease	48	hsa-mir-182	HMDD, dbDEMC, miR2Disease
24	hsa-mir-429	HMDD	49	hsa-mir-221	HMDD, dbDEMC, miR2Disease
25	hsa-mir-205	HMDD, miR2Disease	50	hsa-mir-142	dbDEMC

CASE STUDIES

To further evaluate the predictive ability of MSFSP on inferring diseases potentially related to miRNAs, we selected prostatic neoplasms and lung neoplasms as the case studies with model training and predicting on three independent databases HMDD v3.2 (Huang et al., 2018), dbDEMC 2.0 (Yang et al., 2017), and miR2Disease (Jiang et al., 2009).

Prediction of Potential MiRNA–Disease Associations

The low detection rate of lung neoplasm, making it a common lethal disease, poses a great threat to people's lives especially in developing countries (Torre et al., 2016; Temraz et al., 2017). We used 132 known associations between lung neoplasms and miRNAs as training samples to predict the remaining unknown associations. We found the supporting evidences that 49 out of all the first 50 miRNAs related to lung neoplasms predicted by MSFSP were confirmed on the above-mentioned three databases (HMDD v3.2, dbDEMC 2.0, and miR2Disease), except hsa-mir-384 (as shown in **Table 1**). However, we found the association between hsa-mir-384 and lung neoplasms by searching the latest literature (Guo et al., 2019) whose publication date was after the last update of HMDD v3.2, which further confirmed the

effectiveness of MSFSP for inferring diseases potentially related to miRNAs.

Prostatic neoplasm is a disease occurring in the male reproductive system, especially common in countries with severely aging population, but in recent years, more and more prostatic neoplasms occur in young people (Siegel et al., 2016). We used 118 known associations between prostatic neoplasms and miRNAs as training samples to predict the remaining unknown associations. A total of 48 out of the first 50 miRNAs related to prostatic neoplasms predicted by MSFSP were confirmed on relevant databases (HMDD v3.2, dbDEMC 2.0, and miR2Disease), except hsa-mir-633 and hsa-mir-300 (ranked 46th and 47th, respectively) (as shown in **Table 2**). Although there is no evidence that shows the association between these two miRNAs and prostatic neoplasms by now, we believe that some evidences will be found by scientists in the near future.

Prediction of Isolated Disease-Related MiRNAs

To further evaluate the predictive performance of MSFSP for isolated diseases which are those without any known associations, we removed all 118 known associations related to prostatic neoplasms to simulate the isolated disease condition. The supporting evidences for the top 50 prostatic neoplasm-related miRNAs predicted were all found from the relevant

TABLE 4 | Top 50 isolated disease-related miRNAs (lung neoplasm as a case).

Rank	MiRNA name	Database of evidence	Rank	MiRNA name	Database of evidence
1	hsa-mir-21	HMDD, dbDEMC, miR2Disease	26	hsa-let-7g	HMDD, dbDEMC, miR2Disease
2	hsa-mir-125b	HMDD, dbDEMC, miR2Disease	27	hsa-mir-148a	HMDD, dbDEMC, miR2Disease
3	hsa-mir-155	HMDD, dbDEMC, miR2Disease	28	hsa-let-7d	HMDD, dbDEMC, miR2Disease
4	hsa-mir-34a	HMDD, dbDEMC	29	hsa-mir-101	HMDD, dbDEMC, miR2Disease
5	hsa-mir-375	HMDD, dbDEMC	30	hsa-mir-205	HMDD, dbDEMC, miR2Disease
6	hsa-mir-146a	HMDD, dbDEMC, miR2Disease	31	hsa-let-7e	HMDD, dbDEMC, miR2Disease
7	hsa-mir-1	HMDD, dbDEMC	32	hsa-mir-93	HMDD, dbDEMC, miR2Disease
8	hsa-mir-31	HMDD, dbDEMC, miR2Disease	33	hsa-mir-143	HMDD, dbDEMC, miR2Disease
9	hsa-mir-34c	HMDD, dbDEMC	34	hsa-mir-17	HMDD, dbDEMC, miR2Disease
10	hsa-mir-145	HMDD, dbDEMC, miR2Disease	35	hsa-mir-183	HMDD, dbDEMC, miR2Disease
11	hsa-let-7a	HMDD, dbDEMC, miR2Disease	36	hsa-mir-20a	HMDD, dbDEMC, miR2Disease
12	hsa-mir-221	HMDD, dbDEMC, miR2Disease	37	hsa-mir-200b	HMDD, dbDEMC, miR2Disease
13	hsa-mir-486	HMDD, dbDEMC	38	hsa-mir-133a	HMDD, dbDEMC
14	hsa-mir-100	HMDD, dbDEMC	39	hsa-mir-193b	dbDEMC
15	hsa-mir-16	HMDD, dbDEMC, miR2Disease	40	hsa-mir-27a	HMDD, dbDEMC
16	hsa-mir-126	HMDD, dbDEMC, miR2Disease	41	hsa-let-7c	HMDD, dbDEMC, miR2Disease
17	hsa-let-7b	HMDD, dbDEMC, miR2Disease	42	hsa-mir-196a	HMDD, dbDEMC
18	hsa-mir-200c	HMDD, dbDEMC, miR2Disease	43	hsa-mir-9	HMDD, dbDEMC, miR2Disease
19	hsa-mir-34b	HMDD, dbDEMC	44	hsa-mir-29c	HMDD, dbDEMC, miR2Disease
20	hsa-mir-7	HMDD, dbDEMC, miR2Disease	45	hsa-mir-218	HMDD, dbDEMC, miR2Disease
21	hsa-mir-146b	HMDD, dbDEMC, miR2Disease	46	hsa-mir-130a	HMDD, dbDEMC, miR2Disease
22	hsa-mir-133b	HMDD, dbDEMC, miR2Disease	47	hsa-mir-222	HMDD, dbDEMC
23	hsa-mir-223	HMDD, dbDEMC	48	hsa-mir-15a	HMDD, dbDEMC
24	hsa-mir-199a	HMDD, dbDEMC, miR2Disease	49	hsa-mir-19a	HMDD, dbDEMC, miR2Disease
25	hsa-mir-499a	HMDD, dbDEMC	50	hsa-mir-141	HMDD, dbDEMC, miR2Disease

databases (HMDD v3.2, dbDEMC 2.0, and miR2Disease) (as shown in **Table 3**). Similarly, we removed all 132 known associations related to lung neoplasms to simulate the isolated disease condition. The supporting evidences on the top 50 lung neoplasm-related miRNAs predicted were all found from the above-mentioned three relevant databases (as shown in **Table 4**). The supporting evidences confirmed that the predictive accuracy for the above two simulated objects were both 100%, which further showed the excellent predictive performance of MSFSP on inferring diseases potentially related to miRNAs and isolated diseases related to miRNAs.

DISCUSSION AND CONCLUSION

Considering that the identification of complex disease-related miRNAs is still a key research topic in the bio-medical field, we proposed a computational model called MSFSP that made the following contributions for the identification of miRNA–disease associations: (1) Compared to other methods, MSFSP can enhance the predictive accuracy effectively with an AUC value of 0.9613, which is higher than those of the other current classical computational models; (2) MSFSP implements prediction without needing negative samples; (3) MSFSP solved the inherent limitations of sparsity and incompleteness existing in current datasets *via* multiple similarities fusion; (4) MSFSP can be used to infer new miRNAs and isolated diseases, with AUC values of 0.9493 and 0.8412, respectively; (5) The predicted top 50 results for prostatic neoplasms and lung neoplasms as two cases agree well with the supporting evidences found in HMDD v3.2, dbDEMC 2.0, and miR2Disease, with the consistency of 98 and 96% respectively; (6) The predicted top 50 results for the isolated diseases simulated agree well with the supporting evidences found in HMDD v3.2, dbDEMC 2.0, and miR2Disease, with the consistency of 100% for both.

The reliable performance of MSFSP achieved can be attributed to the following factors: (1) Different biological information data were fused in MSFSP to construct the integrated miRNA similarity network and the integrated disease similarity network; (2) More accurate miRNA–disease correlations were described by weighted networks that were integrated with the disease similarity network, the miRNA similarity network, and the experimentally verified Boolean network of miRNA–disease

associations; (3) MiRNA space projection scores and disease space projection scores were combined to obtain the final prediction scores, which avoided the invalid inference for new miRNAs only with disease space projection scores and the invalid inference for isolated diseases only with miRNA space projection scores.

MSFSP still has some limitations which need to be improved in the future besides its excellent prediction results. Firstly, during miRNA similarity and disease similarity calculation, the known miRNA–disease associations demand extra increase in some amount of overhead because the similarity calculation needs to be redone in LOOCV. Secondly, the construction of miRNA similarity network and disease similarity network is not accurate enough, although the accuracy has been somewhat enhanced by integrating various information. Furthermore, MSFSP can only predict if an association between miRNA and a disease exists or not, but not the specific regulatory mechanism.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

YZ and MC conceived the concept of the work and designed the experiments and wrote the paper. MC, YZ, XC, and HW performed the literature search. MC, YZ, and XC collected and analyzed the data. All authors have approved the manuscript.

FUNDING

The research of this paper has been sponsored by the National Nature Science Foundation of China (Grant Nos. 61772192, 61672223, 61662017, and 61762031), the Nature Science Foundation of Hunan Province, China (Grant Nos. 2018JJ2085 and 2018JJ40064), the Scientific Research Project of the Education Department of Hunan Province, China (19A125), major cultivation projects of Hunan Institute of Technology (Grant No. 2017HGYP001), and Guangxi Key Laboratory Fund of Embedded Technology and Intelligent System (Guilin University of Technology).

REFERENCES

- Bandyopadhyay, S., Mitra, R., Maulik, U., and Zhang, M. Q. (2010). Development of the human cancer microRNA network. *Silence* 1:6. doi: 10.1186/1758-907X-1-6
- Carthew, R. W., and Sontheimer, E. J. (2009). Origins and mechanisms of miRNAs and siRNAs. *Cell* 136:642–655. doi: 10.1016/j.cell.2009.01.035
- Charikar, M. S. (2002). “Similarity estimation techniques from rounding algorithms,” in *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*. (Montreal, QC: ACM). doi: 10.1145/509907.509965
- Chen, M., He, X., Duan, S., and Deng, Y. (2017a). A Novel Gene Selection Method Based on Sparse Representation and Max-Relevance and Min-Redundancy. *Comb. Chem. High Throughput Screen* 20, 158–163. doi: 10.2174/1386207320666170126114051
- Chen, M., Li, Z., Zhang, Y., Chen, X., and Li, A. (2015a). A multiple platform based method for data integration. *J. Comput. Theor. Nanosci.* 12, 4890–4894. doi: 10.1166/jctn.2015.4457
- Chen, M., Liao, B., and Li, Z. (2018a). Global similarity method based on a two-tier random walk for the prediction of microRNA–disease association. *Sci Rep.* 8:6481. doi: 10.1038/s41598-018-24532-7
- Chen, M., Lu, X., Liao, B., Li, Z., Cai, L., and Gu, C. (2016a). Uncover miRNA–Disease Association by Exploiting Global Network Similarity. *PLoS ONE* 11:e0166509. doi: 10.1371/journal.pone.0166509
- Chen, M., Peng, Y., Li, A., Li, Z., Deng, Y., Liu, W., et al. (2018b). A novel information diffusion method based on network consistency

- for identifying disease related microRNAs. *RSC Adv.* 8, 36675–36690. doi: 10.1039/C8RA07519K
- Chen, M., Zhang, Y., Li, A., Li, Z., Liu, W., and Chen, Z. (2019a). Bipartite heterogeneous network method based on Co-neighbour for MiRNA-disease association prediction. *Front. Genet.* 10:385. doi: 10.3389/fgene.2019.00385
- Chen, M., Zhang, Y., Li, Z., Li, A., Liu, W., Liu, L., et al. (2019b). A novel gene selection algorithm based on sparse representation and minimum-redundancy maximum-relevancy of maximum compatibility center. *Curr. Proteomics.* 16, 374–382. doi: 10.2174/1570164616666190123144020
- Chen, M., Zhong, M., Li, Z., Li, X., and Li, A. (2015b). A novel method based on greedy algorithm for informative SNP selection. *J. Comput. Theor. Nanosci.* 12, 4036–4042. doi: 10.1166/jctn.2015.4316
- Chen, X., Cheng, J.-Y., and Yin, J. (2018c). Predicting microRNA-disease associations using bipartite local models and hubness-aware regression. *RNA Biol.* 15, 1192–1205. doi: 10.1080/15476286.2018.1517010
- Chen, X., Gong, Y., Zhang, D. H., You, Z. H., and Li, Z. W. (2018d). DRMDA: deep representations-based miRNA-disease association prediction. *J. Cell. Mol. Med.* 22, 472–485. doi: 10.1111/jcmm.13336
- Chen, X., Guan, N. N., Li, J. Q., and Yan, G. Y. (2018e). GIMDA: graphlet interaction-based MiRNA-disease association prediction. *J. Cell. Mol. Med.* 22, 1548–1561. doi: 10.1111/jcmm.13429
- Chen, X., and Huang, L. (2017). LRSSLMDA: laplacian regularized sparse subspace learning for MiRNA-disease association prediction. *PLoS Comput. Biol.* 13:e1005912. doi: 10.1371/journal.pcbi.1005912
- Chen, X., Huang, L., Xie, D., and Zhao, Q. (2018k). EGBMMDA: extreme gradient boosting machine for MiRNA-disease association prediction. *Cell Death Dis.* 9:3. doi: 10.1038/s41419-017-0003-x
- Chen, X., Jiang, Z. C., Xie, D., Huang, D. S., Zhao, Q., Yan, G. Y., et al. (2017b). A novel computational model based on super-disease and miRNA for potential miRNA-disease association prediction. *Mol. Biosyst.* 13, 1202–1212. doi: 10.1039/C6MB00853D
- Chen, X., Liu, M.-X., and Yan, G.-Y. (2012). RWRMDA: predicting novel human microRNA-disease associations. *Mol. Biosyst.* 8, 2792–2798. doi: 10.1039/c2mb25180a
- Chen, X., Niu, Y. W., Wang, G. H., and Yan, G. Y. (2017c). HAMDA: hybrid approach for MiRNA-disease association prediction. *J. Biomed. Inform.* 76, 50–58. doi: 10.1016/j.jbi.2017.10.014
- Chen, X., and Qu, J. (2018). TLHNMDA: triple layer heterogeneous network based inference for MiRNA-disease association prediction. *Front. Genet.* 9:234. doi: 10.3389/fgene.2018.00234
- Chen, X., Sun, L.-G., and Zhao, Y. (2020). NCMCMDA: miRNA-disease association prediction through neighborhood constraint matrix completion. *Brief. Bioinform.* 12:bbz159. doi: 10.1093/bib/bbz159
- Chen, X., Wang, C.-C., Yin, J., and You, Z.-H. (2018f). Novel human miRNA-disease association inference based on random forest. *Mol. Therapy-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. (2018g). Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* 34, 4256–4265. doi: 10.1093/bioinformatics/bty503
- Chen, X., Wu, Q.-F., and Yan, G.-Y. (2017d). RKNNMDA: ranking-based KNN for MiRNA-disease association prediction. *RNA Biol.* 14, 952–962. doi: 10.1080/15476286.2017.1312226
- Chen, X., Xie, D., Wang, L., Zhao, Q., You, Z.-H., and Liu, H. (2018h). 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. (2019c). 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., Xu, Z., You, Z. H., Yuan, H., and Yan, G. Y. (2016b). HGIMDA: heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget* 7, 65257–65269. doi: 10.18632/oncotarget.11251
- Chen, X., Yan, C. C., Zhang, X., Li, Z., Deng, L., Zhang, Y., et al. (2015c). RBMMMDA: predicting multiple types of disease-microRNA associations. *Sci. Rep.* 5:13877. doi: 10.1038/srep13877
- Chen, X., Yan, C. C., Zhang, X., You, Z. H., Deng, L., Liu, Y., et al. (2016c). WBSMDA: Within and Between Score for MiRNA-Disease Association prediction. *Sci. Rep.* 6:21106. doi: 10.1038/srep21106
- Chen, X., Yin, J., Qu, J., and Huang, L. (2018i). MDHGI: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction. *PLoS Comput. Biol.* 14:e1006418. doi: 10.1371/journal.pcbi.1006418
- Chen, X., Zhou, Z., and Zhao, Y. (2018j). ELLPMDA: Ensemble learning and link prediction for miRNA-disease association prediction. *RNA Biol.* 15, 807–818. doi: 10.1080/15476286.2018.1460016
- Chen, X., Zhu, C.-C., and Yin, J. (2019d). Ensemble of decision tree reveals potential miRNA-disease associations. *PLoS Comput. Biol.* 15:e1007209. doi: 10.1371/journal.pcbi.1007209
- Cheng, A. M., Byrom, M. W., Shelton, J., and Ford, L. P. (2005). Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res.* 33, 1290–1297. doi: 10.1093/nar/gki200
- Fernando, T. R., Rodriguez-Malave, N. I., and Rao, D. S. (2012). MicroRNAs in B cell development and malignancy. *J. Hematol. Oncol.* 5:7. doi: 10.1186/1756-8722-5-7
- Guo, Q., Zheng, M., Xu, Y., Wang, N., and Zhao, W. (2019). MiR-384 induces apoptosis and autophagy of non-small cell lung cancer cells through the negative regulation of Collagen α -1(X) chain gene. *Biosci Reports.* 39:BSR20181523. doi: 10.1042/BSR20181523
- Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., et al. (2018). HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.* 47, D1013–D1017. doi: 10.1093/nar/gky1010
- Iorio, M. V., Ferracin, M., Liu, C.-G., Veronese, A., Spizzo, R., Sabbioni, S., et al. (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* 65, 7065–7070. doi: 10.1158/0008-5472.CAN-05-1783
- Jiang, L., Ding, Y., Tang, J., and Guo, F. J. (2018). MDA-SKF: similarity kernel fusion for accurately discovering miRNA-disease association. *Front. Genet.* 9:618. doi: 10.3389/fgene.2018.00618
- Jiang, L., Xiao, Y., Ding, Y., Tang, J., and Guo, F. (2019). Discovering cancer subtypes via an accurate fusion strategy on multiple profile data. *Front. Genet.* 10:20. doi: 10.3389/fgene.2019.00020
- Jiang, Q., Hao, Y., Wang, G., Juan, L., Zhang, T., Teng, M., et al. (2010a). Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst. Biol.* 28(4 Suppl. 1):S2. doi: 10.1186/1752-0509-4-S1-S2
- Jiang, Q., Wang, G., and Wang, Y. (2010b). “An approach for prioritizing disease-related microRNAs based on genomic data integration,” in *Proceedings of the International Conference on Biomedical Engineering and Informatics*, (Yantai: IEEE) (2010). doi: 10.1109/BMEI.2010.5639313
- Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., et al. (2009). miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* 37, D98–D104. doi: 10.1093/nar/gkn714
- Kozomara, A., and Griffiths-Jones, S. (2013). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181
- Li, G., Luo, J., Xiao, Q., Liang, C., and Ding, P. (2018). Predicting microRNA-disease associations using label propagation based on linear neighborhood similarity. *J. Biomed. Inform.* 82, 169–177. doi: 10.1016/j.jbi.2018.05.005
- Li, J., Wu, Z., Cheng, F., Li, W., Liu, G., and Tang, Y. (2014). Computational prediction of microRNA networks incorporating environmental toxicity and disease etiology. *Sci. Rep.* 4:5576. doi: 10.1038/srep05576
- Li, J. Q., Rong, Z. H., Chen, X., Yan, G. Y., and You, Z. H. (2017). MCMMDA: Matrix Completion for MiRNA-Disease Association prediction. *Oncotarget* 8, 21187–21199. doi: 10.18632/oncotarget.15061
- Li, X., Wang, Q., Zheng, Y., Lv, S., Ning, S., Sun, J., et al. (2011). Prioritizing human cancer microRNAs based on genes' functional consistency between microRNA and cancer. *Nucleic Acids Res.* 39:e153. doi: 10.1093/nar/gkr770
- 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. 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
- Liu, Y., Zeng, X., He, Z., and Zou, Q. (2017). "Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources," in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. (IEEE), 14, 905–915. doi: 10.1109/TCBB.2016.2550432
- Lowe, H. J., and Barnett, G. O. (1994). Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *Jama* 271:1103. doi: 10.1001/jama.1994.03510380059038
- Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W., et al. (2008). An Analysis of Human MicroRNA and Disease Associations. *PLoS ONE* 3:e3420. doi: 10.1371/journal.pone.0003420
- Luo, J., Ding, P., Cheng, L., Cao, B., and Chen, X. (2017a). "Collective Prediction of Disease-Associated miRNAs Based on Transduction Learning," in *IEEE/ACM Transactions on Computational Biology & Bioinformatics*. (IEEE), 14:7. doi: 10.1109/TCBB.2016.2599866
- Luo, J., and Xiao, Q. (2017). A novel approach for predicting microRNA-disease associations by unbalanced bi-random walk on heterogeneous network. *J. Biomed. Inform.* 66, 194–203. doi: 10.1016/j.jbi.2017.01.008
- Luo, J., Xiao, Q., Liang, C., and Ding, P. (2017b). Predicting MicroRNA-Disease associations using kronecker regularized least squares based on heterogeneous omics data. *IEEE Access*. 5, 2503–2513. doi: 10.1109/ACCESS.2017.2672600
- Miska, E. A. (2005). How microRNAs control cell division, differentiation and death. *Curr. Opin. Genet. Dev.* 15:563–568. doi: 10.1016/j.gde.2005.08.005
- Mugunga, I., Ju, Y., Liu, X., and Huang, X. (2017). Computational prediction of human disease-related microRNAs by path-based random walk. *Oncotarget* 8, 58526–58535. doi: 10.18632/oncotarget.17226
- Nalluri, J. J., Kamapantula, B. K., Barh, D., Jain, N., Bhattacharya, A., de Almeida, S. S., et al. (2015). DISMIRA: Prioritization of disease candidates in miRNA-disease associations based on maximum weighted matching inference model and motif-based analysis. *BMC Genomics* 16:S12. doi: 10.1186/1471-2164-16-S5-S12
- Peng, L., Chen, Y., Ma, N., and Chen, X. (2017a). NARRMDA: negative-aware and rating-based recommendation algorithm for miRNA-disease association prediction. *Mol. Biosyst.* 2650–2659. doi: 10.1039/C7MB00499K
- Peng, L., Peng, M., Liao, B., Huang, G., Liang, W., and Li, K. (2017b). Improved low-rank matrix recovery method for predicting miRNA-disease association. *Sci. Rep.* 7:6007. doi: 10.1038/s41598-017-06201-3
- Peng, L.-H., Sun, C.-N., Guan, N.-N., Li, J.-Q., and Chen, X. (2018). HNMDA: heterogeneous network-based miRNA-disease association prediction. *Mol. Genet. Genomics*. 293, 983–995. doi: 10.1007/s00438-018-1438-1
- Peng, L.-H., Zhou, L.-Q., Chen, X., and Piao, X. (2020). A computational study of potential miRNA-disease association inference based on ensemble learning and kernel ridge regression. *Front. Bioeng. Biotechnol.* 8:40. doi: 10.3389/fbioe.2020.00040
- Qu, J., Chen, X., Yin, J., Zhao, Y., and Li, Z.-W. (2019). Prediction of potential miRNA-disease associations using matrix decomposition and label propagation. *Knowl. Based Syst.* 186:104963. doi: 10.1016/j.knsys.2019.104963
- Siegel, R. L., Miller, K. D., and Jemal, A. (2016). Cancer statistics, (2016). *CA Cancer J. Clin.* 66:7–30. doi: 10.3322/caac.21332
- Sun, D., Li, A., Feng, H., and Wang, M. (2016). NTSMDA: prediction of miRNA-disease associations by integrating network topological similarity. *Mol. Biosyst.* 12:2224. doi: 10.1039/C6MB00049E
- Tang, C., Zhou, H., Zheng, X., Zhang, Y., and Sha, X. (2019). Dual Laplacian regularized matrix completion for microRNA-disease associations prediction. *RNA Biol.* 16:601–611. doi: 10.1080/15476286.2019.1570811
- Temraz, S., Charafeddine, M., Mukherji, D., and Shamseddine, A. (2017). Trends in lung cancer incidence in Lebanon by gender and histological type over the period 2005–2008. *J. Epidemiol. Glob. Health.* 7:161–167. doi: 10.1016/j.jegh.2017.04.003
- Torre, L. A., Siegel, R. L., and Jemal, A. (2016). "Lung cancer statistics," in *Lung Cancer and Personalized Medicine*. (Springer). 1–19. doi: 10.1007/978-3-319-24223-1_1
- Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., et al. (2014). Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* 11:333. doi: 10.1038/nmeth.2810
- Wang, C.-C., Chen, X., Yin, J., and Qu, J. (2019). An integrated framework for the identification of potential miRNA-disease association based on novel negative samples extraction strategy. *RNA Biol.* 16, 257–269. doi: 10.1080/15476286.2019.1568820
- 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
- Wang, L., You, Z.-H., Chen, X., Li, Y.-M., Dong, Y.-N., Li, L.-P., et al. (2019). LMTRDA: using logistic model tree to predict MiRNA-disease associations by fusing multi-source information of sequences and similarities. *PLoS Comput. Biol.* 15:e1006865. doi: 10.1371/journal.pcbi.1006865
- Xiao, Q., Luo, J., Liang, C., Cai, J., and Ding, P. (2018). A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics* 34, 239–248. doi: 10.1093/bioinformatics/btx545
- Xu, C., Ping, Y., Li, X., Zhao, H., Wang, L., Fan, H., et al. (2014). Prioritizing candidate disease miRNAs by integrating phenotype associations of multiple diseases with matched miRNA and mRNA expression profiles. *Mol. Biosyst.* 10, 2800–2809. doi: 10.1039/C4MB00353E
- 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.
- Xu, P., Guo, M., and Hay, B. A. (2004). MicroRNAs and the regulation of cell death. *Trends Genet.* 20, 617–624. doi: 10.1016/j.tig.2004.09.010
- Xuan, P., Dong, Y., Guo, Y., Zhang, T., and Liu, Y. (2018). Dual convolutional neural network based method for predicting disease-related miRNAs. *Int. J. Mol. Sci.* 19:3732. doi: 10.3390/ijms19123732
- 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
- Xuan, P., Li, L., Zhang, T., Zhang, Y., and Song, Y. (2019a). Prediction of Disease-related microRNAs through integrating attributes of microRNA Nodes and multiple kinds of connecting edges. *Molecules* 24:3099. doi: 10.3390/molecules24173099
- Xuan, P., Sun, H., Wang, X., Zhang, T., and Pan, S. (2019b). Inferring the Disease-Associated miRNAs Based on Network Representation Learning and Convolutional Neural Networks. *Int. J. Mol. Sci.* 20:3648. doi: 10.3390/ijms20153648
- Xuan, P., Zhang, Y., Zhang, T., Li, L., and Zhao, L. (2019c). Predicting miRNA-disease associations by incorporating projections in low-dimensional space and local topological information. *Genes* 10:685. doi: 10.3390/genes10090685
- Yang, Z., Wu, L., Wang, A., Tang, W., Zhao, Y., Zhao, H., et al. (2017). dbDEM2.0: 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). PBMDA: 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., Liu, L., Lü, L., and Zou, Q. (2018). Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics* 34, 2425–2432. doi: 10.1093/bioinformatics/bty112
- 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., Chen, X., and Yin, J. (2019). Prediction of Potential miRNA-disease associations through a novel unsupervised deep learning framework with variational autoencoder. *Cells* 8:1040. doi: 10.3390/cells8091040
- Zhang, Y., Chen, M., Cheng, X., and Chen, Z. (2019). LSGSP: a novel miRNA-disease association prediction model using a Laplacian score of the graphs and space projection federated method. *RSC Adv.* 9, 29747–29759. doi: 10.1039/C9RA05554A

- Zhao, Y., Chen, X., and Yin, J. (2018). A novel computational method for the identification of potential miRNA-disease association based on symmetric non-negative matrix factorization and Kronecker regularized least square. *Front. Genet.* 9:324. doi: 10.3389/fgene.2018.00324
- Zhao, Y., Chen, X., and Yin, J. (2019). Adaptive boosting-based computational model for predicting potential miRNA-disease associations. *Bioinformatics* 35, 4730–4738. doi: 10.1093/bioinformatics/btz297
- Zhu, L., Zhao, J., Wang, J., Hu, C., Peng, J., Luo, R., et al. (2016). MicroRNAs are involved in the regulation of ovary development in the pathogenic blood fluke *Schistosoma japonicum*. *PLoS Pathog.* 12:e1005423. doi: 10.1371/journal.ppat.1005423

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.

Copyright © 2020 Zhang, Chen, Cheng and Wei. 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.