



MSCHLMDA: Multi-Similarity Based Combinative Hypergraph Learning for Predicting MiRNA-Disease Association

Qingwen Wu¹, Yutian Wang¹, Zhen Gao¹, Jiancheng Ni^{1*} and Chunhou Zheng^{1,2*}

¹ School of Software, Qufu Normal University, Qufu, China, ² School of Computer Science and Technology, Anhui University, Hefei, China

OPEN ACCESS

Edited by:

Wen Zhang,
Huazhong Agricultural
University, China

Reviewed by:

Zhu-Hong You,
Xinjiang Technical Institute of Physics
and Chemistry (CAS), China
ShuLin Wang,
Hunan University, China
Tiantian He,
Nanyang Technological
University, Singapore

*Correspondence:

Jiancheng Ni
nijch@163.com
Chunhou Zheng
zhengch99@126.com

Specialty section:

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

Received: 22 November 2019

Accepted: 23 March 2020

Published: 15 April 2020

Citation:

Wu Q, Wang Y, Gao Z, Ni J and
Zheng C (2020) MSCHLMDA:
Multi-Similarity Based Combinative
Hypergraph Learning for Predicting
MiRNA-Disease Association.
Front. Genet. 11:354.
doi: 10.3389/fgene.2020.00354

Accumulating biological and clinical evidence has confirmed the important associations between microRNAs (miRNAs) and a variety of human diseases. Predicting disease-related miRNAs is beneficial for understanding the molecular mechanisms of pathological conditions at the miRNA level, and facilitating the finding of new biomarkers for prevention, diagnosis and treatment of complex human diseases. However, the challenge for researchers is to establish methods that can effectively combine different datasets and make reliable predictions. In this work, we propose the method of Multi-Similarity based Combinative Hypergraph Learning for Predicting MiRNA-disease Association (MSCHLMDA). To establish this method, complex features were extracted by two measures for each miRNA-disease pair. Then, K-nearest neighbor (KNN) and K-means algorithm were used to construct two different hypergraphs. Finally, results from combinative hypergraph learning were used for predicting miRNA-disease association. In order to evaluate the prediction performance of our method, leave-one-out cross validation and 5-fold cross validation was implemented, showing that our method had significantly improved prediction performance compared to previously used methods. Moreover, three case studies on different human complex diseases were performed, which further demonstrated the predictive performance of MSCHLMDA. It is anticipated that MSCHLMDA would become an excellent complement to the biomedical research field in the future.

Keywords: microRNA, disease, miRNA-disease association, K-nearest neighbor, K-means, combinative hypergraph learning

INTRODUCTION

MicroRNAs(miRNAs) are a class of small endogenous non-coding RNAs that mainly regulate gene expression at the post-transcriptional level, whose length is equivalent to 20–25 nucleotides (Bartel, 2009; Ribeiro et al., 2014). The first miRNA was discovered in the early 1990's. However, miRNAs were not recognized as a distinct class of biological regulators until the early 2000's. Recently, accumulating studies have indicated that more than one-third of genes are regulated by miRNAs (Taguchi, 2012), and that miRNAs participate in various biological processes, such as cell proliferation, tissue development, apoptosis, differentiation and signal transduction (Mattick and Makunin, 2006; Esteller, 2011; Mattick and Rinn, 2015). The deregulation of miRNAs appears to be

associated with various diseases, ranging from common diseases to cancers (Sayed and Abdellatif, 2011; Farazi et al., 2013). For example, based on deep sequencing information and cluster analysis, several miRNAs, including miR-7, miR-95, miR-124, miR-128, and miR-132 were found to be significantly down-regulated in glioblastoma (Skalsky and Cullen, 2011). In addition, Dkk-3 and SMAD4 were identified as potential target genes of miR-183, and the expression of miR-183, miR-146a, and miR-767-5P were significantly higher in prostate cancer tissues (Ueno et al., 2013).

Therefore, predicting potential miRNA-disease associations could not only improve our knowledge of the underlying disease mechanisms at the miRNA level, but also facilitate the finding of novel disease biomarkers for early detection and drug discovery in the contexts of disease prevention, diagnosis, treatment and prognosis. However, compared with the rapidly increasing number of newly discovered miRNAs, only a few miRNA-disease associations have been confirmed. Experimental confirmation of the new disease-related miRNAs is extremely expensive and time-consuming, whose failure rate is also high. Currently, a great quantity of biological data about miRNAs has been generated, and more and more studies have focused on the computational algorithms which can select the most promising miRNAs for further analysis. By decreasing the number of experiments, more effective experimental procedures could be conducted to uncover potential disease-related miRNAs on a large scale.

Mainstream computational methods are roughly grouped into two categories. The first category is based on network analysis (Chen et al., 2012, 2016; Zeng et al., 2016, 2018; Li et al., 2017; Liu et al., 2017; Xiao et al., 2017; Zhong et al., 2017). Jiang et al. (2018) designed the significance SIG of disease pairs or miRNA pairs and then developed a novel miRNA-disease association prediction (ICFMDA) method, which was used to improve the collaborative filtering approach. The collaborative filtering algorithm was further improved by incorporating similarity matrices to enable the prediction of a new miRNA and a particular disease without known associations. Chen et al. (2018) proposed a Two-tier Random Walk method in which they designed a Laplacian score of graphs for the prediction of disease-related miRNAs (GSTRW). This method can predict the correlation of all diseases with miRNAs simultaneously without negative samples. By performing a depth-first search algorithm on the heterogeneous network to infer disease-related miRNAs, You et al. (2017) presented a model called PBMDA, which could be employed in new diseases or miRNAs, greatly improving practicability and reliability. Chen et al. (2018c) designed a Network Distance Analysis method for miRNA-disease Association prediction (NDAMDA), which used the direct network distance and average network distances between two miRNAs or diseases. However, this model might cause a bias toward miRNAs with more known related diseases and might not be applicable to the diseases where associated miRNAs tend to be randomly distributed in the network. Zhao Q. et al. (2018) developed a miRNA-disease association prediction method based on the Spy and super clustering strategy (SSCMDA). They used a Spy strategy to recognize trustworthy negative samples from the uncertain miRNA-disease pairs which could improve prediction accuracy.

However, this method used the Regularized Least Square as the baseline classifier and it was difficult to attain the optimal combining parameters to merge all the developed strategies. Zhao H. C. et al. (2018) proposed a method to predict miRNA-disease associations based on a distance correlation set (DCSMDA). The high point of this approach lay in the construction of a miRNA-lncRNA-disease network that could be applied to predict potential lncRNA-disease associations. Nevertheless, this approach cannot be applied to unknown diseases or miRNAs that are not present in the miRNA-disease or lncRNA-miRNA databases. Later, Zhao et al. (2019) developed a method based on a shortest path algorithm for discovering potential miRNA-disease associations. This method improved the sparseness of known associations and did not require negative samples to predict potential miRNA-disease association simultaneously.

Methods that belong to the second category are adopted machine learning algorithms used to predict miRNA-disease associations (Jiang et al., 2010; Xu et al., 2011; Chen et al., 2015). Chen and Yan (2014) designed a semi-supervised method called RLSMDA. This method could identify disease-related miRNAs without known miRNAs. However, the parameter optimization for RLSMDA was challenging. Chen et al. (2018a) proposed a new machine learning method for miRNA-disease association prediction. They used a stacked auto-encoder to extract deep features and a greedy unsupervised algorithm for a pre-training model. At last, the support vector machine (SVM) was utilized to uncover potential associations. However, the optimization of complex parameters was complicated in this model. Furthermore, Chen et al. (2018b) designed a prediction method named EGBMMDA, which adopted an Extreme Gradient Boosting Machine to predict potential associations. This approach was the first decision tree learning-based method and one of the very few models that achieved a global LOOCV AUC >0.9 at that time. Recently, Xuan et al. (2019) developed a dual convolutional neural network-based method for predicting potential disease-miRNA association (CNNMDA), which was a computational model based on deep learning and used the original and global representation of an miRNA-disease pair to predict disease-related miRNAs. However, this method has many parameters and involves a large number of calculations.

Although the methods mentioned above have made great contributions to the discovery of miRNA-disease associations, there are still some limitations in many aspects. In addition, the limited number of known miRNA-disease associations results in a sparse matrix. Thus, in order to improve the accuracy of the prediction model, we propose a novel prediction method based on a hypergraph and refer to it as MSCHLMDA. The edge of a hypergraph can own more than two vertices, endowing hypergraphs with high flexibility for depicting high-order relationships. Benefitted by this desirable property, hypergraph models have been successfully applied to dozens of computer vision as well as machine learning and pattern recognition areas. The performance of hypergraph learning highly depends on the generated hypergraph structure. A good hypergraph structure can represent the data correlation better. In this study, for all the miRNA-disease pairs, two different measures (graph theoretical and statistical) were utilized to formulate the potential

informative features, and a combinative hypergraph learning model was designed to predict their unknown associations. Experiments with cross validations and case studies fully demonstrated that the performance of our method in predicting the potential disease-miRNA associations has a significant advantage compared to previous methods.

MATERIALS AND METHODS

Method Overview

Our model mainly consists of two steps: (I) data collection and preprocessing, (II) association prediction. First, the feature vector X of all miRNA-disease pairs was constructed; Second, the combinative hypergraph model was designed to learn projection matrices, which were used to map the unknown miRNA-disease pair features to the association scores matrix S .

Data Collection

The raw data used by our method were three matrices: miRNA-disease association matrix A , miRNA similarity matrix SM and disease similarity matrix SD . Matrix A was obtained from the HMDDv2.0 (Li et al., 2014), which contains 5,430 known associations between 495 (nm) miRNAs and 383 (nd) diseases. Concretely, if miRNA $m(i)$ is verified to be associated with disease $d(j)$, the value of $A(m[i], d[j])$ is equal to 1, and 0 otherwise. Our goal is to predict the link between miRNAs and diseases in matrix A . SM was directly downloaded from <http://www.cuilab.cn/files/images/cuilab/misim.zip>. It included similarity scores for all 495 miRNAs, for which the scores were calculated according to the Wang et al. (2010) method. The larger the $SM(m[i], m[j])$ is, the closer their associations will be.

SD contains the similarity scores of different diseases. Based on the disease classification system in the Mesh database, we can use a directed acyclic graph (DAG) to describe the similarity between different diseases. There were two methods to calculate the contribution values of disease $d(t)$ to the semantic value of disease $d(i)$ as follows:

$$D1_{d(i)}(d(t)) = -\log\left(\frac{\text{the number of DAGs including } d(t)}{\text{the number of diseases}}\right) \quad (1)$$

$$DV1(d(i)) = \sum_{d(t) \in T(d(i))} D1_{d(i)}(d(t)) \quad (2)$$

and

$$D2_{d(i)}(t) = \begin{cases} 1 & \text{if } d(t) = d(i) \\ \Delta * D2_{d(i)}(d(t')) & \text{if } d(t) \neq d(i) \end{cases} \quad (3)$$

$$DV2(d(i)) = \sum_{d(t) \in T(d(i))} D2_{d(i)}(d(t)) \quad (4)$$

where Δ represents the semantic contribution factor. It will reduce the contribution of disease $d(t)$ if $d(t)$ is different from $d(i)$.

The disease similarity score was calculated based on the measurement of common subgraphs between disease DAGs. So,

the similarity between disease $d(i)$ and $d(j)$ could be defined as below:

$$SD1(d(i), d(j)) = \frac{\sum_{t \in T(d(i)) \cap T(d(j))} (D1_{d(i)}(t) + D1_{d(j)}(t))}{DV1(d(i)) + DV1(d(j))} \quad (5)$$

and

$$SD2(d(i), d(j)) = \frac{\sum_{t \in T(d(i)) \cap T(d(j))} (D2_{d(i)}(t) + D2_{d(j)}(t))}{DV2(d(i)) + DV2(d(j))} \quad (6)$$

Therefore, by integrated $SD1$ and $SD2$, we could reconstruct a new similarity matrix $SD = \frac{SD1 + SD2}{2}$.

Data Preprocessing

Generally, the similarity of miRNAs, as well as the similarity of diseases, is used to predict the association between miRNAs and diseases directly. However, some unknown interactions might affect the prediction results. To address this limitation, the WKNNP preprocessing method (Xiao et al., 2017) was used to estimate previously unknown but possible interactions between miRNAs and diseases through their known neighbors in the matrix A . If the value of $A(i, j)$ is 0, the role of WKNNP is to update it to a value in the range of 0 to 1. Then the complete matrix A is used to generate Gaussian interaction profile kernel (Gipk) similarity (Laarhoven et al., 2011). For miRNAs, a vector $KS(m[i])$, i.e., the i -th row of matrix A , was utilized as the interaction profiles of miRNA $m(i)$ for denoting the association between $m(i)$ itself and each disease. Thus, the Gipk similarity $GIM(m[i], m[j])$ of miRNA $m(i)$ and miRNA $m(j)$ was defined as:

$$GIM(m(i), m(j)) = \exp(-\gamma_m \|KS(m(i)) - KS(m(j))\|^2) \quad (7)$$

Where $\|\cdot\|^2$ represented l_2 norm, γ_m was a parameter used to control the kernel bandwidth, which was set as

$$\gamma_m = \frac{1}{\frac{1}{nm} \sum_{i=1}^{nm} \|KS(m(i))\|^2} \quad (8)$$

By integrating SM and GIM , a more comprehensive miRNA multi-similarity matrix MMS could be obtained as

$$MMS(m(i), m(j)) = \begin{cases} GIM(m(i), m(j)) & \text{if } SM(m(i), m(j)) = 0 \\ \frac{SM(m(i), m(j)) + GIM(m(i), m(j))}{2} & \text{otherwise} \end{cases} \quad (9)$$

Similarly, we also calculated the Gipk similarity GID for diseases by the follow formulas

$$GID(d(i), d(j)) = \exp(-\gamma_d \|KS(d(i)) - KS(d(j))\|^2) \quad (10)$$

$$\gamma_d = \frac{1}{\frac{1}{nd} \sum_{i=1}^{nd} \|KS(d(i))\|^2} \quad (11)$$

where $KS(d(i))$ and $KS(d(j))$ denoted the i th column and the j th column of A . At last, the disease multi-similarity matrix DMS was obtained by

$$DMS(d(i), d(j))$$

$$= \begin{cases} GID(d(i), d(j)) & \text{if } SD(d(i), d(j)) = 0 \\ \frac{GID(d(i), d(j)) + SD(d(i), d(j))}{2} & \text{otherwise} \end{cases} \quad (12)$$

In the above process, all known miRNA-disease associations in matrix A would be used to calculate the GIPK similarity. Therefore, before the cross validation, the corresponding value of a known miRNA-disease association in matrix A should be set to 0, if it was a test sample.

Feature Construction

Based on the description of the literature (He et al., 2017), there were three types of features to be constructed. Type 1 features summarized A , MMS and DMS from a statistical perspective. For miRNA $m(i)$ /disease $d(j)$, we calculated

- *num. ass*: the number of known association in $A(i,:)/A(:,j)$.
- *me. sim*: for $m(i)$, the mean of $MMS(i,:)$; for $d(j)$, the mean of $DMS(j,:)$.
- *dis. sim*: calculate the distribution of similarity scores for $m(i)/d(j)$. Here, the similarity scores were divided into 5 parts.

Type 2 features described MMS/DMS using graph theories. Graphs for miRNAs and diseases from MMS and DMS were built, respectively. The nodes were representing miRNAs or diseases; if two nodes' similarity scores were greater than the mean value of all entities in MMS/DMS , they would be linked by an edge. For each node, we defined the following features

- *num. nb*: number of neighbors.
- *k. sim*: the similarity values of the k -nearest neighbors of the node (in our study k equal 20).
- *bt, cl*: betweenness, closeness of the node.

Type 3 features focused on matrix A . We defined the following features for each miRNA-disease pair based on statistics and graph theories.

- *m. d. nb*: the number of associations between an miRNA and a disease's neighbors.
- *d. m. nb*: the number of associations between a disease and an miRNA's neighbors.
- *m. d. bt, m. d. cl*: betweenness, closeness of the node.

Feature matrix $X = [x_1, \dots, x_i, \dots, x_n]^T \in \mathbb{R}^{n \times c}$ was generated by selecting both positive samples and negative samples with a ratio of 1:1 and putting them into a feature construction. The known associated miRNA-disease pairs were extracted from the HMDDv2.0 to compose the positive sample set, while the same number of unknown miRNA-disease pairs was randomly selected to constitute a negative sample set. The corresponding labels matrix $Y = [y_1, \dots, y_j, \dots, y_l] \in \mathbb{R}^{n \times l}$, where the j -th category is 1 if x_i belongs to j -th category, and other categories are 0.

Hypergraph Construction

A hypergraph is an extension of graph where an edge (i.e., a hyperedge) can connect more than two vertices and represent the structure of data via measuring the similarity between groups of different points. It has great advantages in complex data modeling. For any application using hypergraph learning approaches, the first step was to construct the corresponding

hypergraph structure. Let $G = (V, E, W)$ denote a hypergraph, which consists of a set of vertices V and a cluster of hyperedge E to which a corresponding weight matrix W is assigned. In this study, the total number of vertices was n , and each vertex represented an miRNA-disease pair in X . We used the K -nearest neighbor (KNN) algorithm and K -means algorithm to generate hyperedges, respectively. For KNN hypergraph G_1 , each time one vertex was selected as a centroid, and one hyperedge was constructed to connect the centroid with its k nearest neighbors in the corresponding feature space. For K -means hypergraph G_2 , we used the K -means algorithm to group all miRNA-disease pairs. If some miRNA-disease pairs are in the same group, they would have been linked by the corresponding edge.

A traditional hypergraph G could be denoted by a $|V| \times |E|$ incidence matrix H

$$h(v, e) = \begin{cases} 1 & \text{if } v \in e \\ 0 & \text{if } v \notin e \end{cases} \quad (13)$$

The degree of a vertex $v \in V$ was obtained by

$$d(v) = \sum_{e \in E} w(e)h(v, e) \quad (14)$$

and the degree of a hypergraph $e \in E$ was obtained by

$$\delta(e) = \sum_{v \in V} h(v, e) \quad (15)$$

D_v denoted the diagonal degree matrix of each vertex, and D_e denoted the diagonal matrix containing the degree of hyperedge.

For the hypergraph G_1 , the weight w_1 of a hyperedge e was estimated by the sum of the distance between two vertexes in the same hyperedge

$$w_1(e) = \sum_{u \in e} dist(v, u) \quad (16)$$

$$dist(v, u) = \exp(-\|x_v - x_u\|^2 / \sigma^2) \quad (17)$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n \|x_i - x_{bar}\|^2}, \quad x_{bar} = \frac{1}{n} \sum_{i=1}^n x_i \quad (18)$$

where v was the centroid of e and u was v 's neighbor.

For the hypergraph G_2 , all the hyperedges were initialized with an equal weight, e.g., $w_2(e) = 1/n_e$, where n_e was the number of hyperedges.

Combinative Hypergraph Learning

There were two hypergraphs in total, denoted by $G_1 = (V_1, E_1, W_1)$ and $G_2 = (V_2, E_2, W_2)$. For each hypergraph, we aimed to learn an individual projection matrix P_i , and the overall combination of all projected matrices could be used to predict the disease-related miRNAs. **Figure 1** illustrated the main framework of our method. We noted that an optimal combination of different hypergraph was also important. Thus, the combination weights $B = [\beta_1, \beta_2]$ were further introduced as another objective

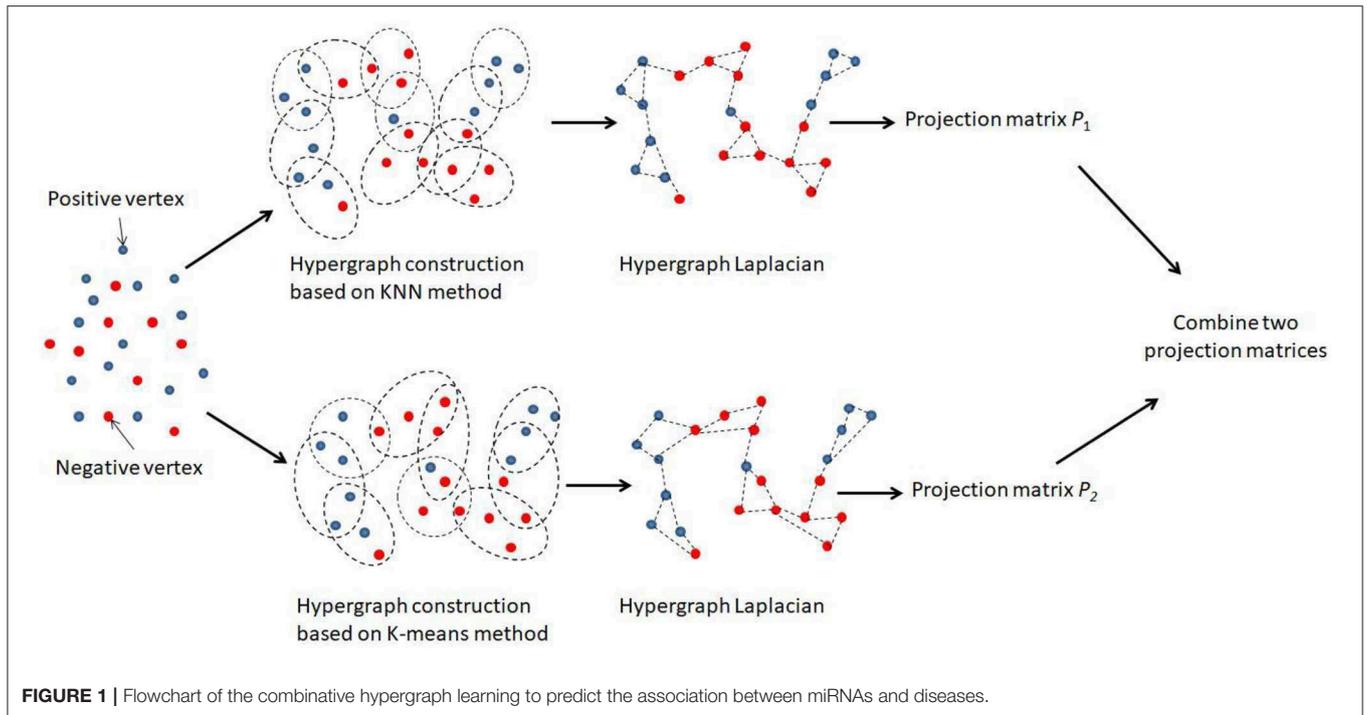


FIGURE 1 | Flowchart of the combinative hypergraph learning to predict the association between miRNAs and diseases.

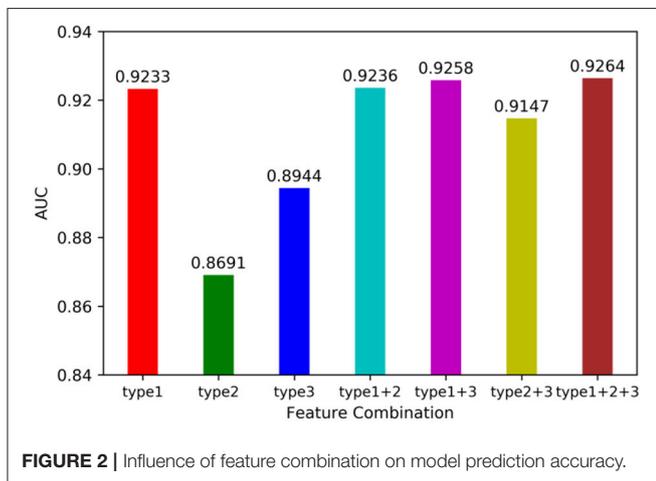


FIGURE 2 | Influence of feature combination on model prediction accuracy.

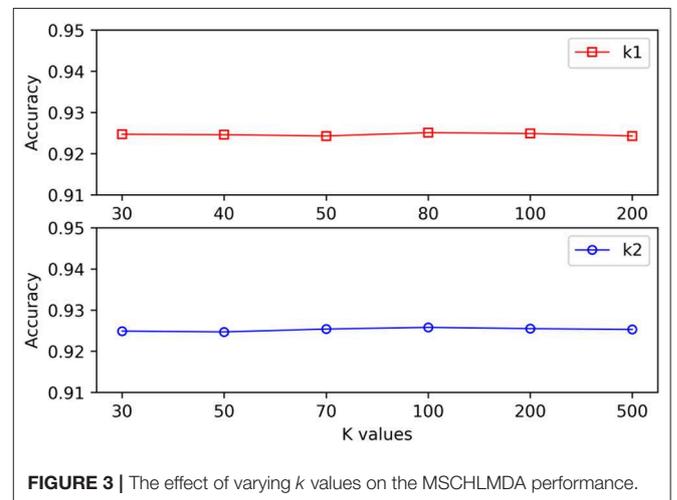


FIGURE 3 | The effect of varying k values on the MSCHLMDA performance.

of the learning task, where β_i was the combination weight for the i -th hypergraph subjecting to $\sum_{i=1}^2 \beta_i = 1$ and $B \geq 0$.

We adopted the objective function proposed in Zhang et al. (2018):

$$\arg \min_{P_i, B \geq 0} \left\{ \sum_{i=1}^2 \beta_i \{ \Omega(P_i) + \lambda R_{emp}(P_i) + \mu \Phi(P_i) \} \right. \quad (19)$$

$$\left. \begin{aligned} & \{ +\eta \Psi(B) \} \\ & \text{s.t. } \sum_{i=1}^2 \beta_i = 1 \end{aligned} \right.$$

Specifically, hypergraph Laplacian regularizer $\Omega(P_i)$ was calculated as

$$\begin{aligned} \Omega(P_i) &= \frac{1}{2} \sum_{k=1}^l \sum_{e \in E_i} \sum_{u, v \in V_i} \frac{W_i(e) H_i(u, e) H_i(v, e)}{\delta_i(e)} \left(\frac{(XP_i)(u, k)}{\sqrt{d_i(u)}} \right. \\ &\quad \left. - \frac{(XP_i)(v, k)}{\sqrt{d_i(v)}} \right)^2 \\ &= \text{tr}(P_i^T X^T \Delta_i X P_i) \end{aligned} \quad (20)$$

where $\Delta_i = I - D_{v_i}^{-1/2} H_i W_i D_{e_i}^{-1} H_i^T D_{v_i}^{-1/2}$ was the hypergraph laplacian matrix, in which I denoted the identity matrix, function $tr(\cdot)$ returned the trace of matrix.

The empirical loss term on P_i was denoted as

$$R_{emp}(P_i) = \|XP_i - Y\|^2 \tag{21}$$

$\Phi(P_i)$ was a l_2 norm regularizer to avoid over-fitting for P_i . $\Phi(P_i)$ was denoted as:

$$\Phi(P_i) = \|P_i\|^2 \tag{22}$$

Here, $\Psi(B)$ was measured as l_2 norm of the hypergraph weights:

$$\Psi(B) = \|B\|^2 \tag{23}$$

The Equation (19) was a multiple variables optimization problem. We noted that it can be split into three independent sub-problems, which were related to each P_i and B , respectively. Therefore, to solve the optimization problem, we first optimized each P_i individually, and then optimized the combination weight B .

Firstly, we optimized each P_i individually. For each hypergraph, the learning task could be rewritten as

$$\arg \min_{P_i} \{ tr(P_i^T X^T \Delta_i X P_i) + \lambda \|XP_i - Y\|^2 + \mu \|P_i\|^2 \} \tag{24}$$

To solve the optimization task in Equation (24), we derived function to P_i . The result could be mathematically denoted as follows

$$P_i = \lambda(X^T \Delta_i X + \lambda X^T X + \mu I)^{-1} X^T Y \tag{25}$$

where I was an identity matrix.

Next, fix each P_i and optimized B . We let $\Theta_i = \Omega(P_i) + \lambda R_{emp}(P_i) + \mu \Phi(P_i)$, and the learning task could be rewritten as

$$\begin{aligned} \arg \min_{B \geq 0} \{ \sum_{i=1}^2 \beta_i \Theta_i + \eta \|B\|^2 \} \\ s.t. \sum_{i=1}^2 \beta_i = 1 \end{aligned} \tag{26}$$

To solve this task, the Lagrange multiplier method was employed and the optimization problem was defined as:

$$\arg \min_{B, \varepsilon} \{ \sum_{i=1}^2 \beta_i \Theta_i + \eta \|B\|^2 \} + \varepsilon (\sum_{i=1}^2 \beta_i - 1) \tag{27}$$

It was derived that

$$\varepsilon = \frac{-\sum_{i=1}^2 \Theta_i - 2\eta}{2}, \beta_i = \frac{1}{2} + \frac{\sum_{i=1}^2 \Theta_i}{4\eta} - \frac{\Theta_i}{2\eta} \tag{28}$$

According to the learned P_i and β_i , the association score of the uncertain miRNA-disease pair x^{un} could be obtained by

$$S(x^{un}) = \sum_{i=1}^2 \beta_i x^{un} \cdot P_i \tag{29}$$

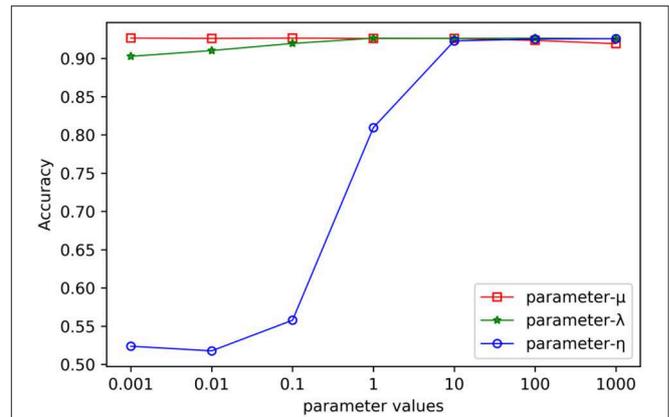


FIGURE 4 | The effect of varying the parameters on the MSCHLMDA performance.

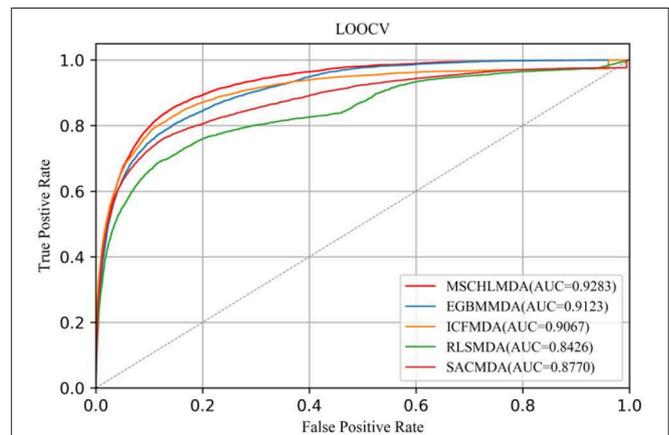
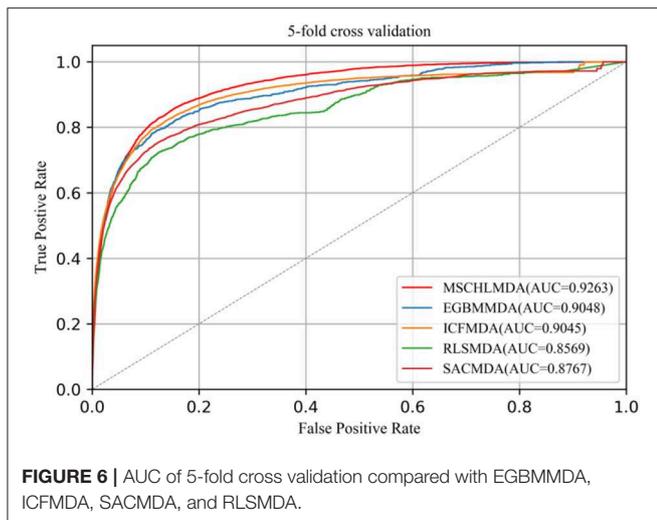


FIGURE 5 | AUC of LOOCV compared with EGBMMDA,ICFMDA, RLSMDA, and SACMDA.

RESULTS

Cross Validation

We utilized leave-one-out cross validation (LOOCV) and 5-fold cross validation (5-CV) to evaluate the performance of MSCHLMDA. A typical machine learning task is to predict the label of a sample by its features. But for a particular learning algorithm, it is unknown which feature is effective. Therefore, it is necessary to select the relevant features that are beneficial to the learning algorithm from all the possible features. In this study, we combined three types of features arbitrarily, forming seven combinations, i.e., type 1; type 2; type 3; type 1, and type 2; type 1 and type 3; type 2 and type 3; type 1, type 2 and type 3. Then we conducted the 5-CV on each combination and calculated the area under curve (AUC) value. The results are shown in Figure 2. Our results indicate that when all three types of features were combined together, the AUC value was the highest. Therefore, for each miRNA-disease pair, we combined type 1, 2, and 3 features into one effective feature vector x , which was used to create the hypergraph to predict miRNA-disease associations.



When different hypergraphs were created, k_1 was adopted to represent the number of neighbors for each vertex, and k_2 was adopted to represent the number of clusters. It is challenging to select the best k value, and thus different k values were used in this study to verify the impact of each value. As shown in **Figure 3**, it is observed that the proposed method could still obtain stable results even when k_1 and k_2 exhibited substantial changes.

In the process of combinative hypergraph learning, the parameters λ , μ and η were the empirical loss, the regularizer on the projection matrices and the regularizer on the hypergraph weights, respectively. They were obtained from the set $\{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$ by cross validating the values of various parameters. We first empirically set them as 10^0 , 10^0 , and 10^3 , respectively. When the influence of one parameters (such as λ) on the prediction performance was being verified, the other two parameters were fixed (such as $\mu = 10^0$, $\eta = 10^3$) while the values of λ were changed from 10^{-3} to 10^3 . **Figure 4** shows the AUC values with varying parameters under cross validation. Our results suggest that the proposed method could achieve relatively stable performance even if λ and μ show in a large range of variability, and that η had a greater impact on the results obtained from this method. It is found that MSCHLMDA achieved the best performance when $\eta = 10^3$. Besides, we ensured more stable results by setting λ to 10^1 and μ to 10^0 .

LOOCV considered each known association as a test sample, while remaining known associations were treated as the training set and all unknown associations were used as candidate samples. When MSCHLMDA completed the forecasting task, the scores of the test sample and candidate samples were compared to iteratively obtain a predicted ranking. The prediction was considered true positive if the rank of the test sample was no lower than the threshold. The prediction was considered false positive if the rank of the candidate sample was no lower than the threshold. The methods of EGBMMDA (Chen et al., 2018b), ICFMDA (Jiang et al., 2018), RLSMDA (Chen and Yan, 2014), and SACMDA (Shao et al., 2018) were implemented on the same dataset, and the parameters were set according to the values

TABLE 1 | The top 50 predicted miRNAs associated with Prostate Neoplasms.

miRNA	Evidence	miRNA	Evidence
hsa-mir-21	miR2Disease;dbDEMC	hsa-mir-223	miR2Disease;dbDEMC
hsa-mir-155	dbDEMC	hsa-mir-133b	dbDEMC
hsa-mir-146a	miR2Disease	hsa-mir-146b	Unconfirmed
hsa-mir-221	miR2Disease;dbDEMC	hsa-mir-181a	miR2Disease;dbDEMC
hsa-mir-122	Unconfirmed	hsa-mir-124	dbDEMC
hsa-mir-16	miR2Disease;dbDEMC	hsa-mir-106b	dbDEMC
hsa-mir-29a	miR2Disease;dbDEMC	hsa-mir-203	Unconfirmed
hsa-mir-15a	miR2Disease;dbDEMC	hsa-let-7a	miR2Disease;dbDEMC
hsa-mir-1	dbDEMC	hsa-mir-196a	dbDEMC
hsa-mir-34a	miR2Disease;dbDEMC	hsa-mir-200b	Unconfirmed
hsa-mir-29b	miR2Disease;dbDEMC	hsa-mir-206	dbDEMC
hsa-mir-133a	dbDEMC	hsa-mir-19b	miR2Disease;dbDEMC
hsa-mir-143	miR2Disease;dbDEMC	hsa-mir-96	miR2Disease;dbDEMC
hsa-mir-126	miR2Disease;dbDEMC	hsa-mir-200c	dbDEMC
hsa-mir-222	miR2Disease;dbDEMC	hsa-mir-181b	miR2Disease;dbDEMC
hsa-mir-31	miR2Disease;dbDEMC	hsa-mir-214	miR2Disease;dbDEMC
hsa-mir-20a	miR2Disease	hsa-mir-34c	dbDEMC
hsa-mir-17	miR2Disease	hsa-mir-195	miR2Disease;dbDEMC
hsa-mir-142	Unconfirmed	hsa-mir-210	miR2Disease
hsa-mir-29c	dbDEMC	hsa-mir-24	miR2Disease;dbDEMC
hsa-mir-92a	Unconfirmed	hsa-mir-18a	Unconfirmed
hsa-mir-199a	miR2Disease;dbDEMC	hsa-let-7b	miR2Disease;dbDEMC
hsa-mir-150	dbDEMC	hsa-mir-148a	miR2Disease
hsa-mir-182	miR2Disease;dbDEMC	hsa-mir-19a	dbDEMC
hsa-mir-15b	dbDEMC	hsa-mir-200a	dbDEMC

given in the original article. Finally, MSCHLMDA obtained the AUC of 0.9283 in LOOCV as shown in **Figure 5**. The AUCs of ICFMDA, EGBMMDA, SACMDA and RLSMDA in LOOCV are 0.9067, 0.9123, 0.8770, and 0.8426, respectively.

In 5-CV, all confirmed associations were randomly divided into five uncrossed subsets with equal sizes. One subset was considered as a test sample and the remaining four subsets as training sets. In this study, we implemented 5-CV 100 times to reduce the bias introduced by random divisions and then calculated the mean and standard deviation of AUCs. The average AUCs of MSCHLMDA, EGBMMDA, ICFMDA, SACMDA, and RLSMDA are 0.9263 (+/-0.0006), 0.9048 (+/-0.0012), 0.9045 (+/-0.0008), 0.8767 (+/-0.0011), and 0.8569 (+/-0.0020), respectively (see **Figure 6**).

Case Studies

To further evaluate the ability of MSCHLMDA to discover potential miRNA-disease associations, case studies of several important human diseases were carried out, such as prostate neoplasms, hepatocellular carcinoma and breast neoplasms. All confirmed associations in the HMDD v2.0 were put into the training set of MSCHLMDA. According to their prediction scores, the top 50 predicted miRNAs were selected, which were associated with the investigated disease. The other databases,

TABLE 2 | The top 50 predicted miRNAs associated with Hepatocellular Carcinoma.

miRNA	Evidence	miRNA	Evidence
hsa-mir-21	HMDD;miR2disease	hsa-mir-15b	HMDD;dbDEMC
hsa-mir-155	HMDD;miR2disease;dbDEMC	hsa-mir-92a	HMDD;miR2disease
hsa-mir-146a	HMDD;miR2disease;dbDEMC	hsa-mir-181a	HMDD;miR2disease;dbDEMC
hsa-mir-125b	HMDD;miR2disease	hsa-mir-182	HMDD;miR2disease
hsa-mir-122	HMDD;miR2disease;dbDEMC	hsa-mir-200b	HMDD;miR2disease
hsa-mir-221	HMDD;miR2disease;dbDEMC	hsa-mir-133b	HMDD
hsa-mir-29a	HMDD;dbDEMC	hsa-let-7a	HMDD;miR2disease;dbDEMC
hsa-mir-34a	HMDD;miR2disease;dbDEMC	hsa-mir-206	Unconfirmed
hsa-mir-16	HMDD;miR2disease;dbDEMC	hsa-mir-196a	HMDD
hsa-mir-1	HMDD;miR2disease	hsa-mir-200a	HMDD;miR2disease;dbDEMC
hsa-mir-15a	HMDD;miR2disease;dbDEMC	hsa-mir-124	HMDD;miR2disease
hsa-mir-133a	miR2disease	hsa-mir-146b	HMDD
hsa-mir-29b	HMDD;dbDEMC	hsa-mir-210	HMDD;dbDEMC
hsa-mir-145	HMDD;miR2disease;dbDEMC	hsa-mir-195	HMDD;miR2disease;dbDEMC
hsa-mir-199a	HMDD;miR2disease;dbDEMC	hsa-mir-214	HMDD;miR2disease;dbDEMC
hsa-mir-126	HMDD;miR2disease;dbDEMC	hsa-mir-34c	HMDD
hsa-mir-29c	HMDD;dbDEMC	hsa-mir-19b	HMDD;miR2disease
hsa-mir-20a	HMDD;miR2disease;dbDEMC	hsa-mir-18a	HMDD;miR2disease;dbDEMC
hsa-mir-150	HMDD;miR2disease;dbDEMC	hsa-mir-9	miR2disease
hsa-mir-17	HMDD;miR2disease	hsa-mir-19a	HMDD;miR2disease;dbDEMC
hsa-mir-31	HMDD;miR2disease	hsa-mir-106b	HMDD;miR2disease;dbDEMC
hsa-mir-222	HMDD;miR2disease;dbDEMC	hsa-mir-181b	HMDD;miR2disease;dbDEMC
hsa-mir-143	miR2disease;dbDEMC	hsa-let-7b	HMDD;miR2disease
hsa-mir-223	HMDD;miR2disease	hsa-mir-148a	HMDD;miR2disease;dbDEMC
hsa-mir-142	HMDD;miR2disease	hsa-mir-24	HMDD;miR2disease

namely dbDEMC (Yang et al., 2010) and miR2Disease (Jiang et al., 2009), were used to validate these findings.

The first experiment was implemented on prostate neoplasms. Prostate neoplasms, also known as the carcinoma of the prostate, are cancers developed from the prostate. The incidence of prostate cancer is 60% higher and the mortality rate is two to three times greater in black vs. white men (Sathegke et al., 2017). Early detection is substantially important for the treatment of prostate tumors. We used MSCHLMDA to predict miRNAs related to prostate neoplasms and considered them as candidate miRNAs. Then, all the candidate miRNAs were ranked in descending order by their predicted scores. Overall, 43 out of the top 50 miRNA predictions were verified by dbDEMC and miR2Disease (See **Table 1**).

In the second experiment using case studies, hepatocellular carcinoma was selected as an example to prove the ability of MSCHLMDA in predicting previously unreported miRNA-disease associations. Hepatocellular carcinoma is a primary liver cancer with a high mortality rate. It is one of the most common malignancies worldwide, especially in Asia, Africa, and southern Europe (Torre et al., 2015). In the first step, all the known hepatocellular carcinoma related miRNAs were removed. Only other disease similarity information and other disease-related miRNAs were used to reveal potentially related miRNAs for hepatocellular carcinoma. When the prediction task was complete, all the miRNAs based on their predicted association

scores were prioritized. Finally, 49 out of the top 50 miRNAs were validated by HMDD v2.0, dbDEMC and miR2Disease (See **Table 2**).

In the final case study, our model was fitted with the miRNA-disease association dataset from HMDD v1.0, which is the old version of HMDD v2.0 and contains less information of miRNA-disease associations. This case study was used to demonstrate MSCHLMDA's robust prediction ability compared to various other datasets. Breast neoplasms were selected as our target disease. Breast neoplasms are the most common malignancies in women, it is also the second leading cause of cancer death among women after lung cancer (Desantis et al., 2016). Here, the whole prediction process was similar to the first experiment of case study. Eventually, 49 out of the top 50 miRNAs in our methods were verified by HMDD v2.0, dbDEMC and miR2Disease (See **Table 3**).

In conclusion, our results show the reliable prediction ability of MSCHLMDA, indicating that MSCHLMDA could be a useful computational mode to investigate a potential disease-related miRNAs association.

DISCUSSION

In recent years, finding novel miRNAs associated with specific diseases has attracted increasing attention in understanding the

TABLE 3 | The top 50 predicted miRNAs associated with Breast Neoplasms.

miRNA	Evidence	miRNA	Evidence
hsa-let-7i	HMDD;miR2Disease;dbDEMC	hsa-mir-32	dbDEMC
hsa-let-7e	HMDD;dbDEMC	hsa-mir-448	dbDEMC
hsa-mir-223	HMDD;dbDEMC	hsa-mir-29c	HMDD;miR2Disease;dbDEMC
hsa-let-7c	HMDD;dbDEMC	hsa-mir-181a	HMDD;miR2Disease;dbDEMC
hsa-mir-126	HMDD;miR2Disease;dbDEMC	hsa-mir-150	dbDEMC
hsa-let-7b	HMDD;dbDEMC	hsa-mir-30e	Unconfirmed
hsa-mir-182	HMDD;miR2Disease;dbDEMC	hsa-mir-30a	HMDD;miR2Disease
hsa-mir-191	HMDD;miR2Disease;dbDEMC	hsa-mir-98	miR2Disease;dbDEMC
hsa-mir-92b	dbDEMC	hsa-mir-203	HMDD;miR2Disease;dbDEMC
hsa-mir-101	HMDD;miR2Disease;dbDEMC	hsa-mir-199b	HMDD;dbDEMC
hsa-mir-130a	dbDEMC	hsa-mir-659	dbDEMC
hsa-mir-532	dbDEMC	hsa-mir-521	dbDEMC
hsa-mir-16	HMDD;dbDEMC	hsa-mir-23b	HMDD;dbDEMC
hsa-let-7g	HMDD;dbDEMC	hsa-mir-130b	dbDEMC
hsa-mir-373	HMDD;miR2Disease;dbDEMC	hsa-mir-196b	dbDEMC
hsa-mir-92a	HMDD	hsa-mir-335	HMDD;miR2Disease;dbDEMC
hsa-mir-24	HMDD;dbDEMC	hsa-mir-26a	HMDD;miR2Disease;dbDEMC
hsa-mir-99b	dbDEMC	hsa-mir-224	HMDD;dbDEMC
hsa-mir-18b	HMDD;dbDEMC	hsa-mir-192	dbDEMC
hsa-mir-15b	dbDEMC	hsa-mir-195	HMDD;miR2Disease;dbDEMC
hsa-mir-99a	dbDEMC	hsa-mir-328	HMDD;miR2Disease;dbDEMC
hsa-mir-372	dbDEMC	hsa-mir-135a	HMDD;dbDEMC
hsa-mir-106a	dbDEMC	hsa-mir-27a	HMDD;miR2Disease;dbDEMC
hsa-mir-520b	HMDD;dbDEMC	hsa-mir-452	HMDD;dbDEMC
hsa-mir-100	HMDD;dbDEMC	hsa-mir-186	dbDEMC

pathophysiology of the diseases and discovery of new drugs to establish effective treatment strategies. In this study, we proposed a combinative hypergraph learning (CHL) method called MSCHLMDA to effectively define miRNA/disease similarity for predicting underlying miRNA-disease associations. CHL captures the similarity between two samples in the same category by KNN hypergraph and K-means hypergraph. MSCHLMDA's performance was verified by cross validation and case studies. These results indicate that MSCHLMDA is able to generate reliable candidate miRNA-disease associations for further validation by biologists.

The improved performance of our model could be mainly attributed to the following two aspects. First, an informative feature vector was created from a statistical analysis and a graph theoretic. The statistical features recorded the sum, the mean, the histogram distributions of the similarity scores, the neighbor count and the neighbor's similarity scores. For miRNAs and diseases, the graph theoretic features contained the betweenness and closeness centrality measures of the network graphs. Second, we used hypergraph learning to design a predictive model. Hypergraph-based models have proven to be beneficial for a variety of classification/clustering tasks, because it can represent the information that three or more vertices have the same semantic attribute, which common graphs are unable to describe. Hypergraphs can model the high-order relationships between their vertices by hyperedges, whose influence can be assessed by

properly estimating their weights. Furthermore, we employed the neighborhood-based formulation and the clustering techniques to generate the hyperedges.

In our previous model of HGMDA (Wu et al., 2019), we also used hypergraph learning, but there are many differences between the implementation process of these two models. First, the hypergraph construction was different. In HGMDA, we only used the K-means algorithm for clustering, which means that known miRNA-disease associations were not utilized to extract the clustering relationship of miRNA-disease pairs. In the current study, KNN and a K-means algorithm was used to seek the relationship between miRNA-disease pairs, which was more comprehensive because KNN was a supervised learning method. Second, the weights of the hyperedges were different. To generate a better hypergraph representation, different hyperedges should have different influences. In HGMDA, all hyperedges had the same weight failing to reflect the importance of different hyperedges. However, in this work, we assigned different weights to each hyperedge based on the distance of each vertex from its neighborhood; this can help to improve the representation ability of the hypergraph structure. Third, the projection matrix was different. In HGMDA, it was required to iterate multiple times to get a stable projection matrix, while in this work we could obtain two projection matrices directly, then combine them into a comprehensive mapping matrix, which was scored higher in efficiency and accuracy.

This method still has some limitations. First, it is required to add negative samples in the training datasets to train the predictive model. Second, due to the computational cost of the hypergraph construction, our method fails to efficiently deal with large-scale samples. Besides, with newly discovered miRNAs, the originally learned projection matrices may be unable to represent the data distribution well. These shortages limit the application range of our model. In future study, we will further investigate the online updates of the learned hypergraph embedding results.

DATA AVAILABILITY STATEMENT

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

REFERENCES

- Bartel, D. P. (2009). MicroRNAs: Target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002
- Chen, M., Liao, B., and Li, Z. J. (2018). 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, X., Gong, Y., Zhang, D. H., You, Z. H., and Li, Z. W. (2018a). DRMDA: deep representations-based miRNA-disease association prediction. *J. Cell. Mol. Med.* 22, 472–485. doi: 10.1111/jcmm.13336
- Chen, X., Huang, L., Xie, D., and Zhao, Q. (2018b). EGBMMDA: Extreme gradient boosting machine for MiRNA-disease association prediction. *Cell Death Dis.* 9, 3–19. doi: 10.1038/s41419-017-0003-x
- 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., Wang, L. Y., and Huang, L. (2018c). NDAMDA: network distance analysis for MiRNA-disease association prediction. *J. Cell Mol. Med.* 22, 2884–2895. doi: 10.1111/jcmm.13583
- Chen, X., Yan, C. C., Zhang, X., You, Z. H., Deng, L. X., Liu, Y., et al. (2016). WBSMDA: within and between score for MiRNA-disease association prediction. *Sci. Rep.* 6:21106. doi: 10.1038/srep21106
- Chen, X., Yan, C. C., Zhang, X. T., Li, Z. H., Deng, L. X., Zhang, Y. D., et al. (2015). RBMMMDA: predicting multiple types of disease-microRNA associations. *Sci. Rep.* 5, 13877–13890. doi: 10.1038/srep13877
- Chen, X., and Yan, G. Y. (2014). Semi-supervised learning for potential human microRNA-disease associations inference. *Sci. Rep.* 4, 5501–5511. doi: 10.1038/srep05501
- Desantis, C. E., Fedewa, S. A., Goding Sauer, A., Kramer, J. L., Smith, R. A., and Jemal, A. (2016). Breast cancer statistics, 2015: convergence of incidence rates between black and white women. *CA Cancer J. Clin.* 66, 31–42. doi: 10.3322/caac.21320
- Esteller, M. (2011). Non-coding RNAs in human disease. *Nat. Rev. Genet.* 12, 861–874. doi: 10.1038/nrg3074
- Farazi, T. A., Hoell, J. I., Morozov, P., and Tuschl, T. (2013). MicroRNAs in human cancer. *Adv. Exp. Med. Biol.* 774, 1–20. doi: 10.1007/978-94-007-5590-1_17
- He, T., Heidemeyer, M., Ban, F. Q., Cherkasov, A., and Ester, M. (2017). SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines. *J. Cheminform.* 9, 24–38. doi: 10.1186/s13321-017-0209-z
- Jiang, Q. H., Wang, G. H., and Wang, Y. D. (2010). An approach for prioritizing disease-related microRNAs based on genomic data integration. *Int. Confer. Biomed. Eng. Inform.* 6, 2270–2274. doi: 10.1109/BMEI.2010.5639313
- Jiang, Q. H., Wang, Y. D., Hao, Y. Y., Juan, L. R., Teng, M. X., Zhang, X. J., 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

AUTHOR CONTRIBUTIONS

JN and CZ conceived and supervised the entire project. QW developed the prediction method. YW and ZG undertook data collection and designed the experiments. QW, YW, and ZG analyzed the result. JN, CZ, and QW wrote the paper. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Nature Science Foundation of China (Nos. 61873001, 61872220, 61672037, 61861146002, and 61732012).

- Jiang, Y. T., Liu, B. T., Yu, L. H., Yan, C. G., and Bian, H. J. (2018). Predict MiRNA-disease association with collaborative filtering. *Neuroinformatics* 16, 363–372. doi: 10.1007/s12021-018-9386-9
- Laarhoven, T. V., Nabuurs, S. B., and Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* 27, 3036–3043. doi: 10.1093/bioinformatics/btr500
- Li, G. H., Luo, J. W., Xiao, Q., Liang, C., Ding, P. J., and Cao, B. W. (2017). Predicting MicroRNA-disease associations using network topological similarity based on deepwalk. *IEEE Access* 5, 24032–24039. doi: 10.1109/ACCESS.2017.2766758
- Li, Y., Qiu, C. X., Tu, J., Geng, B., Yang, J. C., Jiang, T. Z., et al. (2014). HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* 42, D1070–D1074. doi: 10.1093/nar/gkt1023
- Liu, Y. S., Zeng, X. X., He, Z. Y., and Zou, Q. (2017). Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources. *IEEE/ACM Trans. Comput. Boil. Bioinform.* 14, 905–915. doi: 10.1109/TCBB.2016.2550432
- Mattick, J. S., and Makunin, I. V. (2006). Non-coding RNA. *Hum. Mol. Genet.* 15, 17–29. doi: 10.1093/hmg/ddl046
- Mattick, J. S., and Rinn, J. L. (2015). Discovery and annotation of long noncoding RNAs. *Nat. Struct. Mol. Biol.* 22, 5–7. doi: 10.1038/nsmb.2942
- Ribeiro, A. O., Schoof, C. R., Izzotti, A., Pereira, L. V., and Vasques, L. R. (2014). MicroRNAs: modulators of cell identity, and their applications in tissue engineering. *MicroRNA* 3, 45–53. doi: 10.2174/2211536603666140522003539
- Satheke, M., Lengana, T., Maes, A., Vorster, M., Zeevaart, J. R., Lawal, I., et al. (2017). 68ga-psma-11 pet/ct in primary staging of prostate carcinoma: preliminary results on differences between black and white south-africans. *Eur. J. Nuclear Med. Mol. Imag.* 45, 226–234. doi: 10.1007/s00259-017-3852-8
- Sayed, D., and Abdellatif, M. (2011). MicroRNAs in development and disease. *Physiol. Rev.* 91, 827–887. doi: 10.1152/physrev.00006.2010
- Shao, B. Y., Liu, B. T., and Yan, C. G. (2018). SACMDA: MiRNA-disease association prediction with short acyclic connections in heterogeneous graph. *Neuroinformatics* 16, 373–382. doi: 10.1007/s12021-018-9373-1
- Skalsky, R. L., and Cullen, B. R. (2011). Reduced expression of brain-enriched microRNAs in glioblastomas permits targeted regulation of a cell death gene. *PLoS ONE* 6:e24248. doi: 10.1371/journal.pone.0024248
- Taguchi, Y. H. (2012). Inference of target gene regulation via mirnas during cell senescence by using the mirage server. *Aging Dis.* 3, 301–306. doi: 10.1007/978-3-642-31837-5_64
- Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA A Cancer J. Clin.* 65, 87–108. doi: 10.3322/caac.21262
- Ueno, K., Hirata, H., Shahryari, V., Deng, G., Tanaka, Y., Tabatabai, Z. L., et al. (2013). microRNA-183 is an oncogene targeting Dkk-3 and SMAD4 in prostate cancer. *Br. J. Cancer* 108, 1659–1667. doi: 10.1038/bjc.2013.125
- Wang, D., Wang, J., Lu, M., Song, F., and Cui, Q. H. (2010). Inferring the human microRNA functional similarity and functional

- network based on microRNA-associated diseases. *Bioinformatics* 26, 1644–1650. doi: 10.1093/bioinformatics/btq241
- Wu, Q. W., Wang, Y. T., Gao, Z., Zhang, M. W., Ni, J. C., and Zheng, C. H. (2019). “HGMDA: hypergraph for predicting MiRNA-disease association,” in *International Conference on Intelligent Computing* (Cham: Springer), 265–271.
- Xiao, Q., Luo, J. W., Liang, C., Cai, J., and Ding, P. J. (2017). A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics* 34, 239–248. doi: 10.1093/bioinformatics/btx545
- Xu, J., Li, C. X., Lv, J. Y., Li, Y. S., Xiao, Y., Shao, T. T., et al. (2011). Prioritizing candidate disease miRNAs by topological features in the miRNA target-dysregulated network: case study of prostate cancer. *Mol. Cancer Ther.* 10, 1857–1866. doi: 10.1158/1535-7163.MCT-11-0055
- Xuan, P., Sun, H., Wang, X., Zhang, T. G., and Pan, S. X. (2019). Inferring the disease-associated miRNAs based on network representation learning and convolutional neural networks. *Int. J. Mol. Sci.* 20:3648. doi: 10.3390/ijms20153648
- Yang, Z., Ren, F., Liu, C. N., He, S. M., Sun, G., Gao, Q., et al. (2010). dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genom.* 11, S5–S12. doi: 10.1186/1471-2164-11-S4-S5
- You, Z. H., Huang, Z. A., Zhu, Z. X., Yan, G. Y., Li, Z. W., Wen, Z. K., 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. X., Liu, L., Lü, L. Y., 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. X., Zhang, X., and Zou, Q. (2016). 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, Z. Z., Lin, H. J., Zhao, X. B., Ji, R. R., and Gao, Y. (2018). Inductive multi-hypergraph learning and its application on view-based 3D object classification. *IEEE Trans. Image Proces.* 27, 5957–5968. doi: 10.1109/TIP.2018.2862625
- Zhao, H. C., Kuang, L. A., Feng, X., Zou, Q., and Wang, L. (2019). A novel approach based on a weighted interactive network to predict associations of MiRNAs and diseases. *Int. J. Mol. Sci.* 20, 110–130. doi: 10.3390/ijms20010110
- Zhao, H. C., Kuang, L. A., Wang, L., Ping, P. Y., Xuan, Z. W., Pei, T. R., et al. (2018). Prediction of microRNA-disease associations based on distance correlation set. *BMC Bioinform.* 19, 141–156. doi: 10.1186/s12859-018-2146-x
- Zhao, Q., Xie, D., Liu, H. S., Wang, F., Yan, G. Y., and Chen, X. (2018). SSCMDA: spy and super cluster strategy for MiRNA-disease association prediction. *Oncotarget* 9, 1826–1842. doi: 10.18632/oncotarget.22812
- Zhong, Y. L., Xuan, P., Wang, X., Zhang, T. G., Li, J. Z., Liu, Y., et al. (2017). A non-negative matrix factorization based method for predicting disease-associated miRNAs in miRNA-disease bilayer network. *Bioinformatics* 34, 267–277. doi: 10.1093/bioinformatics/btx546

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 Wu, Wang, Gao, Ni and Zheng. 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.