


Adaptive multi-source multi-view latent feature learning for inferring potential disease-associated miRNAs

Qiu Xiao , Ning Zhang, Jiawei Luo, Jianhua Dai and Xiwei Tang

Corresponding authors: Jiawei Luo, College of Computer Science and Electronic Engineering, Hunan University, Changsha, 410082, China. Tel: +86-731-88821971; E-mail: luojiawei@hnu.edu.cn; Jianhua Dai, College of Information Science and Engineering, Hunan Normal University, Changsha, 410081, China. Tel.: +86-731-88872564; E-mail: jhdai@hunnu.edu.cn

Abstract

Accumulating evidence has shown that microRNAs (miRNAs) play crucial roles in different biological processes, and their mutations and dysregulations have been proved to contribute to tumorigenesis. *In silico* identification of disease-associated miRNAs is a cost-effective strategy to discover those most promising biomarkers for disease diagnosis and treatment. The increasing available omics data sources provide unprecedented opportunities to decipher the underlying relationships between miRNAs and diseases by computational models. However, most existing methods are biased towards a single representation of miRNAs or diseases and are also not capable of discovering unobserved associations for new miRNAs or diseases without association information. In this study, we present a novel computational method with adaptive multi-source multi-view latent feature learning (M2LFL) to infer potential disease-associated miRNAs. First, we adopt multiple data sources to obtain similarity profiles and capture different latent features according to the geometric characteristic of miRNA and disease spaces. Then, the multi-modal latent features are projected to a common subspace to discover unobserved miRNA-disease associations in both miRNA and disease views, and an adaptive joint graph regularization term is developed to preserve the intrinsic manifold structures of multiple similarity profiles. Meanwhile, the $L_{p,q}$ -norms are imposed into the projection matrices to ensure the sparsity and improve interpretability. The experimental results confirm the superior performance of our proposed method in screening reliable candidate disease miRNAs, which suggests that M2LFL could be an efficient tool to discover diagnostic biomarkers for guiding laborious clinical trials.

Key words: microRNAs; disease miRNA inference; multi-source learning; feature extraction; adaptive learning

Qiu Xiao is an assistant professor in the College of Information Science and Engineering, Hunan Normal University, China. His current research interests include bioinformatics, complex network and artificial intelligence.

Ning Zhang is a student in the College of Information Science and Engineering, Hunan Normal University, China. Her current research interests include deep learning, data mining and bioinformatics.

Jiawei Luo is a professor in the College of Computer Science and Electronic Engineering, Hunan University, China. Her primary research interests include bioinformatics, data mining and deep learning.

Jianhua Dai is a professor in the College of Information Science and Engineering, Hunan Normal University, Changsha, China. His current research interests include artificial intelligence, bioinformatics and evolutionary computation.

Xiwei Tang is an associate professor in the School of Information Science and Engineering, Hunan First Normal University, China. His current research interests include computational genomics and proteomics.

Submitted: 13 December 2019; **Received (in revised form):** 16 February 2020

Introduction

MicroRNAs (miRNAs) are a group of short non-coding RNAs that regulate gene expression at the post-transcriptional level, which have been reported to be involved in tumorigenesis and progression [1, 2]. Existing studies revealed that miRNAs are important regulators and participate in different biological processes or pathways, including cell growth, cell proliferation, cellular apoptosis and metabolism [3–5]. The majority of well-characterized miRNAs have indicated their significant roles in many human complex diseases or tumors, such as lung neoplasms, laryngeal squamous cell carcinoma, colorectal cancer and so on [6, 7]. For example, miR-367 is observed to promote hepatocellular carcinoma cell proliferation and show aberrant expressions in hepatocellular cancer patient tumor samples or tissues than that in normal tissues [8]. Therefore, identifying the underlying relationships between miRNAs and diseases will undoubtedly facilitate the understanding of pathogenic mechanisms and contribute to the development of personalized treatment. In recent years, although some efforts have been devoted to discover disease miRNAs by biological experiments, the roles and activities related to miRNAs in human diseases are still insufficiently understood. Thus, it is imperative to reveal potential miRNA and disease relationships by computational models.

With the rapid accumulation of genomic data, it creates huge opportunities to construct high-efficiency and low-cost *in silico* methods for help to study various biological problems, such as complex biological network construction [9, 10], functional module identification [11, 12], disease candidate molecule prioritization [13–15] and DNA-protein binding site prediction [16, 17]. It is well known that the traditional biological experiment methods usually time consuming, expensive, small scale and susceptible to outside world. Therefore, computationally prioritizing the most possible disease miRNA candidates can help to guide and accelerate those costly and tremendously laborious clinical experiments. Considering miRNAs might be biomarkers for tumor diagnosis and treatment, some computational models have been developed to decipher those underlying miRNA-disease associations [18]. The network- and machine learning-based approaches are popular for prioritizing disease-associated miRNAs, and many of them are based on the assumption that functional similar miRNAs tend to be related to phenotypically similar diseases [19]. By constructing miRNA-disease heterogeneous network, Luo et al. [20] developed a method called BRWH to infer disease-related miRNA candidates by unbalanced bi-random walk. Chen et al. [21] adopted the matrix decomposition and heterogeneous graph inference to discover those missing associations between miRNAs and diseases. Yu et al. [22] first constructed a miRNA-disease heterogeneous network, and then introduced a method of global linear neighborhoods for scoring disease and miRNA pairs. Chen et al. [23] designed a bipartite network projection model to infer associations based on the known miRNA-disease network and the integrated similarity of miRNAs and diseases. You et al. [24] proposed a path-based model to uncover miRNA-disease associations in a heterogeneous graph network through the fusing of multiple biological data. Moreover, a series of other network-based models have also been presented to infer disease candidate miRNAs [25, 26]. However, the performance of this type of models is unusually influenced by the quality of those constructed networks.

In addition, many previous computational approaches take advantage of machine learning algorithms and various

genomic data for prediction of disease miRNAs. For example, Xiao et al. [27] introduced a graph regularized non-negative matrix factorization method (GRNMF) for discovering the latent associations between diseases and miRNAs. By integrating heterogeneous omics data, Luo et al. [28] presented a model named KRLSM to prioritize disease-related miRNAs through Kronecker regularized least squares. Xiao et al. [29] proposed a multi-view manifold regularized learning-based method (MRSLA) to predict candidate disease miRNAs. Chen et al. [30] developed a semi-supervised learning method to discover candidate disease-associated miRNAs. Zhao et al. [31] introduced an adaptive boosting-based method to predict associations between miRNAs and diseases. Chen et al. [32] proposed a computational framework with decision tree to predict miRNA-disease associations, and the feature vectors were extracted by utilizing ensemble learning and dimensionality reduction. Chen et al. [33] introduced another method called RKNMMDA to predict potential disease-related miRNAs based on k nearest neighbor strategy. Li et al. [34] first calculated pairwise linear neighborhood similarities for miRNAs and disease and they developed a computational model for scoring miRNA and disease pairs using label propagation. By fully exploiting the statistical and graph theoretical profiles obtained from similarity networks, Chen et al. [35] adopted a subspace learning model with Laplacian regularization to uncover miRNA-disease associations. Liang et al. [36] presented a model for screening disease miRNAs based on data integration by using multi-label learning. It is well known that deep learning shows superior performance in many applications. Peng et al. [37] and Zeng et al. [38] presented two learning-based frameworks to identify potential disease-related miRNAs based on neural networks. By combining multi-label learning with semi-supervised graph convolutional networks, Pan et al. [39] proposed a method to infer disease-associated miRNAs by fusing multiple data sources. Li et al. [40] introduced a topology-based similarity measure method with DeepWalk to reveal unobserved associations between miRNAs and diseases. More recently, Chen et al. [41] presented a novel computational model called NCMCMDA for miRNA-disease association prediction based on neighborhood constraint with matrix completion. Meanwhile, they also developed some models with logistic regression [42], extreme gradient boosting machine [43], label propagation [44] and restricted Boltzmann machine [45], to discover those missing associations between miRNAs and diseases. Overall, all these abovementioned models have provided great insights into the disease mechanisms at miRNA level as well as the detection of potential prognostic biomarkers.

Despite great efforts have been made to explore the underlying miRNA and disease relationships, some limitations still exist for those existing computational models. First, many network-based methods strongly rely on the experimentally verified miRNA-disease associations and cannot be applied for new diseases or miRNAs that without any known association information. In addition, the integration of multiple complementary data sources can usually improve performance, and some previous methods often obtained different similarity matrices for both miRNAs and diseases, but most of them generally adopted a simple average or linear weighting strategy to combine different miRNA similarity matrices and disease similarity matrices, which may not be the best choice for model learning. What is more, a number of methods only focus on a single representation for disease or miRNA spaces. Existing studies show that many real-world data are comprised of different features and incorporating multiple representations

or multi-modal features could often strengthen the prediction ability.

In this work, we introduce a novel integrative framework to discover potential disease-associated miRNAs with adaptive multi-source multi-view latent feature learning (M2LFL), which projected the multi-modal latent features into a common subspace for both miRNA and disease spaces. First, the framework comprehensively estimates similarities for both miRNA–miRNA pairs and disease–disease pairs and extracts multiple latent features of miRNAs and diseases by fully exploiting multiple biological data sources, including miRNA sequence information, known miRNA–disease associations, gene ontology (GO), miRNA target interactions and disease semantic information. Then, inspired by [46, 47], we infer potential disease-related miRNAs by a feature-based ensemble learning method. Additionally, we perform the learning task by incorporating adaptive graph regularization and $L_{p,q}$ -norm constrain to improve the interpretability and performance. Finally, we use an alternating iteration algorithm to solve the optimization problem of M2LFL, and the results show that M2LFL achieves superior performance compared with other state-of-the-art methods in discovering candidate disease miRNAs.

The main contributions of this paper include as follows:

- We present a multi-source multi-view learning approach to infer disease-associated miRNAs from both miRNA and disease perspectives, which incorporates multiple similarity profiles, multi-modal latent features as well as known association information.
- By using manifold theory, an adaptive joint graph regularization is introduced to efficiently integrate diverse similarity profiles and preserve those manifold characteristics for data spaces, which could automatically update the weights and ensure that all graph Laplacian items contribute to the learning task.
- To improve interpretability and alleviate the effects of the inherent noise in both miRNA and disease feature spaces, the $L_{p,q}$ mixed-norms is imposed on the projection matrices to select the most representative or discriminative sparse features.
- The proposed method combines different representations related to diseases and miRNAs, which could be used to infer miRNA–disease associations for the scenarios involving new miRNAs or diseases.

The rest of this paper is organized as follows. In ‘Materials and methods,’ multiple similarity profiles and features for miRNAs and diseases are introduced, and then the mathematical formulation of M2LFL is presented. Finally, the experimental results with additional analysis and the study’s conclusions are given in ‘Results and discussion.’

Materials and methods

Human miRNA–disease associations

We downloaded the experimentally supported miRNA–disease associations from HMDD v2.0 [48], which is a manually collected database of observed human disease and miRNA associations, and each entry includes detailed information on a miRNA–disease association, including miRNA and disease names, PubMed id of reference and a brief functional description. Meanwhile, we also downloaded the disease descriptors from the U.S. National Library of Medicine (MeSH) [49], in which the semantic information of diseases can be explained by

directed acyclic graphs (DAGs), as shown in Figure 1, and the disease semantic similarity can be estimated according to the disease DAG hierarchical structures. Additionally, the miRNA annotations are obtained from miRBase [50] to maintain the consistency of data from different data sources. After filtering out associations that their corresponding miRNA names or disease names were absent in the MeSH descriptors or miRBase records, we finally retained 6088 known associations involving 550 miRNAs and 328 diseases as the benchmark dataset for model learning.

Multi-source similarity calculation and representation

In practice, various heterogeneous data sources often contain different representation or complementary information [51–53]. In this section, we incorporate multiple biological omics data (Table 1) to comprehensively estimate disease–disease similarity and miRNA–miRNA similarity from different modalities and present two types of disease similarity profiles and four types of miRNA similarity profiles for M2LFL to lead to high accuracy.

Disease similarity profiles

Disease semantic similarity. As described in Zou et al. [54], the MeSH disease descriptors have been widely adopted to estimate disease semantic similarity based on their corresponding DAG structures. Formally, a disease d can be denoted as $DAG_d = (d, T(d), E(d))$, where $T(d)$ represents d itself and all its ancestor nodes, $E(d)$ is an edge set that contains all the direct links between parent nodes and child nodes in DAG_d . The semantic contribution of disease t related to d was calculated as follows:

$$C_d(t) = \begin{cases} 1 & \text{if } t = d \\ \max\{0.5 * C_d(t') \mid t' \in \text{children of } t\}, & \text{if } t \neq d \end{cases} \quad (1)$$

Then, the pairwise disease semantic similarity was defined by:

$$SS_d(d_i, d_j) = \frac{\sum_{t \in T(d_i) \cap T(d_j)} (C_{d_i}(t) + C_{d_j}(t))}{SV(d_i) + SV(d_j)} \quad (2)$$

where $SV(d_i) = \sum_{t \in T(d_i)} C_{d_i}(t)$ denotes the semantic value of diseases d_i . Obviously, according to the Equation (2), we can know that two diseases sharing a greater overlap of their DAGs are usually more similar.

Disease Gaussian interaction profile kernel similarity. It has been shown that the Gaussian interaction profile (GIP) can be used to capture the topological characteristics of interaction networks, which was successfully adopted to estimate kernel similarity for each pair of biomolecules or items in computational biology [13, 55]. Let $IP(d_i)$ and $IP(d_j)$ denote the binary interaction profile vectors of diseases d_i and d_j , which correspond to the i -th and j -th rows in the adjacency matrix of known association network, respectively, the GIP kernel similarity between diseases d_i and d_j was defined as follows:

$$GS_d(d_i, d_j) = \exp\left(-\gamma_d \|IP(d_i) - IP(d_j)\|^2\right) \quad (3)$$

where n denotes the number of human diseases, $\gamma_d = \gamma_d' / (\frac{1}{n} \sum_{i=1}^n \|IP(d_i)\|^2)$ controls the kernel bandwidth and γ_d' was set to 1 as suggested in Xiao et al. [29].

MiRNA similarity profiles

MiRNA functional similarity. Based on the commonly accepted assumption that miRNAs with similar functions are more likely

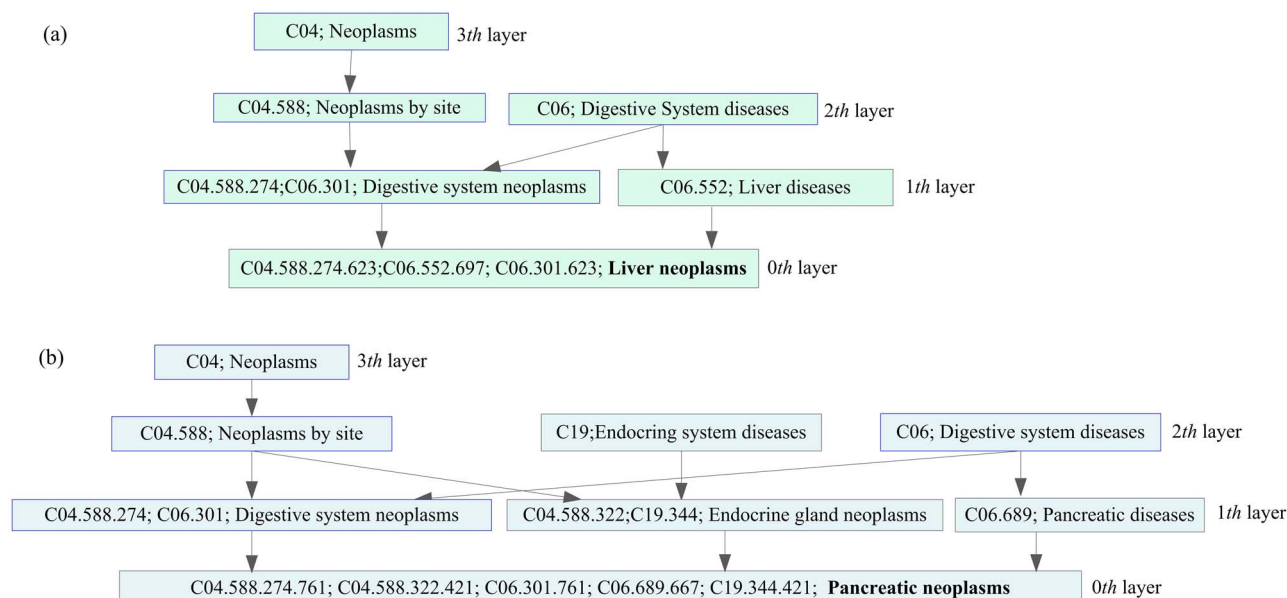


Figure 1. The directed acyclic subgraphs structures of two diseases.

Table 1. The details of data sources used in M2LFL

Name	Website	Reference	Description
HMDD	http://www.cuilab.cn/hmdd	[44]	The experimentally supported miRNA-disease associations
MeSH	https://www.nlm.nih.gov/mesh	[45]	Disease descriptors
miRBase	http://www.mirbase.org/	[46]	miRNA sequences and annotation
Gene ontology annotations	http://www.bioconductor.org/	—	R package 'org.Hs.eg.db'
mirTarBase	http://mirtarbase.mbc.nctu.edu.tw	[48]	The experimentally validated miRNA-target gene interactions
miRCancer	http://mircancer.ecu.edu/	[53]	MicroRNA cancer association database
dbDEMC	http://www.picb.ac.cn/dbDEMC	[54]	Differentially expressed miRNAs in human cancers
miR2Disease	http://www.mir2disease.org/	[55]	A manually curated database for miRNA deregulation in human disease

to be associated with similar diseases and vice versa, we utilize the most popular strategy, as illustrated in literature [54], to quantify the functional similarity for miRNA-miRNA pairs m_i and m_j as follows:

$$FS_m(m_i, m_j) = \frac{\sum_{d \in T(m_i)} ST(d, T(m_j)) + \sum_{d \in T(m_j)} ST(d, T(m_i))}{|T(m_i)| + |T(m_j)|} \quad (4)$$

where $T(m_i)$ and $T(m_j)$ represent the disease sets related to m_i and m_j , respectively; $|T(\cdot)|$ is the cardinality of disease set, and $ST(d, T(m_i)) = \max_{d_x \in T(m_i)} (SS_d(d, d_x))$.

MiRNA sequence similarity. To measure miRNA sequence similarity, we obtain their sequence information from miRBase [50] and use the *pairwiseAlignment* function in R package *Biostrings* to quantify similarities for any two miRNAs. The similarity values are calculated according to the entire mature sequences with a gap extension penalty of 2 and a gap opening penalty of 5. The match and mismatch scores are set to 1 and -1 [36], respectively, to obtain a substitution matrix for sequence alignment. Meanwhile, the sequence similarity score, $Score(m_i, m_j)$, can be further normalized to range $[0,1]$ based on min-max normalization as

follows:

$$SC_m(m_i, m_j) = \frac{Score(m_i, m_j) - SC_{\min}}{SC_{\max} - SC_{\min}} \quad (5)$$

where SC_{\max} and SC_{\min} denote the maximum and minimum similarity scores, respectively.

MiRNA semantic similarity. The aforementioned miRNA functional similarities rely on those miRNA-related diseases, and thus cannot be used for the novel miRNAs without any known associations. Therefore, we also integrate miRNA targets and those gene-related annotation items to measure miRNA semantic similarity. The verified miRNA-target gene interaction information are downloaded from mirTarBase [56], and GO annotation information are acquired based on the R package 'org.Hs.eg.db.' The semantic similarity score between two miRNAs is calculated according to their related target-gene lists as well as those GO annotations related to the two group genes. Here, we use the *clusterSim* function in *GOSemSim* package to estimate semantic similarity between two gene sets, and adopt *best-match average* (BMA) method to combine semantic similarity values of multiple GO terms [57]. Finally, the miRNA semantic similarity matrix can be obtained and the corresponding similarity score for m_i and m_j pair is denoted as $SS_m(m_i, m_j)$.

MiRNA GIP kernel similarity. Similarly, the GIP kernel similarity between miRNAs m_i and m_j also can be estimated as follows:

$$GS_m(m_i, m_j) = \exp\left(-\gamma_m \|\text{IP}(m_i) - \text{IP}(m_j)\|^2\right)$$

$$\gamma_m = \gamma'_m / \left(\frac{1}{m} \sum_{i=1}^m \|\text{IP}(m_i)\|^2\right) \quad (6)$$

where $\text{IP}(m_i)$ and $\text{IP}(m_j)$ are the interaction profiles of m_i and m_j , which correspond to the i -th and j -th columns in the adjacency matrix of association network, respectively; m represents the number of miRNAs, γ'_m was set to 1 as in previous studies.

Network-based multi-modal feature extraction

In this section, we mainly describe the multi-modal feature extraction for miRNAs and diseases according to the known association network and similarity networks. As we know, a complex human disease is usually associated with many miRNAs, and a miRNA is also involved in many different diseases. Here, we extract the first type of feature vectors from the known miRNA-disease association network, which is represented by a

matrix $X = \begin{bmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mn} \end{bmatrix}$ and the entity $x_{ij} = 1$ if miRNA

m_i was known to be associated with disease d_j , otherwise $x_{ij} = 0$. Then, the feature vectors of miRNAs (or diseases) can be obtained through the corresponding rows (or columns) in matrix X , i.e. each row binary vector $X(m_i) = \{x_{i1}, x_{i2}, \dots, x_{in}\}$ corresponds to a miRNA while each column vector $X(d_j) = \{x_{1j}, x_{2j}, \dots, x_{mj}\}^T$ represents a disease.

Additionally, the miRNA (or disease) similarity network reflect the degree of closeness among different miRNAs (or diseases), which also contain valuable information for generating latent features. Considering the sparseness of different similar networks, and if a sparse network is used to capture latent features, it is not conducive to performance improvement. Here, the GIP similarity networks in both miRNA and disease spaces are relatively dense, we utilize them to extract the other type of latent features for miRNAs and diseases. Inspired by Lu et al. [58], the principal component analysis (PCA) is used to capture low-dimensional features, and the singular value decomposition (SVD) is adopted to perform the PCA. Since the GIP kernel similarity matrices GS_d and GS_m are symmetric, both of them can be expressed as the form of $U \Sigma U^T$, where U is an unitary matrix and Σ is an diagonal matrix with singular values deposited on its diagonal. Let $F_m(m_i) = \{f_{i1}, f_{i2}, \dots, f_{ik_m}\}$ and $F_d(d_i) = \{f_{1i}, f_{2i}, \dots, f_{k_d i}\}^T$ be the low-dimensional latent feature vectors for miRNA m_i and disease d_i , respectively, we calculate the dimensions k_d and k_m of them based on the dominating energy strategy [59] as follows:

$$k_d = \arg \min_k \left(\frac{\sum_{i=1}^k (GS_d)_{ii}}{\sum_{j=1}^n (GS_d)_{jj}} \geq \sigma_d \right) \quad (7)$$

and

$$k_m = \arg \min_k \left(\frac{\sum_{i=1}^k (GS_m)_{ii}}{\sum_{j=1}^m (GS_m)_{jj}} \geq \sigma_m \right) \quad (8)$$

where σ_d and σ_m are the free parameters controlling k_d and k_m , respectively. Finally, we obtain two types of miRNA and disease latent feature vectors to implement the learning task.

Adaptive M2LFL framework

Problem statement and notations

Suppose that we have m miRNAs and n diseases, which are denoted by $T_m = \{m_1, m_2, \dots, m_m\}$ and $T_d = \{d_1, d_2, \dots, d_n\}$. As aforementioned, the known miRNA-disease association network is represented as $X = \{x_{ij}\}_{i=1, j=1}^{m, n}$, and the entity is defined by

$$x_{ij} = \begin{cases} 1 & \text{if miRNA } m_i \text{ is known associated with disease } d_i \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

Assuming that we have obtained N (or N') types of similarity profiles and M (or M') types of latent features for miRNAs (or diseases), and denote the similarity profiles and feature matrices as $S = \{S_m^{(i)} \in R^{m \times m}\}_{i=1}^N$ (or $S' = \{S'_d \in R^{n \times n}\}_{i=1}^{N'}$) and $F = \{F_m^{(i)} \in R^{m \times k_m^{(i)}}\}_{i=1}^M$ (or $F' = \{F'_d \in R^{n \times k_d^{(i)}}\}_{i=1}^{M'}$) for miRNAs (or diseases), respectively. In this study, we obtain four and two types of similarity profiles for miRNAs and diseases (i.e. $N=4$ and $N'=2$), respectively, and the number of feature matrices is two for both miRNA and disease spaces (i.e. $M=2$ and $M'=2$). Since the available experimentally verified associations are still very limited, the main purpose of this work is to discover the unobserved associations based on the above information, and the predicted association matrix is denoted by $G \in R^{m \times n}$, which has the same dimensionality with matrix X . The main notations and abbreviations are listed in Table 2.

Mathematical formulation

In this study, we present a model named M2LFL to predict the most possible disease-associated miRNAs. Many studies indicate that implementing learning tasks by integrating different views can help to improve performance [29, 46]. In addition, it has also been shown that incorporating various features or exploiting complementary information can usually contribute to high-accuracy models [46, 60]. Therefore, we formulate the model of M2LFL by utilizing multiple similarity profiles and multi-modal latent features and perform the disease-associated miRNA inferring task based on the views of miRNAs and diseases.

As aforementioned, due to the insufficient experimentally validated miRNA-disease associations, the corresponding adjacency matrix X is sparse and those values in the interaction profiles $\text{IP}(m_i)$ and $\text{IP}(d_i)$ for novel miRNAs or diseases are all zeros. Therefore, inspired by the successful application of weighted K nearest known neighbors (WKNKN) in the previous study [27], we perform a pre-processing procedure with WKNKN to add more interaction information to infer the initial probability values for these interacting or non-interacting pairs. For each miRNA m_i or disease d_i , the interaction profile is updated according to the weighted average of its neighbors' profile information, and the WKNKN is formulated as follows:

$$X_m(m_j, :) = \frac{1}{Q_m} \sum_{i=1}^K w_i^{(m)} X(m_i)$$

$$X_d(:, d_j) = \frac{1}{Q_d} \sum_{i=1}^K w_i^{(d)} X(d_i) \quad (10)$$

where m_1 (or d_1) to m_K (or d_K) denote the K nearest known neighbors of m_j (or d_j) in the kernel similarity network; $Q_m = \sum_{i=1}^K GS_m(m_i, m_j)$ and $Q_d = \sum_{i=1}^K GS_d(d_i, d_j)$ are the normalization terms; $w_i^{(m)}$ and $w_i^{(d)}$ represent the weight coefficients. As suggested in Xiao et al. [27], here we also set K to 5 in both miRNA and

Table 2. The main notations and abbreviations

Notation	Definition
m	Number of miRNAs
n	Number of diseases
$X \in \mathbb{R}^{m \times n}$	Known miRNA-disease association matrix
$SS_d \in \mathbb{R}^{n \times n}, GS_d \in \mathbb{R}^{n \times n}$	Two types of disease similarity profiles
$FS_m \in \mathbb{R}^{m \times m}, SC_m \in \mathbb{R}^{m \times m}, SS_m \in \mathbb{R}^{m \times m}, GS_m \in \mathbb{R}^{m \times m}$	Four types of miRNA similarity profiles
$S_m^{(i)} \in \mathbb{R}^{m \times m}$ (or $S_d^{(i)} \in \mathbb{R}^{n \times n}$)	The i -th miRNA (or disease) similarity profile
$F_m^{(i)} \in \mathbb{R}^{m \times k_m^{(i)}}$ (or $F_d^{(i)} \in \mathbb{R}^{n \times k_d^{(i)}}$)	The i -th miRNA (or disease) feature matrix
$k_m^{(i)}$ (or $k_d^{(i)}$)	Dimension of the i -th feature matrix
$H_m^{(i)} \in \mathbb{R}^{n \times k_m^{(i)}}$ (or $H_d^{(i)} \in \mathbb{R}^{m \times k_d^{(i)}}$)	The projection matrix of the i -th miRNA (or disease) feature
$\alpha^{(i)}, \lambda, \mu$ and η	Weight coefficient and regularization parameters
$G_m \in \mathbb{R}^{m \times n}$ (or $G_d \in \mathbb{R}^{n \times m}$)	The common latent interaction matrix in miRNA (or disease) view
$G \in \mathbb{R}^{m \times n}$	The predicted miRNA-disease association matrix

disease spaces. Finally, the entities in matrix X can be updated by taking the average of the updated interaction likelihood profiles.

In the miRNA perspective, to combine different features of miRNAs and projected those corresponding feature matrices $\{F_m^{(i)}\}_{i=1}^M$ into a common latent interaction subspace, we adopt a linear transformation to achieve this goal by using $F_m^{(i)}(H_m^{(i)})^T$, where $H_m^{(i)} \in \mathbb{R}^{n \times k_m^{(i)}}$ is the projection matrix of the i -th miRNA feature matrix, $i = 1, 2, \dots, M$. Then, the common latent interaction matrix $G_m \in \mathbb{R}^{m \times n}$ of miRNA view can be approximated by $F_m^{(i)}(H_m^{(i)})^T$, which can be written as $\|F_m^{(i)}(H_m^{(i)})^T - G_m\|_F^2, i = 1, 2, \dots, M$, where $\|\cdot\|_F$ is the Frobenius norm. Meanwhile, the matrix G_m is also encouraged to approximate to the original observed association matrix X , and thus we can mathematically formulate the following objective function:

$$\begin{aligned} \min_{G_m, H_m^{(i)}} \sum_{i=1}^M \left\| F_m^{(i)}(H_m^{(i)})^T - G_m \right\|_F^2 + \lambda \|G_m - X\|_F^2 \\ \text{s.t. } H_m^{(i)} \geq 0 \end{aligned} \quad (11)$$

where $\lambda \geq 0$ is a regularization parameter.

It has been shown that the intrinsic characteristics of different data spaces usually provide complementary or underlying correlated information and the graph regularization can be used to preserve the local structures of data spaces [27]. Here, to keep the geometric structures for the miRNA similarity profiles $\{S_m^{(i)}\}_{i=1}^N$ and maintain the similarity or correlation relationships of miRNAs, a non-negative weight vector $\alpha = [\alpha^{(1)}, \alpha^{(2)}, \dots, \alpha^{(N)}]$ is introduced, and we define a novel joint graph regularization as follows:

$$\begin{aligned} \sum_{i=1}^N \alpha^{(i)} \text{tr} \left(G_m^T (D_m^{(i)} - S_m^{(i)}) G_m \right) = \text{tr} \left(G_m^T \left(\sum_{i=1}^N \alpha^{(i)} L_m^{(i)} \right) G_m \right) \\ \text{s.t. } \sum_{i=1}^N \alpha^{(i)} = 1, \alpha^{(i)} \in [0, 1] \end{aligned} \quad (12)$$

where $L_m^{(i)} = D_m^{(i)} - S_m^{(i)}$ is the graph Laplacian matrix of the i -th miRNA similarity profile, and $(D_m^{(i)})_{ii} = \sum_{j=1}^m (S_m^{(i)})_{ij}$ is a diagonal matrix; the weight vector α is used to control the contributions of different similarity profiles.

To efficiently integrate multi-modal latent features and diverse similarity profiles of miRNA view, we can reformulated the objective function by combining the function in Equation

(11) and the regularization term of Equation (12) as follows:

$$\begin{aligned} \min_{G_m, H_m^{(i)}, \alpha^{(i)}} \sum_{i=1}^M \left\| F_m^{(i)}(H_m^{(i)})^T - G_m \right\|_F^2 + \lambda \|G_m - X\|_F^2 \\ + \mu \text{tr} \left(G_m^T \left(\sum_{i=1}^N \alpha^{(i)} L_m^{(i)} \right) G_m \right) \\ \text{s.t. } H_m^{(i)} \geq 0, \sum_{i=1}^N \alpha^{(i)} = 1, \alpha^{(i)} \in [0, 1] \end{aligned} \quad (13)$$

where μ is a parameter that control graph regularization term of miRNA space.

For the above objective function of Equation (13), the solution of α is $\alpha^{(i)} = 1$ when $\text{tr}(G_m^T L_m^{(i)} G_m)$ is the minimum one over the N items, and other entries in the weight vector $\alpha = \{\alpha^{(1)}, \alpha^{(2)}, \dots, \alpha^{(N)}\}$ are all zeros. To conquer this limitation and ensure that all graph Laplacian items could contribute effectively for the maintaining of graph local structures, we adopt a self-adaptive parameter adjustment strategy to guarantee that each similarity profile contributes to the learning task and transform the third term in Equation (13) into a non-linear programming problem by enforcing parameter $\mu > 1$ as the exponent of $\alpha^{(i)}$. Then, Equation (13) becomes:

$$\begin{aligned} \min_{G_m, H_m^{(i)}, \alpha^{(i)}} \sum_{i=1}^M \left\| F_m^{(i)}(H_m^{(i)})^T - G_m \right\|_F^2 + \lambda \|G_m - X\|_F^2 \\ + \text{tr} \left(G_m^T \left(\sum_{i=1}^N (\alpha^{(i)})^\mu L_m^{(i)} \right) G_m \right) \\ \text{s.t. } H_m^{(i)} \geq 0, \sum_{i=1}^N \alpha^{(i)} = 1, \alpha^{(i)} \in [0, 1] \end{aligned} \quad (14)$$

In addition, the elements of projection matrix $H_m^{(i)} \in \mathbb{R}^{n \times k_m^{(i)}}$ in Equation (13) describe the weights of miRNA features. To alleviate the influence of inherent noise in miRNA space and improve the interpretability for the prediction results, we also impose the $L_{p,q}$ mixed-norm on the projection matrices. The $L_{p,q}$ mixed-norm could help to obtain the most representative and discriminative sparse features [47], and the definition of $L_{p,q}$ norm is as follows:

$$\|H_m^{(i)}\|_{p,q} = \left(\sum_{x=1}^n \left(\sum_{y=1}^{k_m^{(i)}} (H_m^{(i)})_{xy}^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \quad (15)$$

Here, as done in Qiu et al. [47], we utilize $L_{1,2}$ -norm to regularize the sparsity of the projection matrix $H_m^{(i)}$. Finally, the objective function of M2LFL is mathematically formulated as follows:

$$\begin{aligned} \min_{G_m, H_m^{(i)}, \alpha^{(i)}} & \sum_{i=1}^M \left\| F_m^{(i)} (H_m^{(i)})^T - G_m \right\|_F^2 + \lambda \|G_m - X\|_F^2 \\ & + \text{tr} \left(G_m^T \left(\sum_{i=1}^N (\alpha^{(i)})^\mu L_m^{(i)} \right) G_m \right) + \eta \sum_{i=1}^M \left\| H_m^{(i)} \right\|_{1,2}^2 \\ \text{s.t. } & H_m^{(i)} \geq 0, \sum_{i=1}^N \alpha^{(i)} = 1, \alpha^{(i)} \in [0, 1] \end{aligned} \quad (16)$$

where η is a regularization coefficient to control the sparse constraint for $H_m^{(i)}, i = 1, 2, \dots, M$. The first term is the subspace regression term, which was used to combine different features of miRNAs (or diseases) and projected those corresponding feature matrices $\{F_m^{(i)}\}_{i=1}^M$ into a common latent interaction subspace. The second term makes the predicted association matrix G_m approximate to the original observed association matrix X . The third term is a joint graph Laplacian regularization, which could fully exploit complementation information in different similarity profiles to boost the performance. The detailed optimization procedures of the above objective function are provided in the following section.

Optimization

To solve the optimization problem in Equation (16), we present an efficient alternating strategy to obtain the optimal solution by updating $G_m, H_m^{(i)}$ and $\alpha^{(i)}$ iteratively. The solution processes are shown as follows:

Fix $H_m^{(i)}$ and $\alpha^{(i)}$, and solve for G_m . The optimization problem reduces to the following sub-problem for M_m :

$$\begin{aligned} \min_{G_m} & \sum_{i=1}^M \left\| F_m^{(i)} (H_m^{(i)})^T - G_m \right\|_F^2 + \lambda \|G_m - X\|_F^2 \\ & + \text{tr} \left(G_m^T \left(\sum_{i=1}^N (\alpha^{(i)})^\mu L_m^{(i)} \right) G_m \right) \end{aligned} \quad (17)$$

By setting the derivative of Equation (17) respected to G_m to zero. Then, the updating rule for G_m can be obtained as:

$$G_m = \left(\sum_{i=1}^N (\alpha^{(i)})^\mu L_m^{(i)} + (M + \lambda) I \right)^{-1} \left(\lambda X + \sum_{i=1}^M F_m^{(i)} (H_m^{(i)})^T \right) \quad (18)$$

Fix G_m and $\alpha^{(i)}$, and solve for $H_m^{(i)}$. Accordingly, when G_m and $\alpha^{(i)}$ are fixed, the sub-problem of the objective function in Equation (16) can be rewritten as:

$$\begin{aligned} \min_{H_m^{(i)}} & \sum_{i=1}^M \left\| F_m^{(i)} (H_m^{(i)})^T - G_m \right\|_F^2 + \eta \sum_{i=1}^M \left\| H_m^{(i)} \right\|_{1,2}^2 \\ \text{s.t. } & H_m^{(i)} \geq 0 \end{aligned} \quad (19)$$

To optimize Equation (19), we introduce the Lagrange multiplier $\varphi^{(i)}$ for the constraint of $H_m^{(i)}$, and thus the sub-problem is formulated as:

$$\begin{aligned} f(H_m^{(i)}, \varphi^{(i)}) & = \sum_{i=1}^M \left\| F_m^{(i)} (H_m^{(i)})^T - G_m \right\|_F^2 + \eta \sum_{i=1}^M \left\| H_m^{(i)} \right\|_{1,2}^2 \\ & - \text{tr} \left(\varphi^{(i)} \sum_{i=1}^M H_m^{(i)} \right) \end{aligned} \quad (20)$$

Then, the partial derivative of $f(H_m^{(i)}, \varphi^{(i)})$ with respect to $G_m^{(i)}$ is:

$$\frac{\partial f(H_m^{(i)}, \varphi^{(i)})}{\partial H_m^{(i)}} = 2H_m^{(i)} (F_m^{(i)})^T F_m^{(i)} - 2G_m^T F_m^{(i)} + 2\eta H_m^{(i)} e e^T - \varphi^{(i)} \quad (21)$$

When employing the KKT condition $(\varphi^{(i)})_{jk} (H_m^{(i)})_{jk} = 0$, and setting the derivative to zero, we have:

$$(H_m^{(i)})_{jk} = (H_m^{(i)})_{jk} \odot \frac{(G_m^T F_m^{(i)})_{jk}}{(H_m^{(i)}) \left((F_m^{(i)})^T F_m^{(i)} + \eta e e^T \right)_{jk}} \quad (22)$$

where \odot denotes the Hadamard product; $e \in \{1\}^{k_m^{(i)} \times 1}$ is a vector and $k_m^{(i)}$ represents the number of features in feature matrix $F_m^{(i)}$. To ensure that all elements of $H_m^{(i)}$ are non-negative, the updating rule for $H_m^{(i)}$ can be rewritten as:

$$(H_m^{(i)})_{jk} = (H_m^{(i)})_{jk} \odot \frac{\left(H_m^{(i)} \left((F_m^{(i)})^T F_m^{(i)} + \eta e e^T \right)^+ + (G_m^T F_m^{(i)})^- \right)_{ij}}{\sqrt{\left(H_m^{(i)} \left((F_m^{(i)})^T F_m^{(i)} + \eta e e^T \right)^- + (G_m^T F_m^{(i)})^+ \right)_{ij}}} \quad (23)$$

where the matrices with negative and positive symbols are defined as $A_{jk}^- = (|A_{jk}| - A_{jk})/2$ and $A_{jk}^+ = (|A_{jk}| + A_{jk})/2$.

Fix G_m and $H_m^{(i)}$, and solve for $\alpha^{(i)}$. By fixing G_m and $H_m^{(i)}$, the objective function in Equation (16) can be simplified as:

$$\begin{aligned} \min_{\alpha^{(i)}} & \text{tr} \left(G_m^T \left(\sum_{i=1}^N (\alpha^{(i)})^\mu L_m^{(i)} \right) G_m \right) \\ \text{s.t. } & \sum_{i=1}^N \alpha^{(i)} = 1, \alpha^{(i)} \in [0, 1] \end{aligned} \quad (24)$$

Similarly, the Lagrange function of Equation (24) is:

$$f(\alpha^{(i)}, \xi) = \text{tr} \left(G_m^T \left(\sum_{i=1}^N (\alpha^{(i)})^\mu L_m^{(i)} \right) G_m \right) - \xi \left(\sum_{i=1}^N \alpha^{(i)} - 1 \right) \quad (25)$$

where ξ is the Lagrange multiplier for the constraint of $\sum_{i=1}^N \alpha^{(i)} = 1$.

By setting the derivative with respect to $\alpha^{(i)}$ and ξ to 0, we have:

$$\begin{cases} \mu (\alpha^{(i)})^{\mu-1} \text{tr} \left(G_m^T L_m^{(i)} G_m \right) - \xi = 0 \\ \sum_{i=1}^N \alpha^{(i)} - 1 = 0 \end{cases} \quad (26)$$

Then, the updating rule for $\alpha^{(i)}$ is obtained as follows:

$$\alpha^{(i)} = \left(\sum_{i=1}^N \left(\frac{1}{\text{tr} \left(G_m^T L_m^{(i)} G_m \right)} \right)^{\frac{1}{\mu-1}} \right)^{-1} \left(\frac{1}{\text{tr} \left(G_m^T L_m^{(i)} G_m \right)} \right)^{\frac{1}{\mu-1}} \quad (27)$$

Thus, we can update $G_m, H_m^{(i)}$ and $\alpha^{(i)}$ by using the rules of Equations (18), (23) and (27) alternately until convergence. Since the model formulation and optimization procedures in both miRNA and disease spaces are similar, as shown in Figure 2, the objective function of disease perspective also can be formulated and solved based on the known association information $X^T \in \mathbb{R}^{n \times m}$, disease features $F' = \{F_d^{(i)} \in \mathbb{R}^{n \times k_d^{(i)}}\}_{i=1}^M$ as well as disease similarity profiles $S' = \{S_d^{(i)} \in \mathbb{R}^{n \times n}\}_{i=1}^M$. Finally, considering that the proposed method already contains several parameters that need to be tuned, for simplicity, we combine the predicted results of both miRNA and disease views with an average weighting

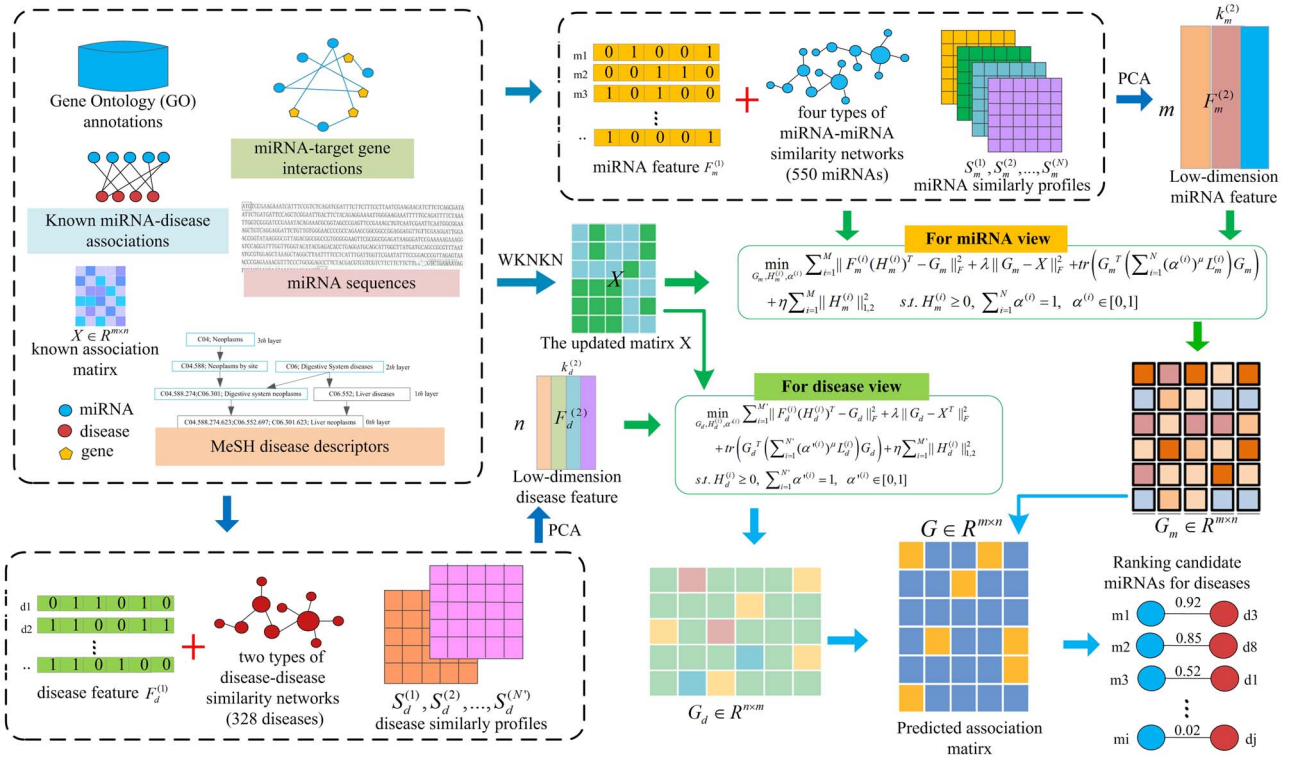


Figure 2. The workflow of M2FL for inferring potential disease-associated miRNAs.

strategy as $G = (G_m + (G_d)^T)/2$ to prioritize disease-associated miRNAs, and the values of the entities in matrix G stand for the pairwise correlation scores between miRNAs and diseases. Algorithm 1 summarizes the procedure of M2FL to infer candidate disease miRNAs.

Algorithm 1 M2FL Algorithm

Input: miRNA similarity profiles $\{S_m^{(1)}, S_m^{(2)}, \dots, S_m^{(N)}\}$, miRNA feature matrices $\{F_m^{(1)}, F_m^{(2)}, \dots, F_m^{(N)}\}$, disease similarity profiles $\{S_d^{(1)}, S_d^{(2)}, \dots, S_d^{(N)}\}$, disease feature matrices $\{F_d^{(1)}, F_d^{(2)}, \dots, F_d^{(N)}\}$, known miRNA-disease association matrix $X \in R^{m \times n}$, parameters λ , μ , and η

Output: probability matrix G

1. Initialize $\{H_m^{(l)} \in R^{m \times m}\}_{l=1}^M$ and $\{H_d^{(l)} \in R^{n \times n}\}_{l=1}^M$ with random values in the interval $[0,1]$
2. Initialize the values of $\{\alpha^{(l)}\}_{l=1}^M$ and $\{\alpha^{(l)}\}_{l=1}^M$ as $\frac{1}{M}$ and $\frac{1}{N}$, respectively
3. **for** each miRNA $m_i \in T_m$ and each disease $d_j \in T_d$ **do**
4. update their interaction profiles with WKNKN by Eq.(10)
5. **end for**
6. $X_{mn} = (a_i X_{ij} + a_j X_{ji}) / \sum_{l=1}^2 a_l$
7. update the entities in association matrix X by $X_j = \max(X_j, (X_{mn})_j)$
8. **repeat** // In miRNA view
9. update G_m by Eq.(18) with fixing $\{H_m^{(l)}\}_{l=1}^M$ and $\{\alpha^{(l)}\}_{l=1}^M$
10. **for** $i \leftarrow 1$ to M **do**
11. update $H_m^{(i)}$ by Eq.(23) with fixing G_m and $\alpha^{(i)}$
12. **end for**
13. **for** $i \leftarrow 1$ to N **do**
14. update $\alpha^{(i)}$ by Eq.(27) with fixing G_m and $H_m^{(i)}$
15. **end for**
16. **until convergence**
17. **repeat** // In disease view
18. update G_d , $H_d^{(l)}$ and $\alpha^{(l)}$ for disease space by the corresponding rules
19. **until convergence**
20. obtain matrix G according to $G = (G_m + (G_d)^T)/2$
21. **return** G .

Results and discussion

Experiment settings

In this study, we perform 5-fold cross validation (5CV) to comprehensively investigate the effectiveness of M2FL for inferring potential disease-associated miRNAs. In each fold, all of the

known miRNA-disease associations in the benchmark dataset are randomly divided into five equal size subsets, four of them are used for model training and the remaining subset is adopted for testing. These known associations in the testing subset are regarded as positive samples, and those unobserved ones without experimental validation are regarded as negative samples. As aforementioned, we have collected 6088 known associations between 550 miRNAs and 328 diseases from HMDD v2.0, and the ratio between observed associations and unobserved associations is about 1–30. Since the positive samples are much less than negative samples, as done in many imbalanced data classes [29], the area under the precision-recall curve (AUPR) is used as the primary metric, which could punish false positive more in performance evaluation. Meanwhile, the area under ROC curve (AUC) is also adopted to estimate the global performance of different computational models in screening disease-associated miRNAs. The precision-recall and receiver operating characteristic (ROC) curves can be drawn based on the precisions and recalls as well as the true positive rates (TPRs) and false positive rates (FPRs) with various thresholds, respectively. The precision and recall are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (28)$$

while the TPR and FPR can be represented as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (29)$$

where TP, TN, FP and FN denote true positive, true negative, false positive and false negative, respectively. In addition, several common metrics, including precision, sensitivity, accuracy,

Table 3. Results of paired t-test on the AUPRs and AUCs between M2LFL and other methods

M2LFL versus		IMCMDA	GRNMF	MKRMDA	BRWH	MRSLA
P-value	AUPR	1.3957e-26	3.4240e-27	1.0766e-27	2.2908e-32	7.6507e-29
	AUC	1.1798e-24	1.8652e-19	5.0465e-23	8.2798e-29	7.3957e-23

specificity and F1-measure, are also adopted to make a comparison for measuring the performance.

Performance evaluation

In order to systematically evaluate the performance of M2LFL, we compare the proposed method with other state-of-the-art computational models, including IMCMDA [61], GRNMF [27], MKRMDA [62], BRWH [20] and MRSLA [29]. Specifically, IMCMDA is an inductive matrix completion-based method for discovering disease miRNAs [61]. GRNMF prioritizes disease-associated miRNAs with graph regularized non-negative matrix factorization [27], and MKRMDA is a multiple kernel learning-based method for predicting potential miRNA-disease associations [62]. BRWH is a method for disease miRNA prediction by using unbalanced bi-random walk [20], and MRSLA prioritizes candidate disease miRNAs based on multi-view manifold regularized learning [29]. The comparison between M2LFL and other compared methods is carried out based on the benchmark dataset by 5CV. We perform 20 runs of 5CV for each model and obtain the averaging results for a convincing evaluation. Those miRNA and disease similarities or features that depend on the known miRNA-disease association information are reacquired in each iteration of cross validation. To make a fair comparison, we set the parameters of these existing state-of-the-art methods to their optimal values that were recommended by their authors. Moreover, since some compared methods utilize earlier versions of databases for prediction, here we replace these data sources with the same version datasets used in M2LFL, which can help to improve the performance for the compared methods.

As shown in Figures 3 and 4, the overall performance of M2LFL is superior to other compared methods according to the PR and ROC curves. The average AUPR values of M2LFL, IMCMDA, GRNMF, MKRMDA, BRWH and MRSLA are 0.590, 0.230, 0.279, 0.247, 0.035 and 0.156, respectively. M2LFL achieves the best performance, and it outperforms IMCMDA by 36%, GRNMF by 31.1%, MKRMDA by 34.3%, BRWH by 58.9% and MRSLA by 43.4% in terms of AUPR under 5CV. The AUC of M2LFL is 0.969, as shown in Figure 4, which makes 6.8, 3.8, 7.8, 15 and 6.1% improvements compared with other methods (IMCMDA: 0.901, GRNMF: 0.931, MKRMDA: 0.891, BRWH: 0.819 and MRSLA: 0.908), respectively. In particular, we also perform a statistical analysis of paired t-tests to evaluate the difference between M2LFL and other methods. We implement 20 independent runs of 5CV for all computational methods, and then measure the significance difference by using the function t-test (X, Y) in the Matlab software. As shown in Table 3, the paired t-test results show that the proposed method achieves significantly ($P < 0.05$) better performance over these compared methods in terms of AUPR and AUC.

In addition, as shown in Table 4, the comparison of several other common metrics (i.e. precision, sensitivity, accuracy, specificity and F1-measure) further demonstrates the superior performance of M2LFL in prioritizing candidate disease miRNAs. To be more specific, M2LFL obtains the highest precision of 0.663, and its precision value is 37.9, 34.2, 35.9, 66.2 and 47.6% better than IMCMDA, GRNMF, MKRMDA, BRWH and MRSLA, respectively.

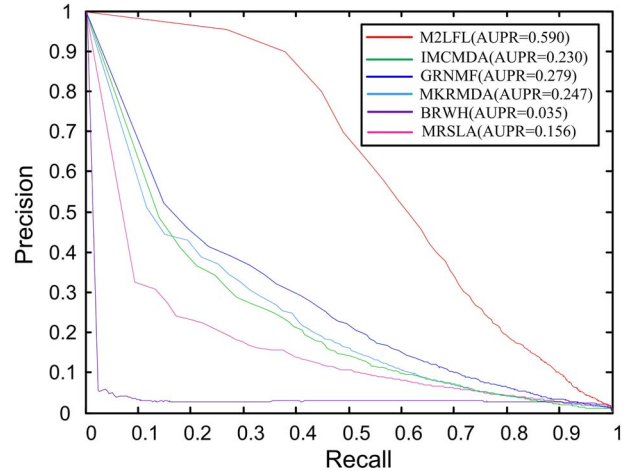


Figure 3. Comparison of M2LFL and other models in terms of PR curves under 5CV.

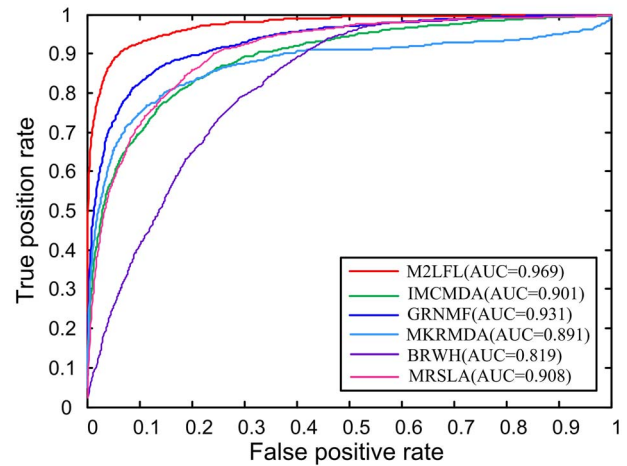
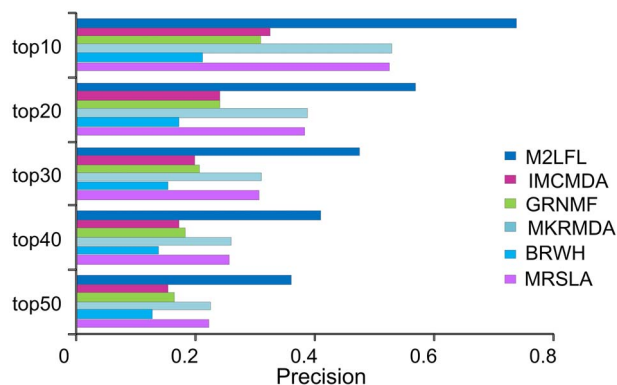


Figure 4. Comparison of M2LFL and other models in terms of ROC curves under 5CV.

Meanwhile, M2LFL also produces the best performance in terms of sensitivity and reaches the value of 0.499, which outperforms IMCMDA by 17.6%, GRNMF by 15.0%, MKRMDA by 18.4%, BRWH by 27.1% and MRSLA by 22.7%. In particular, the proposed method obtains a significant improvement of F1-score, and the F1-scores of M2LFL, IMCMDA, GRNMF, MKRMDA, BRWH and MRSLA are 0.568, 0.302, 0.334, 0.309, 0.057 and 0.220, respectively. In addition, M2LFL is also consistently better than the other compared methods in terms of accuracy and specificity, and the values of M2LFL are 0.995 and 0.998, while the corresponding values of the second best models (GRNMF and MKRMDA) are 0.990 and 0.995, respectively. These results confirm the effectiveness and robustness of the proposed method, which suggest that M2LFL could infer disease-associated miRNAs with a higher confidence than other computational models.

Table 4. Performance of different metrics of different methods

Method	AUPR	AUC	Precision	Sensitivity	Accuracy	Specificity	F1
M2LFL	0.590	0.969	0.663	0.499	0.995	0.998	0.568
IMCMDA	0.230	0.901	0.284	0.326	0.987	0.993	0.302
GRNMF	0.279	0.931	0.321	0.349	0.990	0.995	0.334
MKRMDA	0.247	0.891	0.304	0.315	0.990	0.995	0.309
BRWH	0.035	0.819	0.033	0.228	0.948	0.953	0.058
MRS�A	0.156	0.908	0.187	0.272	0.987	0.992	0.220

**Figure 5.** Comparison of precisions for M2LFL and other models under various top k cutoffs.

The precision within top-k ranking list means reliability of discovered disease miRNA candidates. As shown in Figure 5, the results also show that M2LFL consistently outperforms MCMDA, GRNMF, MKRMDA, BRWH and MRS�A in terms of precision from top 10 to top 50. M2LFL produces the highest precision under different top k cutoffs and obtains 74.7% in top 10, 57.7% in top 20, 48.2% in top 30, 41.6% in top 40 and 36.5% in top 50. MKRMDA achieves the second highest precisions, while BRWH performs the worst precisions at various top k rankings. This may be due to the sparseness of the known association network and BRWH is a network-based method. The precision values of the former are 52.5, 38.4, 30.8, 25.9 and 22.5% in top 10 to top 50, respectively, and the corresponding precisions of the latter are 21.3% in top 10, 17.3% in top 20, 15.5% in top 30, 13.9% in top 40 and 12.9% in top 50. All the above comparison results demonstrate that M2LFL exhibits superior prediction performance compared with the other methods in screening the reliable candidate miRNAs for diseases, suggesting that the prediction accuracy can be improved by fully capturing the multi-modal miRNA and disease latent features.

Parameter analysis

In the experiment, the influences of parameters on the performance are estimated by 5CV. As shown in Equation (16), there are three regularization coefficients (i.e. λ , μ and η) are introduced in the objective function to balance the contributions of different terms. Here, the optimal parameter combination is determined from the following values: $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^1, 10^2\}$ for λ , $\{2^{-2}, 2^{-1}, 2^1, 2^2, 2^3, 2^4\}$ for μ and $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^1, 10^2\}$ for η . We analyze all the combinations with a grid-search strategy and choose the most suitable hyper-parameters for predicting candidate disease miRNAs. Here, we fix the value of one parameter and analyze the effects of the

remaining two parameters on the prediction performance. As shown in Figure 6, we fix parameter λ and found that a greater value of μ is more likely to obtain better performance, M2LFL reaches the lowest AUPR when $\mu = 2^{-2}$ and produces better result as the value of μ increases from 2^{-2} to 2^2 . Meanwhile, the AUPR of M2LFL improves as parameter η decreases to 10^{-3} . M2LFL is inclined to use a large value for μ and a small value for η . In addition, when analyzing the influence of parameter λ on the performance, we found that a larger value of the parameter λ tends to produce better results for M2LFL, and the performance increases slightly after reaching a peak. After implementing a series of experiments, M2LFL achieves the best AUPR when $\lambda = 10^2$, $\mu = 2^2$ and $\eta = 10^{-3}$. Moreover, to extract the dimensions of the latent features for miRNAs and diseases, we determine the parameter values of δ_m and δ_d in Equations (7) and (8) from the range of $\{0.1, 0.2, 0.3, \dots, 0.9\}$. As shown in Figure 7, the best AUPR of M2LFL is reached when $\delta_m = 0.7$ and $\delta_d = 0.7$. Finally, we set $\lambda = 10^2$, $\mu = 2^2$, $\eta = 10^{-3}$, $\delta_m = 0.7$ and $\delta_d = 0.7$ for M2LFL to discover disease-associated miRNAs in the following studies.

As aforementioned, we adopt a self-adaptive parameter learning strategy to ensure that all graph Laplacian items or similarity profiles could contribute to the disease-related miRNA prediction task. To investigate the effectiveness of the self-adaptive learning on performance, we perform an experiment to set the contribution values $\{\alpha^{(i)}\}_{i=1}^N$ and $\{\alpha'^{(i)}\}_{i=1}^{N'}$ in both miRNA and disease spaces as mentioned in Algorithm 1 to fixed values instead of automatically updating the weights. Here, we carried out the experiment with several representative settings for $\{\alpha^{(i)}\}_{i=1}^N$ and $\{\alpha'^{(i)}\}_{i=1}^{N'}$, and the results (Table 5) show that the AUPR of M2LFL with an adaptive parameter learning strategy is higher than that of in other fixed weight scenarios under 5CV, which suggest the effectiveness of the self-adaptive learning for M2LFL in predicting disease-associated miRNAs.

Case studies: hepatocellular carcinoma and renal cell carcinoma

We also perform case studies for several popular human diseases to illustrate the ability of M2LFL in discovering disease-related miRNAs. We measure the performance of M2LFL by observing the number of experimentally validated candidate miRNAs for these diseases in various top-ranked predictions. Here, the known associations related to each disease in HMDD v2.0 are used for training the model, and all miRNAs are sorted in descending order according to their predicted probability scores, which suggest that how likely they would correlate with the test diseases. For all diseases, the percentages of successfully retrieved miRNA-disease associations at different top k cutoffs are given in Figure 8. It can be observed that M2LFL recovers high rates of known associations in different top k thresholds. For example, among all of the 6088 known associations, M2LFL correctly retrieved 31.3% (1904/6088) and 56.7% (3450/6088) in

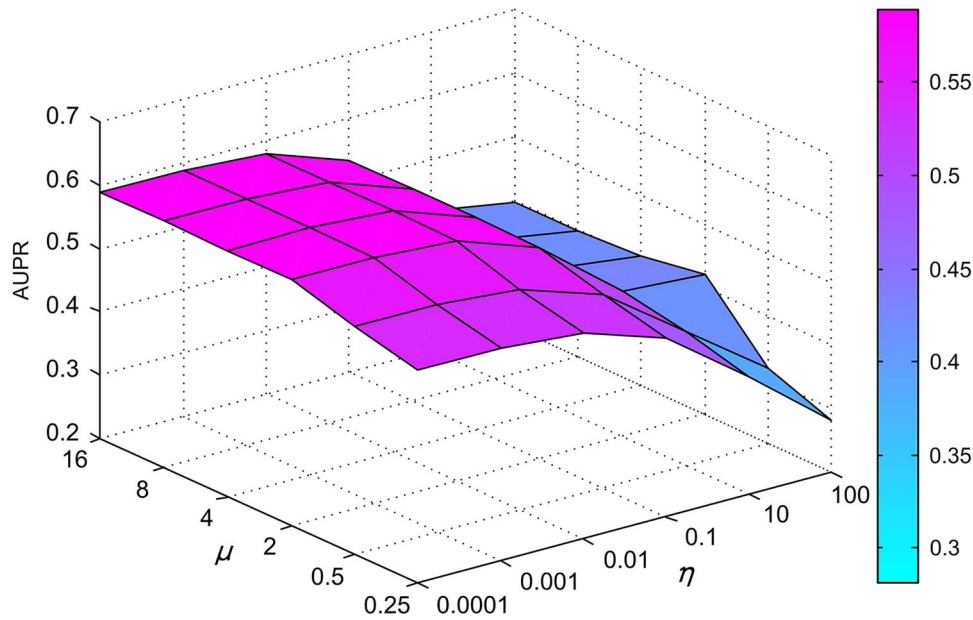


Figure 6. The influence of parameters μ and η on the performance of M2LFL under 5-fold cross validation.

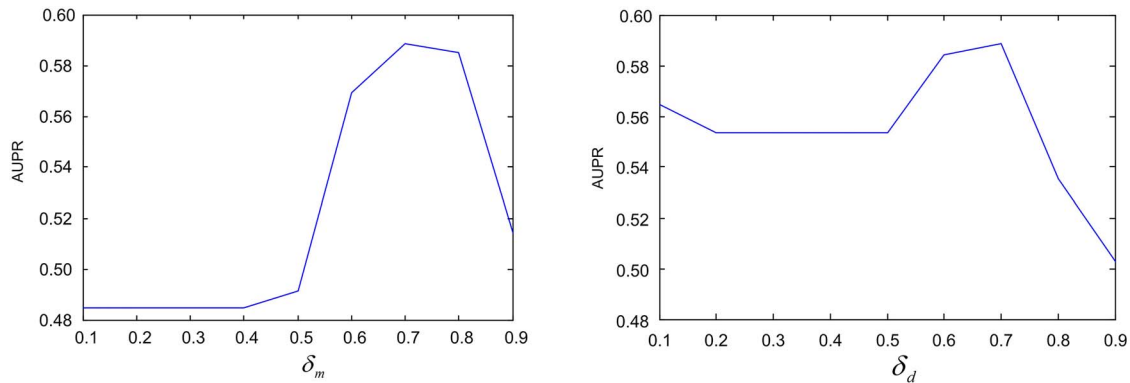


Figure 7. The influence of parameters δ_m and δ_d on the performance of M2LFL under 5-fold cross validation.

Table 5. Performance of M2LFL with different fixed values of $\alpha^{(i)}$ and $\alpha'^{(i)}$

Number of scenario	Experimental settings		AUPR	AUC
	miRNA space	Disease space		
Scenario 1	$\alpha^{(1)} = \alpha^{(2)} = \alpha^{(3)} = \alpha^{(4)} = 0.25$	$\alpha'^{(1)} = \alpha'^{(2)} = 0.5$	0.576	0.968
Scenario 2	$\alpha^{(1)} = \alpha^{(2)} = \alpha^{(3)} = \alpha^{(4)} = 0.25$	$\alpha'^{(1)} = 0, \alpha'^{(2)} = 1$	0.540	0.963
Scenario 3	$\alpha^{(1)} = \alpha^{(2)} = \alpha^{(3)} = \alpha^{(4)} = 0.25$	$\alpha'^{(1)} = 1, \alpha'^{(2)} = 0$	0.566	0.967
Scenario 4	$\alpha^{(1)} = 1, \alpha^{(2)} = \alpha^{(3)} = \alpha^{(4)} = 0$	$\alpha'^{(1)} = 1, \alpha'^{(2)} = 0$	0.482	0.955
Scenario 5	$\alpha^{(1)} = \alpha^{(2)} = 0.5, \alpha^{(3)} = \alpha^{(4)} = 0$	$\alpha'^{(1)} = \alpha'^{(2)} = 0.5$	0.524	0.961
Scenario 6	$\alpha^{(1)} = \alpha^{(2)} = \alpha^{(3)} = 0, \alpha^{(4)} = 1$	$\alpha'^{(1)} = \alpha'^{(2)} = 0.5$	0.506	0.958

top 10 and top 30, respectively. In addition, after excluding those observed miRNAs associated with specific diseases in the ranking lists, we further validate the top 20 predicted miRNA candidates for two representative diseases (hepatocellular carcinoma and renal cell carcinoma) by manually mining literatures and other public experimentally supported databases, namely, miRCancer [63], dbDEMC [64] and miR2Disease [65].

Hepatocellular carcinoma is one of the most common cancer and lead to a large number of deaths every year. Accumulating

evidence has been reported to indicate that miRNAs play an important role in the development of hepatocellular cancer [48]. With the implementation of M2LFL, all candidate miRNAs related to hepatocellular carcinoma are prioritized according to their scores. The top 20 miRNAs predicted by M2LFL are listed in Table 6. We can see that 18 of top 20 predicted miRNAs are confirmed to be related to the disease by experimentally supported databases and recently published literatures. For example, hsa-mir-367 [8] and hsa-mir-452 [66] promote hepatocellular

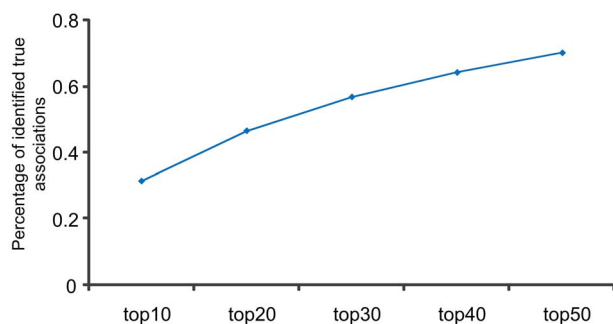


Figure 8. Percentage of true known associations identified by M2LFL with various ranking cutoffs.

cancer cell proliferation, and existing evidence indicates that they have great potential as new promising biomarkers for disease diagnosis. In addition, some other miRNAs including hsa-mir-143, hsa-mir-133a-1 and hsa-mir-193b have also been proved to be significantly downregulated [63], while hsa-mir-9-2, hsa-mir-367 and hsa-mir-429 are confirmed to be upregulated in hepatocellular cancer cell.

Renal cell carcinoma is a cancer that starts in the lining of very small tubes (tubules) in the kidney. Here, we also perform case studies involving renal cell carcinoma to estimate the prediction performance of M2LFL. Table 7 provides the top 20 predicted candidates of this cancer, and we can observe that most of the candidates are validated based on different evidences. For example, hsa-mir-20a, hsa-mir-125b-1, hsa-mir-222 and hsa-mir-31 within top 5 are validated to be involved in renal cell carcinoma by dbDEM. Meanwhile, researchers

also reported that miRNAs hsa-mir-17, hsa-mir-221 and hsa-mir-125b-1 show higher expression levels in renal cell patient tumor tissues compared with those of in normal controls [67, 68]. Existing evidence also show that some predicted miRNA candidates (e.g. hsa-mir-183, hsa-mir-221, hsa-mir-29b-1 and hsa-mir-29a) with high rankings may promote cell proliferation in renal cell carcinoma and function as new promising biomarkers for its diagnosis. In conclusion, the case studies further validate the capability of M2LFL in screening disease miRNA candidates with high confidence.

Inferring novel disease-associated miRNAs

After evaluating the performance of M2LFL with 5CV and case studies, we further verify the practical application of this computational model for discovering novel disease-associated miRNAs. In general, it is also effective to conduct validation by using different versions of the same database to confirm those actual potential of the discovery. Specifically, as done in previous studies [27], we train M2LFL by using all the observed associations in the older version database (i.e. HMDD v2.0) and validate the predicted disease miRNA candidates by the latest version of HMDD. As aforementioned, HMDD v2.0 is adopted as the benchmark dataset and 6088 known associations between 328 diseases and 550 miRNAs are retained. Currently, HMDD v3.2 was released, which contains 35 547 miRNA-disease association entries involving 893 diseases and 1206 miRNAs. Then, we collect those recently added associations that were not in the benchmark dataset to validate the prediction results. As expected, we found that most of potential disease miRNAs predicted by M2LFL within the top rankings are directly confirmed by the latest version dataset. As shown in Figure 9, it displays the percentages

Table 6. The top 20 candidate miRNAs identified by M2LFL for hepatocellular carcinoma

Ranking	miRNA	Evidence	Ranking	miRNA	Evidence
1	hsa-mir-143	dbDEM, miR2Disease, miRCancer	11	hsa-mir-27b	dbDEM, miRCancer
2	hsa-mir-34b	miRCancer	12	hsa-mir-429	miRCancer
3	hsa-mir-9-1	Unconfirmed	13	hsa-mir-452	miRCancer
4	hsa-mir-9-3	Unconfirmed	14	hsa-mir-193b	miRCancer
5	hsa-mir-23b	miR2Disease, miRCancer	15	hsa-mir-215	miRCancer
6	hsa-mir-9-2	miR2Disease	16	hsa-mir-137	miRCancer
7	hsa-mir-133a-1	miRCancer	17	hsa-mir-204	PMID:23282077
8	hsa-mir-133a-2	miRCancer	18	hsa-mir-132	miRCancer
9	hsa-mir-135b	PMID:26429530	19	hsa-mir-26b	dbDEM, miR2Disease, miRCancer
10	hsa-mir-367	miRCancer	20	hsa-mir-149	miRCancer

Table 7. The top 20 candidate miRNAs identified by M2LFL for renal cell carcinoma

Ranking	miRNA	Evidence	Ranking	miRNA	Evidence
1	hsa-mir-17	PMID:25011053	11	hsa-mir-29a	dbDEM
2	hsa-mir-20a	dbDEM	12	hsa-mir-29b-1	dbDEM
3	hsa-mir-125b-1	miRCancer	13	hsa-mir-19b-1	PMID:20964835
4	hsa-mir-222	dbDEM	14	hsa-mir-25	dbDEM
5	hsa-mir-31	dbDEM	15	hsa-mir-182	dbDEM, miRCancer
6	hsa-mir-183	dbDEM, miRCancer	16	hsa-mir-1-1	Unconfirmed
7	hsa-mir-221	miRCancer	17	hsa-mir-96	miRCancer
8	hsa-mir-18a	dbDEM	18	hsa-mir-181a-1	dbDEM
9	hsa-mir-15a	dbDEM, miRCancer	19	hsa-mir-106a	dbDEM, miRCancer
10	hsa-mir-125b-2	miRCancer	20	hsa-mir-181b-1	Unconfirmed

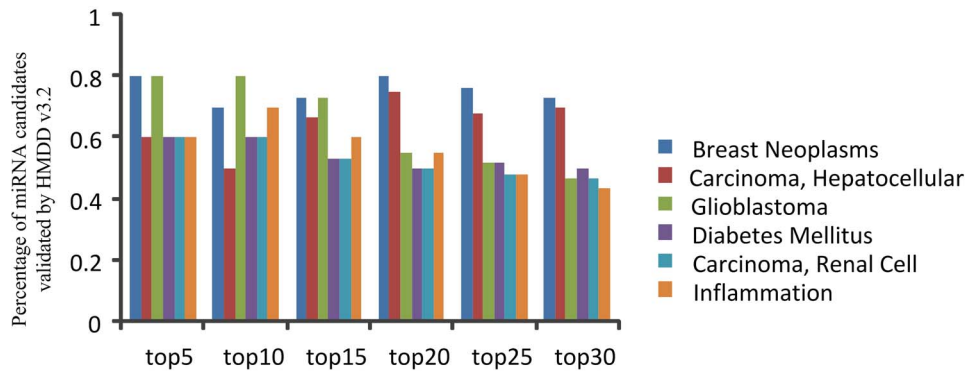


Figure 9. Percentage of candidate miRNAs validated by HMDD v3.2 for different ranking thresholds.

Table 8. The top 20 candidate miRNAs identified by M2LFL for breast neoplasms

Ranking	miRNA	Evidence	Ranking	miRNA	Evidence
1	hsa-mir-142	HMDD v3.2	11	hsa-mir-15b	HMDD v3.2
2	hsa-mir-150	HMDD v3.2	12	hsa-mir-186	Unconfirmed
3	hsa-mir-106a	HMDD v3.2	13	hsa-mir-185	HMDD v3.2
4	hsa-mir-138-1	Unconfirmed	14	hsa-mir-196b	HMDD v3.2
5	hsa-mir-99a	HMDD v3.2	15	hsa-mir-98	HMDD v3.2
6	hsa-mir-138-2	Unconfirmed	16	hsa-mir-212	HMDD v3.2
7	hsa-mir-130a	HMDD v3.2	17	hsa-mir-542	HMDD v3.2
8	hsa-mir-19b-2	Unconfirmed	18	hsa-mir-372	HMDD v3.2
9	hsa-mir-378a	HMDD v3.2	19	hsa-mir-92b	HMDD v3.2
10	hsa-mir-192	HMDD v3.2	20	hsa-mir-30e	HMDD v3.2

of predicted miRNAs validated by HMDD v3.2 for several diseases under different ranking thresholds, and the results show that M2LFL can effectively discover novel disease-associated miRNAs. For example, as shown in Table 8, 16 miRNAs in top 20 are proved to be involved in breast neoplasms by HMDD v3.2. The above observations imply that M2LFL has potential to be a promising computational tool to discover prognostic miRNAs for guiding laborious clinical trials, which could help to explore the pathogenesis of diseases at the level of miRNAs.

Conclusions

In this study, we have systematically studied the problem of discovering unobserved associations between miRNAs and diseases, and presented a method of M2LFL to prioritize candidate disease miRNAs. M2LFL fully integrates multiple complementary data sources with different representations and captures latent features from miRNA and disease spaces, and then aims to identify those most promising disease-associated miRNAs for further biological investigation. An adaptive M2LFL framework is formulated to obtain those miRNA candidates for specific diseases by utilizing the known miRNA-disease associations, multiple similarity profiles and multi-modal latent features of miRNAs and diseases. In addition, to capture different manifold characteristics and structure information of data spaces, the adaptive joint graph regularization and mixed-norm constraint terms are incorporated into this model to contribute to the learning task. The results indicate that M2LFL achieves comparable performance under different metrics and can be effectively adopted to discover disease-related miRNAs. The case studies further confirm the promising capability of M2LFL in discovering the reliable candidate miRNAs for human diseases, which is helpful to guide laborious clinical biological experiments.

However, M2LFL still has some limitations that require further research. First, the similarity measurement and feature extraction strategies in miRNA and disease spaces might not be optimal, and many other data can offer beneficial feature representations. Second, it is challenge and time consuming to determine the optimal parameters in different circumstances. On the other hand, with the rapid development of sequencing technology, more and more data sources are available, and how to integrate different biological data sources more reasonably deserves further study. What is more, combining linear and non-linear features and learning more data-consistent representations may provide a way to boost the performance.

Key Points

- In silico identification of promising diagnostic biomarkers is critical for precision medicine.
- Large and diverse biomedical data create unprecedented opportunities to decipher biological problems.
- An integrative framework is developed to infer disease miRNAs with adaptive multi-source multi-view learning.
- Experimental validations demonstrated its effectiveness in prioritizing disease-associated miRNAs.

Funding

This work was supported by the National Natural Science Foundation of China under (grant numbers: 61873089, 61976089, 61473259, 61472133), the Hunan Provincial Science and Technology Project Foundation under (grant numbers:

2018TP1018, 2018RS3065) and the Hunan Provincial Natural Science Foundation of China under (grant number: 2019JJ40049).

Conflict of interest

None declared.

References

- Carrington JC, Ambros V. Role of microRNAs in plant and animal development. *Science* 2003;**301**:336–8.
- Iorio MV, Ferracin M, Liu CG, et al. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* 2005;**65**:7065–70.
- Ebert MS, Sharp PA. Roles for microRNAs in conferring robustness to biological processes. *Cell* 2012;**149**:515–24.
- Le TD, Liu L, Zhang J, et al. From miRNA regulation to miRNA-TF co-regulation: computational approaches and challenges. *Brief Bioinform* 2015;**16**:475–96.
- Yi H-C, You Z-H, Huang D-S, et al. A deep learning framework for robust and accurate prediction of ncRNA-protein interactions using evolutionary information. *Mol Ther Nucleic Acids* 2018;**11**:337–44.
- Chen X, Xie D, Zhao Q, et al. MicroRNAs and complex diseases: from experimental results to computational models. *Brief Bioinform* 2019;**20**:515–39.
- Chen X, Wang C-C, Yin J, et al. Novel human miRNA-disease association inference based on random forest. *Mol Ther Nucleic Acids* 2018;**13**:568–79.
- Meng X, Lu P, Fan Q. miR-367 promotes proliferation and invasion of hepatocellular carcinoma cells by negatively regulating PTEN. *Biochem Biophys Res Commun* 2016;**470**:187–91.
- Zheng R, Li M, Chen X, et al. BiXGBoost: a scalable, flexible boosting-based method for reconstructing gene regulatory networks. *Bioinformatics* 2019;**35**:1893–900.
- Li M, Gao H, Wang J, et al. Control principles for complex biological networks. *Brief Bioinform* 2019;**20**:2253–66.
- Xiao Q, Luo J, Liang C, et al. CeModule: an integrative framework for discovering regulatory patterns from genomic data in cancer. *BMC Bioinform* 2019;**20**:67.
- Xiao Q, Luo J, Liang C, et al. Identifying lncRNA and mRNA co-expression modules from matched expression data in ovarian cancer. *IEEE/ACM Trans Comput Biol Bioinform* 2018. doi: [10.1109/TCBB.2018.2864129](https://doi.org/10.1109/TCBB.2018.2864129).
- Xiao Q, Luo J, Dai J. Computational prediction of human disease-associated circRNAs based on manifold regularization learning framework. *IEEE J Biomed Health Inform* 2019;**23**:2661–9.
- Chen X, Zhou Z, Zhao Y. ELLPMDA: ensemble learning and link prediction for miRNA-disease association prediction. *RNA Biol* 2018;**15**:807–18.
- Deng S-P, Zhu L, Huang D-S. Predicting hub genes associated with cervical cancer through gene co-expression networks. *IEEE/ACM Trans Comput Biol Bioinform* 2016;**13**:27–35.
- Zhang Q, Zhu L, Bao W, et al. Weakly-supervised convolutional neural network architecture for predicting protein-DNA binding. *IEEE/ACM Trans Comput Biol Bioinform* 2018. doi: [10.1109/TCBB.2018.2864203](https://doi.org/10.1109/TCBB.2018.2864203).
- Zhang Q, Zhu L, Huang D-S. High-order convolutional neural network architecture for predicting DNA-protein binding sites. *IEEE/ACM Trans Comput Biol Bioinform* 2019;**16**:1184–92.
- Wang CC, Chen X, Yin J, et al. An integrated framework for the identification of potential miRNA-disease association based on novel negative samples extraction strategy. *RNA Biol* 2019;**16**:257–69.
- Zeng XX, Zhang X, Zou Q. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Brief Bioinform* 2016;**17**:193–203.
- Luo J, Xiao Q. A novel approach for predicting microRNA-disease associations by unbalanced bi-random walk on heterogeneous network. *J Biomed Inform* 2017;**66**:194–203.
- Chen X, Yin J, Qu J, et al. MDHGI: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction. *Plos Comput Biol* 2018;**14**:e1006418.
- Yu SP, Liang C, Xiao Q, et al. GLNMDA: a novel method for miRNA-disease association prediction based on global linear neighborhoods. *RNA Biol* 2018;**15**:1215–27.
- Chen X, Xie D, Wang L, et al. BNPMDA: bipartite network projection for MiRNA-disease association prediction. *Bioinformatics* 2018;**34**:3178–86.
- You ZH, Huang ZA, Zhu Z, et al. PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction. *Plos Comput Biol* 2017;**13**:e1005455.
- Chen X, Cheng JY, Yin J. Predicting microRNA-disease associations using bipartite local models and hubness-aware regression. *RNA Biol* 2018;**15**:1192–205.
- Ding P, Luo J, Liang C, et al. Human disease MiRNA inference by combining target information based on heterogeneous manifolds. *J Biomed Inform* 2018;**80**:26–36.
- Xiao Q, Luo JW, Liang C, et al. A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics* 2018;**34**:239–48.
- Luo JW, Xiao Q, Liang C, et al. Predicting MicroRNA-disease associations using Kronecker regularized least squares based on heterogeneous omics data. *IEEE Access* 2017;**5**:2503–13.
- Xiao Q, Dai JH, Luo JW, et al. Multi-view manifold regularized learning-based method for prioritizing candidate disease miRNAs. *Knowl-Based Syst* 2019;**175**:118–29.
- Chen X, Yan GY. Semi-supervised learning for potential human microRNA-disease associations inference. *Sci Rep* 2014;**4**:5501.
- Zhao Y, Chen X, Yin J. Adaptive boosting-based computational model for predicting potential miRNA-disease associations. *Bioinformatics* 2019;**35**:4730–8.
- Chen X, Zhu CC, Yin J. Ensemble of decision tree reveals potential miRNA-disease associations. *Plos Comput Biol* 2019;**15**:e1007209.
- Chen X, Wu QF, Yan GY. RKNMMDA: ranking-based KNN for MiRNA-disease association prediction. *RNA Biol* 2017;**14**:952–62.
- Li G, Luo J, Xiao Q, et al. Predicting microRNA-disease associations using label propagation based on linear neighborhood similarity. *J Biomed Inform* 2018;**82**:169–77.
- Chen X, Huang L. LRSSLMDA: Laplacian regularized sparse subspace learning for MiRNA-disease association prediction. *Plos Computat Biol* 2017;**13**:e1005912.
- Liang C, Yu S, Luo J. Adaptive multi-view multi-label learning for identifying disease-associated candidate miRNAs. *Plos Comput Biol* 2019;**15**:e1006931.
- Peng J, Hui W, Li Q, et al. A learning-based framework for miRNA-disease association identification using neural networks. *Bioinformatics* 2019;**35**:4364–71.

38. Zeng X, Wang W, Deng G, et al. Prediction of potential disease-associated microRNAs by using neural networks. *Mol Ther Nucleic Acids* 2019;**16**:566–75.
39. Pan X, Shen H-B. Inferring disease-associated microRNAs using semi-supervised multi-label graph convolutional networks. *iScience* 2019;**20**:265–77.
40. Li GH, Luo JW, Xiao Q, et al. Predicting microRNA-disease associations using network topological similarity based on DeepWalk. *IEEE Access* 2017;**5**:24032–9.
41. Chen X, Sun L-G, Zhao Y. NCMCMDA: miRNA-disease association prediction through neighborhood constraint matrix completion. *Brief Bioinform* 2020. <https://doi.org/10.1093/bib/bbz159>.
42. Wang L, You ZH, Chen X, et al. LMTRDA: using logistic model tree to predict miRNA-disease associations by fusing multi-source information of sequences and similarities. *PLoS Comput Biol* 2019;**15**:e1006865.
43. Chen X, Huang L, Xie D, et al. EGBMMDA: extreme gradient boosting machine for miRNA-disease association prediction. *Cell Death Dis* 2018;**9**:3.
44. Qu J, Chen X, Yin J, et al. Prediction of potential miRNA-disease associations using matrix decomposition and label propagation. *Knowl-Based Syst* 2019;**186**:104963.
45. Chen X, Yan CC, Zhang X, et al. RBMMMDA: predicting multiple types of disease-microRNA associations. *Sci Rep* 2015;**5**:13877.
46. Shi C, Ruan Q, An G, et al. Semi-supervised sparse feature selection based on multi-view Laplacian regularization. *Image Vision Comput* 2015;**41**:1–10.
47. Qiu X, Chen Z, Zhao L, et al. Unsupervised multi-view non-negative for law data feature learning with dual graph-regularization in smart internet of things. *Future Gener Comp Syst* 2019;**100**:523–30.
48. Li Y, Qiu CX, Tu J, et al. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res* 2014;**42**:D1070–4.
49. Lipscomb CE. Medical subject headings (MeSH). *Bull Med Libr Assoc* 2000;**88**:265–6.
50. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014;**42**:D68–73.
51. Yuan L, Guo L-H, Yuan C-A, et al. Integration of multi-omics data for gene regulatory network inference and application to breast cancer. *IEEE/ACM Trans Comput Biol Bioinform* 2018. doi: [10.1109/TCBB.2018.2866836](https://doi.org/10.1109/TCBB.2018.2866836).
52. Peng C, Zheng Y, Huang D-S. Capsule network based modeling of multi-omics data for discovery of breast cancer-related genes. *IEEE/ACM Trans Comput Biol Bioinform* 2019. doi: [10.1109/TCBB.2019.2909905](https://doi.org/10.1109/TCBB.2019.2909905).
53. Yu SP, Liang C, Xiao Q, et al. MCLPMDA: a novel method for miRNA-disease association prediction based on matrix completion and label propagation. *J Cell Mol Med* 2019;**23**:1427–38.
54. Zou Q, Li J, Song L, et al. Similarity computation strategies in the microRNA-disease network: a survey. *Brief Funct Genomics* 2016;**15**:55–64.
55. Bao W, Jiang Z, Huang D-S. Novel human microbe-disease association prediction using network consistency projection. *BMC Bioinform* 2017;**18**:543–3.
56. Hsu SD, Tseng YT, Shrestha S, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res* 2014;**42**:D78–85.
57. Yu GC, Li F, Qin YD, et al. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* 2010;**26**:976–8.
58. Lu C, Yang M, Luo F, et al. Prediction of lncRNA-disease associations based on inductive matrix completion. *Bioinformatics* 2018;**34**:3357–64.
59. Hao J, Yu W, Li Y. A rank revealing randomized singular value decomposition (R3SVD) algorithm for low-rank matrix approximations. *arXiv* 2016; arXiv:1605:08134. preprint.
60. Xu W, Zhu L, Huang D-S. DCDE: an efficient deep convolutional divergence encoding method for human promoter recognition. *IEEE Trans Nanobiosci* 2019;**18**:136–45.
61. Chen X, Wang L, Qu J, et al. Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* 2018;**34**:4256–65.
62. Chen X, Niu YW, Wang GH, et al. MKRMDA: multiple kernel learning-based Kronecker regularized least squares for miRNA-disease association prediction. *J Transl Med* 2017;**15**:251.
63. Xie BY, Ding Q, Han HJ, et al. miRCancer: a microRNA-cancer association database constructed by text mining on literature. *Bioinformatics* 2013;**29**:638–44.
64. Yang Z, Wu L, Wang A, et al. dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res* 2017;**45**:D812–8.
65. Jiang Q, Wang Y, Hao Y, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res* 2009;**37**:D98–104.
66. Zheng QL, Sheng Q, Jiang CY, et al. MicroRNA-452 promotes tumorigenesis in hepatocellular carcinoma by targeting cyclin-dependent kinase inhibitor 1B. *Mol Cell Biochem* 2014;**389**:187–95.
67. Lu GJ, Dong YQ, Zhang QM, et al. miRNA-221 promotes proliferation, migration and invasion by targeting TIMP2 in renal cell carcinoma. *Int J Clin Exp Pathol* 2015;**8**:5224–9.
68. Jin L, Zhang Z, Li Y, et al. miR-125b is associated with renal cell carcinoma cell migration, invasion and apoptosis. *Oncol Lett* 2017;**13**:4512–20.