LLCMDA: A Novel Method for Predicting miRNA Gene and Disease Relationship Based on Locality-Constrained Linear Coding

Yu Qu, Huaxiang Zhang*, Chen Lyu and Cheng Liang*

School of Information Science and Engineering, Shandong Normal University, Jinan, China

MiRNAs are small non-coding regulatory RNAs which are associated with multiple diseases. Increasing evidence has shown that miRNAs play important roles in various biological and physiological processes. Therefore, the identification of potential miRNA-disease associations could provide new clues to understanding the mechanism of pathogenesis. Although many traditional methods have been successfully applied to discover part of the associations, they are in general time-consuming and expensive. Consequently, computational-based methods are urgently needed to predict the potential miRNA-disease associations in a more efficient and resources-saving way. In this paper, we propose a novel method to predict miRNA-disease associations based on Locality-constrained Linear Coding (LLC). Specifically, we first reconstruct similarity networks for both miRNAs and diseases using LLC and then apply label propagation on the similarity networks to get relevant scores. To comprehensively verify the performance of the proposed method, we compare our method with several state-of-the-art methods under different evaluation metrics. Moreover, two types of case studies conducted on two common diseases further demonstrate the validity and utility of our method. Extensive experimental results indicate that our method can effectively predict potential associations between miRNAs and diseases.

Keywords: miRNA gene–disease relationship, similarity measure, association prediction, locality-constrained linear coding, label propagation

INTRODUCTION

MiRNAs are small non-coding regulatory RNAs. Since the first miRNA lin-4 (Lee et al., 1993) was found, a plenty of miRNAs have been discovered. Accumulating evidence has shown that miRNAs play a critical role in many biological processes, such as cell proliferation, differentiation, aging, and apoptosis (Ambros, 2004; Xu et al., 2004; Cheng et al., 2005; Miska, 2005; Huang et al., 2016). With the deepening of the research, researchers found that the dysfunctions of miRNAs are closely related to various diseases (Mei et al., 2016; Zou et al., 2016; Liao et al., 2018; Qu et al., 2018b; Tang et al., 2018), which sent an important signal to scientists from all around the world that exploring the associations between miRNAs and diseases is of great significance. Some experimental methods, such as PCR and Microarray (Thomson et al., 2007; Mohammadi-Yeganeh et al., 2013), have been able to successfully identify certain miRNAs related with diseases. However, it is unrealistic to use these traditional experimental methods to predict miRNA-disease associations at a large scale

for their limitations of being time-consuming and expensive. To resolve this situation, multiple computational methods were proposed to efficiently uncover the potential associations between miRNAs and diseases.

Based on the assumption that miRNAs with similar functions are usually related to similar diseases (Zeng et al., 2016; Chen et al., 2017c), Jiang et al. (2010) proposed a network-based method to predict miRNA-disease associations using a hypergeometric distribution scoring system by constructing a miRNA functional similarity network and a human phenome-microRNAome network. Xuan et al. (2013) developed a method named HDMP based on weighted k most similar neighbors. They calculated miRNA functional similarity according to disease terms and disease phenotype similarity. In addition, the miRNAs within same families or clusters were assigned higher weights. Shi et al. (2013) performed random walk to predict miRNA-disease associations on protein–protein interaction (PPI) networks and achieved a satisfactory performance. Mørk et al. (2014) proposed a novel protein-driven method named miRPD to predict potential associations between miRNAs and diseases, where they presented a scoring scheme to efficiently predict and rank miRNA-disease associations. Considering that the global network-based methods could achieve better performance than local network-based methods, Chen et al. (2012) proposed a global similarity measure named RWRMDA. They applied random walk with restart to uncover miRNAs related with diseases on miRNA–miRNA functional similarity network. However, RWRMDA could not predict for diseases without any known related miRNAs. Li et al. (2017) proposed another method named MCMDA. In this method, they applied the matrix completion algorithm to update the known miRNA-disease associations matrix and predict the potential associations. Liu et al. (2017) also applied random walk to predict miRNA-disease associations on a heterogeneous network which was constructed by integrating multiple data sources. Similarly, Luo and Xiao (2017) used an imbalanced bi-random walk to predict miRNA-disease associations on a heterogeneous network consisting of miRNA functional similarity network, disease semantic network and known miRNA-disease association network. Chen et al. (2016a) presented another method WBSMDA to identify the associations between miRNAs and diseases by calculating Gaussian interaction profile kernel similarity for both miRNAs and diseases. Specifically, a within-score and a between-score were calculated and combined to gain a prediction score for each miRNA-disease pair. Using the same data, Chen et al. (2016b) presented HGIMDA which iteratively update an optimization function to uncover potential relations between miRNAs and diseases. Zeng et al. (2018) used structural consistency as an indicator to estimate the link predictability of the bilayer network and further predicted the potential associations between miRNAs and diseases based on Structural Perturbation Method (SPM). According to the lengths of different walks, Zou et al. (2015) introduced a path-based method using KATZ model and obtained reliable results. Similarly, You et al. (2017) proposed another effective path-based method named PBMDA. PBMDA also constructed a heterogeneous network and applied depth-first search algorithm to predict miRNA-disease associations.

Although effective, the length of the paths in the searching process is limited to three. Qu et al. (2018a) presented a novel method SNMDA to identify potential diseases-related miRNAs based on sparse neighborhood and achieved comparable results. In recent years, several models based on machine learning have also been developed to predict the relationships between miRNAs and diseases (Chen et al., 2017b, 2018a,d). Based on semi-supervised learning framework, a model of Regularized Least Squares for MiRNA-Disease Association (RLSMDA) prediction was proposed by Chen and Yan (2014). Xiao et al. (2018) utilized graph-regularized non-negative matrix factorization to effectively predict for diseases without any related miRNAs based on heterogeneous omics data. Chen et al. (Zou et al., 2017) proposed an effective method ELLPMDA based on ensemble learning and link prediction. They integrated the results given by three classical similarity-based algorithms using ensemble learning. Li et al. (2018) presented a Kronecker kernel matrix dimension reduction (KMDR) model to predict miRNA-disease associations which integrates miRNA space and disease space into a larger miRNA-disease associations space. Chen et al. (2017a) proposed another model called MKRMDA that automatically optimizes the combination of multiple kernels. Recently, Chen et al. (2018b) presented EGBMMDA based on the model of extreme gradient boosting machine. Notably, EGBMMDA was the first decision tree learning-based model to uncover disease-related miRNAs and achieved favorable performance.

Although great efforts have been made to reliably predict miRNA-disease associations, there is still room for improvement. In this paper, we propose a novel method called LLCMDA for predicting miRNA-disease associations based on Locality-constrained Linear Coding (LLC). We apply four different cross-validation frameworks to comprehensively evaluate the performance of our method. The comparison results between LLCMDA and five state-of-the-art computational models demonstrate the utility of the proposed method. Besides, case studies on two common neoplasms further prove the effectiveness of our method. In summary, LLCMDA is an effective model for predicting potential miRNA–disease associations.

MATERIALS AND METHODS

Known miRNA-Disease Associations

HMDD (Li et al., 2014) is a database that records known experimentally-verified miRNA-disease associations, which contains 5,430 associations between 383 diseases and 495 miRNAs. For simplicity, an adjacency matrix A of dimension 495 * 383 is defined to describe the known miRNA-disease associations used in this paper. If miRNA $m(i)$ has been confirmed to be related to $d(j)$, $A(i, j) = 1$; otherwise $A(i, j) = 0$.

MiRNA Functional Similarity

Wang et al. (2010b) proposed an informative measure to calculate miRNA functional similarities. Benefitting from previous researches, we downloaded miRNA similarity scores

directly from http://www.cuilab.cn/files/images/cuilab/misim.zip. Similarly, we constructed a miRNA functional similarity matrix *FMS* to represent similarity scores, where *FMS* (i, j) represents the similarity score between miRNA i and miRNA j. A larger value indicates more similar function between two miRNAs.

Disease Semantic Similarity

According to the Mesh descriptor, each disease can be described as a corresponding Directed Acyclic Network (DAG) (Wang et al., 2010a), i.e., DAG$(A) = (A, T(A), E(A))$, where $T(A)$ is the node set including itself as well as its ancestor nodes, and $E(A)$ represents the link set of A. Suppose disease t belongs to $T(A)$, then the contribution of disease t to A can be calculated by:

$$\begin{cases} D_A(t) = 1 & if\ t = A \\ D_A(t) = \max\left\{0.5 * D_A(t')\ \middle|\ t' \in child\ of\ t\right\} & if\ t \neq A \end{cases} \quad (1)$$

Besides, the semantic of A can be calculated by:

$$DV(A) = \sum_{t \in T_{(A)}} D_A(t) \quad (2)$$

For disease A and B, the semantic similarity is calculated through the following formula:

$$S(A, B) = \frac{\sum_{t \in T(A) \cap T(B)} (D_A(t) + D_B(t))}{DV(A) + DV(B)} \quad (3)$$

where t is a common disease both in $T(A)$ and $T(B)$. $D_A(T)$ and $D_B(T)$ represent the contribution of disease t to the disease A and B, respectively. Therefore, for each disease pair, we can calculate their semantic similarity according to Equation (3). For convenience, we use an adjacency matrix *DSS* to denote the obtained semantic similarities for all disease pairs.

Methods

In this paper, we predict potential associations between miRNAs and diseases based on LLC and label propagation. Specifically, the LLC algorithm is first used to reconstruct similarity networks for both miRNAs and diseases and then label propagation is applied on the similarity networks to obtain reliable predicted labels. An overall workflow of LLCMDA is illustrated in **Figure 1**.

Locality-Constrained Liner Coding

Locality-constrained linear coding was first proposed by Wang et al. (2010b) and has been successfully applied to image classification. Compared with sparse representation, LLC is more computationally efficient and can preserve local information during the coding process (Saffari and Ebrahimi-Moghadam, 2015; Zhu et al., 2018). The objective function of LLC algorithm is defined as:

$$\underset{w_i}{\arg\min}\ \|x_i - Dw_i\|_2^2 + \lambda_1 \|P_i \odot w_i\|_2^2\ s.t.\ I^T w_i = 1 \quad (4)$$

Where x_i is the i-th sample, D represents a dictionary matrix and P_i is a local adapter vector representing the distances between

the i-th sample and the other samples. λ_1 is a regularization parameter. The sign of $\odot$ denotes element-wise multiplication. Our goal is to find the optimized reconstructed similarities w_i for each sample x_i. The Lagrangian function of Equation (4) can be obtained as follows:

$$\underset{w_i}{\arg\min}\ \|x_i - Dw_i\|_2^2 + \lambda_1 \|P_i \odot w_i\|_2^2 + \lambda_2\left(I^T w_i - 1\right) \quad (5)$$

Where λ_2 is the Lagrange multiplier. With simple algebra, the above equation can be further transformed into:

$$L(w_i; \eta) = w_i^T C w_i + \lambda_1 w_i^T \{diag(P_i)\}^2 w_i + \lambda_2\left(I^T w_i - 1\right) \quad (6)$$

where $C = \left(x_i I^T - D\right)\left(x_i I^T - D\right)$ and diag (P_i) is a diagonal matrix whose (j,j)-th diagonal elements equals to the j-th element of vector P_i. Specifically, we use the following formula to calculate the local distances between samples for P_i:

$$P_i = \{P_{ij}\}_{j=1,\dots,n} = \left\{\exp\left(\frac{\|x_i - x_j\|_2}{\gamma}\right)\right\}_{j=1,\dots,n} \quad (7)$$

Where γ is a positive parameter controlling the bandwidth.

By taking the derivative of Equation (6) with respect to w_i and setting it to zero, we have:

$$\frac{\partial}{\partial w_i} L(w_i; \eta) = 0 \Rightarrow S w_i + \lambda_2 \mathbf{1} = 0 \quad (8)$$

where $S = 2\left(C + \lambda_1 \{diag(P_i)\}^2\right)$. By multiplying both sides of Equation (8) by $1^T S^{-1}$ and considering the LLC constraint $1^T w_i = 1$, we can derive the optimal solution for w_i as follows:

$$\begin{cases} w_i = \left(C + \{(diag(P_i))\}^2\right) \setminus I \\ w_i = w_i / I^T w_i \end{cases} \quad (9)$$

To obtain feature vectors as the input for LLC algorithm, we applied interaction profile to construct the feature vectors for miRNAs and diseases according to the known miRNA-disease associations (Zang and Zhang, 2012; Zhang et al., 2017). Specifically, the i-th row of adjacency matrix A represents the feature vector of miRNA i and the j-th column represents the feature vector of disease j. As a result, we can obtain two reconstructed similarity networks *RMS* and *RDS* for miRNAs and diseases according to Equation (9), respectively.

Label Propagation

In this section, we adopt label propagation to obtain relevant scores of miRNA-disease pairs. In the process of label propagation, the known miRNA-disease associations are regarded as initial labels and label propagation is used to iteratively update labels (Zhang et al., 2018). Each point receives information not only from its neighbors but also its initial information. Here, we set a parameter α to control the rate. Therefore, the iteration equation on miRNA functional similarity network can be written as follows:

$$F_M(t+1) = \alpha * FMS * F_M(t) + (1 - \alpha) * Y \quad (10)$$
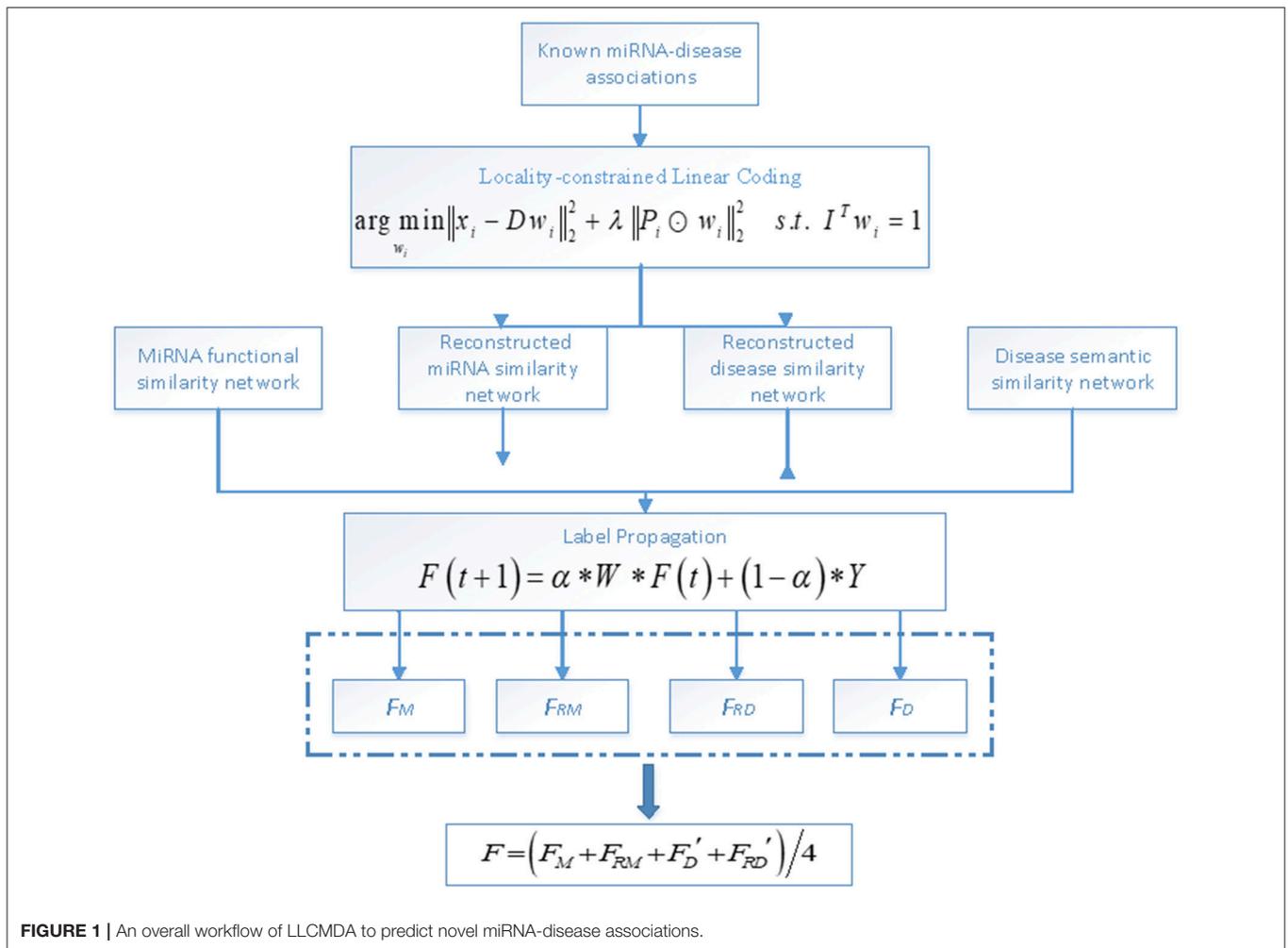
FIGURE 1 | An overall workflow of LLCMDA to predict novel miRNA-disease associations.

Here, *FMS* represents miRNA similarity network while *Y* represents the initial labels and $F_M(0) = Y$. We used Equation (10) to update the label information. When the iteration equation converges, $F_M(t+1)$ is regarded as the relevant score matrix. Therefore, we can sort the miRNAs by relevant scores for each disease. According to previous studies (Zhou et al., 2003), *FMS* is guaranteed to converge if it is properly normalized as follows:

$$FMS = D^{-1/2} * FMS * D^{1/2} \qquad (11)$$

where *D* is a diagonal matrix, the values on the diagonal correspond to the sum of all elements in each row. Similarly, we apply label propagation on the other three similarity networks *RMS*, *DSS*, and *RDS* to obtain three relevant score matrixes F_{RM}, F_D, and F_{RD}. At last, we integrate the four prediction results and take the average as the final output *F*.

$$F = \left(F_M + F_{RM} + F'_D + F'_{RD}\right)/4 \qquad (12)$$

Implementation Details
LLCMDA is implemented in MATLAB under the MATLAB R2016b programming environment. All the experiments are

performed on a desktop with an i7-6700 3.40 GHz CPU and 16G RAM. The source code of LLCMDA is freely available at: https://github.com/misitequ/LLCMDA.

RESULTS

Evaluation
In this section, three cross-validation frameworks are applied to test the performance of our algorithm: global LOOCV, local LOOCV, and five-fold cross-validation. In the framework of global LOOCV, each known miRNA-disease association is left out in turn as a test sample, and the other associations are regarded as training samples. After prediction, each miRNA-disease pair would obtain a score accordingly. If its ranking is higher than a given threshold, the prediction is regarded as a successful prediction. In the framework of local LOOCV, a disease is given in advance and then each miRNA associated with this disease is left out in turn as a test sample while the rest of miRNAs associated with the disease are set as seed samples. The only difference between global LOOCV and local LOOCV is that whether we simultaneously consider the candidates from all diseases (Chen et al., 2018a,c). Five-fold cross validation is

FIGURE 2 | The comparison results between LLCMDA and other four methods (SPM, HGIMDA, EGBMMDA, PBMDA, MKRMDA) in terms of global LOOCV.

FIGURE 4 | The comparison results between LLCMDA and other four methods (SPM, HGIMDA, EGBMMDA, PBMDA, MKRMDA) in terms of five-fold cross-validation.

FIGURE 3 | The comparison results between LLCMDA and other four methods (SPM, HGIMDA, EGBMMDA, PBMDA, MKRMDA) in terms of local LOOCV.

FIGURE 5 | The comparison results between LLCMDA, SPM and HGIMDA in terms of LODOCV.

also implemented to verify the utility of our method. Concretely, the 5,430 known associations are randomly divided into five subsets, each subset is taken as test samples in turn and the others are considered as training samples. To avoid the bias caused by random division of samples, we repeat five-fold cross-validation 20 times and take the average as the final result. Receiver-Operating Characteristics (ROC) curves are plotted by calculating True Positive Rate (TPR) and False Positive Rate (FPR) at varying thresholds. We then calculate the Area Under the ROC Curve (AUC) to quantitatively evaluate the performance of prediction models. AUC = 1 means the model is perfect while AUC = 0.5 denotes a random prediction.

As a result, LLCMDA obtained the AUCs of 0.924, 0.870, and 0.919 in global LOOCV, local LOOCV, and five-fold cross-validation, respectively. To further illustrate the effectiveness

of our algorithm, we compared LLCMDA with five state-of-the-art methods, i.e., SPM, HGIMDA, PBMDA, MKRMDA, EGBMMDA. In the framework of global LOOCV, SPM, HGIMDA, PBMDA, MKRMDA, and EGBMMDA achieved AUCs of 0.942,0.875, 0.922, 0.904, and 0.912 (**Figure 2**). In local LOOCV, the AUCs obtained by SPM, HGIMDA, PBMDA, MKRMDA, and EGBMDA were 0.814, 0.823, 0.853, 0.827, and 0.807 (**Figure 3**). In addition, they obtained AUC-values of 0.865, 0.867, 0.916, 0.884, and 0.904 in five-fold cross-validation (**Figure 4**), respectively. As can be seen from the results, the AUCs of LLCMDA were higher than that of the other methods in all three cross-validation frameworks except the global LOOCV. In conclusion, our method is reliable to predict the potential miRNA-disease associations.

To further test the performance of our method in predicting new associations for diseases without any known related miRNAs, we adopted another evaluation metric called Leave

FIGURE 6 | The parameter effects on the prediction performance in: **(A)** five-fold cross-validation; **(B)** LODOCV.
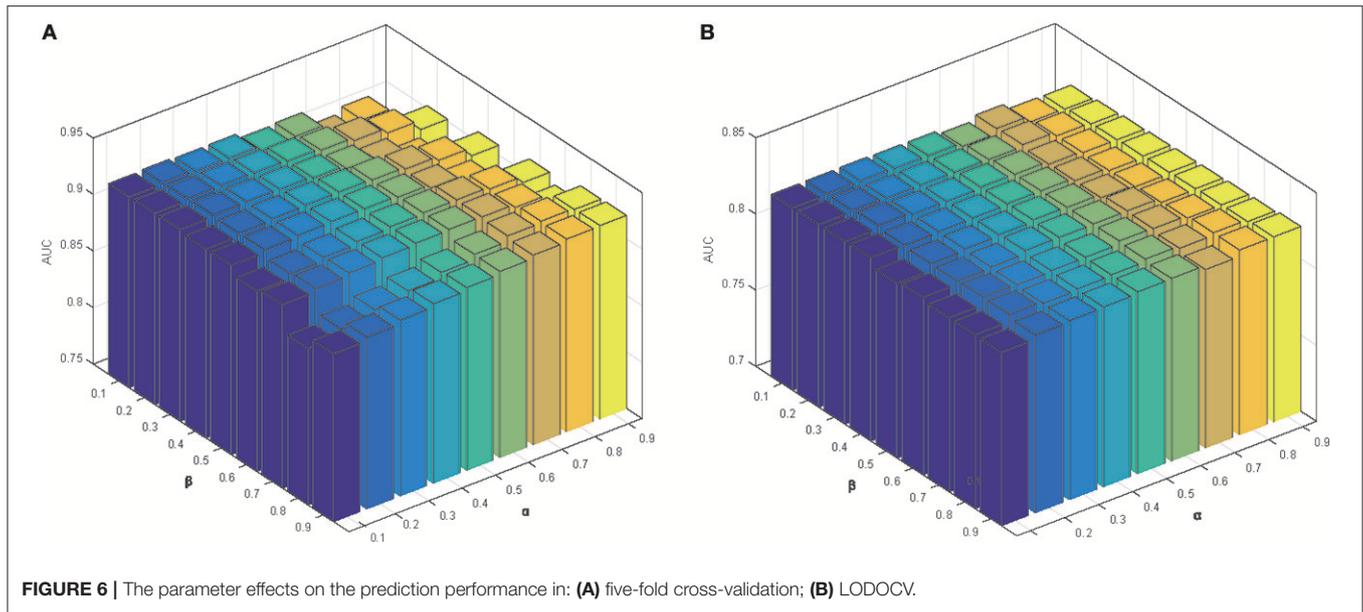
One Disease Out Cross Validation (LODOCV) (Fu and Peng, 2017). In particular, we removed all the associated miRNAs for a given disease and then prioritized all the candidate miRNAs based on the known associations of other diseases. LODOCV is considerably more stringent than the afore mentioned cross-validation frameworks since there is no prior association information available for the given disease. We also compared LLCMDA with the five state-of-the-art methods in terms of the AUC-values. As shown in **Figure 5**, LLCMDA achieved the highest AUC-value of 0.822 in LODOCV framework. Here, we only demonstrate the performances of LLCMDA, SPM, and HGIMDA in the figure as the AUC-values obtained by the other three methods were lower than 0.6. The experimental results indicate that LLCMDA has better generalization ability in predicting new miRNA-disease associations.

Parameter Analysis

Parameter α was used to control the rate of the initial labels on the prediction results for miRNA in Equation (10). Similarly, we used another parameter β to control the effects of initial labels for diseases. To explore the impact of the two parameters, we set different values (0.1–0.9) for both parameters to obtain the prediction results in five-fold cross-validation and LODOCV frameworks (**Figure 6**). It can be seen that parameter α and β only have minor effects on the final prediction accuracies. Similar trends were also observed in global LOOCV and local LOOCV. Consequently, both parameters were set to 0.5.

Case Study

In recent years, substantial evidence suggests that miRNAs are associated with various neoplasms, such as breast neoplasms, lung neoplasms, and etc. Here, we conducted two types of case studies to validate the utility of LLCMDA on two common neoplasms, lung neoplasms and lymphomas. The case studies on other diseases can be found at

TABLE 1 | Top 50 predicted miRNAs associated with Lung Neoplasms based on known associations in HMDD.

miRNA (1–25)	Evidence	miRNA (26–50)	Evidence
hsa-mir-16	I;II;III;	hsa-mir-488	I;
hsa-mir-106b	I;	hsa-mir-376c	I;III;
hsa-mir-429	I;II;	hsa-mir-451	I;II;
hsa-mir-195	I;II;	hsa-mir-302d	I;
hsa-mir-141	I;II;III;	hsa-mir-449a	I;
hsa-mir-130a	I;II;	hsa-mir-520b	I;
hsa-mir-15a	I;II;III;	hsa-mir-139	I;II;
hsa-mir-151	unconfirmed;	hsa-mir-193b	I;
hsa-mir-302a	I;	hsa-mir-383	I;
hsa-mir-373	I;	hsa-mir-194	I;III;
hsa-mir-20b	I;	hsa-mir-149	I;
hsa-mir-296	unconfirmed;	hsa-mir-10a	I;III;
hsa-mir-302c	I;	hsa-mir-452	I;III;
hsa-mir-92b	I;	hsa-mir-491	I;
hsa-mir-339	I;II;	hsa-mir-144	I;III;
hsa-mir-372	I;II;	hsa-mir-520c	unconfirmed;
hsa-mir-28	I;	hsa-mir-449b	I;
hsa-mir-23b	I;	hsa-mir-484	I;
hsa-mir-367	I;	hsa-mir-299	unconfirmed;
hsa-mir-99b	I;	hsa-mir-204	I;II;
hsa-mir-130b	I;	hsa-mir-382	I;
hsa-mir-15b	I;II;	hsa-mir-129	I;
hsa-mir-99a	I;II;III;	hsa-mir-432	I;
hsa-mir-215	I;	hsa-mir-301b	I;
hsa-mir-342	I;	hsa-mir-423	II;

I, II and, III represent dbDEMC, miR2Disease, and miRwayDB, respectively. The first and third columns record the 1–25 and 26–50 related miRNAs, respectively.

https://github.com/misitequ/LLCMDA. We selected the top 50 miRNAs predicted by our model for each disease. The prediction results were then verified by another three databases,

TABLE 2 | Top 50 predicted miRNAs associated with Lymphomas based on known associations in the older version of HMDD.

miRNA (1–25)	Evidence	miRNA (26–50)	Evidence
hsa-mir-21	HMDDv2.0;I;II;III;	hsa-mir-668	HMDD;I;
hsa-mir-155	HMDDv2.0;I;II;III;	hsa-mir-339	I;
hsa-mir-221	HMDDv2.0;I;II;	hsa-mir-143	HMDDv2.0;I;
hsa-mir-146a	HMDDv2.0	hsa-mir-10a	HMDDv2.0
hsa-mir-222	HMDDv2.0;I;	hsa-mir-30d	I;II;
hsa-let-7e	HMDDv2.0;I;II;	hsa-mir-187	I;
hsa-let-7d	HMDDv2.0;I;	hsa-mir-205	I;
hsa-mir-34a	HMDDv2.0;I;	hsa-mir-93	HMDDv2.0;I;
hsa-let-7g	HMDDv2.0;I;	hsa-mir-34c	HMDDv2.0;I;
hsa-mir-200b	HMDDv2.0;I;	hsa-mir-15b	unconfirmed;
hsa-let-7b	HMDDv2.0;I;	hsa-mir-429	I;
hsa-mir-223	HMDDv2.0;I;	hsa-mir-142	unconfirmed;
hsa-mir-29a	HMDDv2.0;I;	hsa-mir-25	HMDDv2.0;III;
hsa-mir-29c	HMDDv2.0;I;	hsa-mir-106a	I;
hsa-mir-145	HMDDv2.0;I;II;III;	hsa-mir-373	I;II;
hsa-let-7c	HMDDv2.0;I;II;	hsa-mir-200c	HMDDv2.0;I;
hsa-let-7i	HMDDv2.0;I;	hsa-mir-302c	HMDDv2.0;I;III;
hsa-mir-146b	I;	hsa-mir-34b	I;
hsa-mir-127	HMDDv2.0;II;	hsa-mir-302d	I;II;
hsa-mir-106b	I;III;IV;	hsa-mir-191	I;
hsa-mir-200a	HMDDv2.0;I;II;	hsa-mir-150	I;
hsa-mir-126	HMDDv2.0;I;	hsa-mir-30e	HMDDv2.0;I;II;III;
hsa-mir-141	I;	hsa-mir-367	HMDDv2.0;I;
hsa-mir-135b	HMDDv2.0;I;III;	hsa-mir-215	I;
hsa-mir-125a	HMDDv2.0;I;II;III;	hsa-mir-19b	I;

I, II, and III represent dbDEMC, miR2Disease, and miRwayDB, respectively. The first and third columns record the 1–25 and 26–50 related miRNAs, respectively.

i.e., mir2disease (Jiang et al., 2009), dbDEMC (Yang et al., 2017), and miRwayDB (Das et al., 2018), which all record experimentally-validated miRNA-disease associations.

Lung neoplasms is one of the malignant tumors with the fastest increase in morbidity and mortality and the greatest threat to human health and life (Yanaihara et al., 2006). Therefore, there is an urgent need to identify prognostic and predictive markers for early detection. We used our method to uncover the potential miRNAs and listed the top 50 predicted candidate miRNAs. As a result (**Table 1**), 46 out of the top 50 miRNAs were verified to be associated with lung neoplasms by at least one database from Mir2disease, dbDEMC, and miRwayDB. For instance, studies have shown that hsa-mir-16(1st in **Table 1**) and hsa-mir-429 (3rd in **Table 1**) are closely related to the diagnosis and treatment of lung cancer (Reid et al., 2013; Ren et al., 2016).

To verify the potency of our method on real datasets, we conducted the second type of case study where we used older version of HMDD (v 1.0) as input to predict potential associations and test whether LLCMDA could uncover the newly-added ones in the latest version of HMDD (v 2.0). Specifically, HMDD v 1.0 contains 1,395 associations between 271 miRNAs and 137 diseases (Zhao et al., 2018). Here, we chose Lymphomas for validation. As shown in **Table 2**, 48 out of the top 50 candidate

miRNAs have been confirmed by dbDEMC, miR2Disease or/and miRwayDB. In particular, 31 miRNAs were found in HMDD 2.0. Taken together, these evidence further showed that our prediction method can effectively predict potential associations between miRNAs and diseases.

DISCUSSION

Nowadays, identifying potential disease-associated miRNAs could provide new insights into the role of miRNA as valuable biomarkers for clinical measure, diagnosis and treatment. However, it is impossible to predict the associations between miRNA-disease relying on traditional experimental-based methods. Consequently, great numbers of computational methods have been proposed to solve this challenging problem in recent years. In this paper, we presented a novel method to predict potential miRNA-disease associations based on locality-constrained liner coding. We first applied LLC algorithm to reconstruct similarity networks for miRNAs and diseases. The label propagation was then applied on the similarity networks to retrieve relevant scores for each miRNA-disease association. The final results were calculated as the average of the predicted results from both miRNA space and disease space, respectively. To comprehensively verify the performance of our method, we compared LLCMDA with five state-of-the-art computational model under four different cross-validation frameworks. The experimental results demonstrated powerful evidence that our method could effectively predict miRNA-disease associations. In addition, case studies on two common diseases also gave a strong confirmation to the prediction ability of our method.

The success of our method is mainly due to the following two reasons. First, the reconstructed similarity networks for both miRNAs and diseases are more robust as the LLC algorithm regards the local information in the coding process. Second, we applied label propagation on the reconstructed similarity networks as well as the original similarity networks to calculate reliable relevant scores for the final output. Nonetheless, more informative data sources should be integrated into our model to further improve the prediction performance. Besides, the final outcome was simply taken as the average from the prediction scores from different similarity networks, which may lead to sub-optimal results. Therefore, a more appropriate way to incorporate the prediction results needs to be put forward.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

REFERENCES

Ambros, V. (2004). The functions of animal microRNAs. *Nature* 431, 350–355. doi: 10.1038/nature02871

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. 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., Niu, Y. W., Wang, G. H., and Yan, G. Y. (2017a). MKRMDA: multiple kernel learning-based Kronecker regularized least squares for MiRNA-disease association prediction. *J. Transl. Med.* 15:251. doi: 10.1186/s12967-017-1340-3

Chen, X., Qu, J., and Yin, J. (2018c). TLHNMDA: triple layer heterogeneous network based inference for MiRNA-disease association prediction. *Front. Genet.* 9:234. doi: 10.3389/fgene.2018.00234

Chen, X., Wu, Q. F., and Yan, G. Y. (2017b). RKNNMDA: ranking-based KNN for MiRNA-disease association prediction. *RNA Biol.* 14, 952–962. doi: 10.1080/15476286.2017.1312226

Chen, X., Xie, D., Wang, L., Zhao, Q., You, Z. H., and Liu, H. (2018d). BNPMDA: bipartite network projection for MiRNA-disease association prediction. *Bioinformatics* 34, 3178–3186. doi: 10.1093/bioinformatics/bty333

Chen, X., Xie, D., Zhao, Q., and You, Z. H. (2017c). MicroRNAs and complex diseases: from experimental results to computational models. *Brief. Bioinform.* doi: 10.1093/bib/bbx130. [Epub ahead of print].

Chen, X., Yan, C. C., Zhang, X., You, Z. H., Deng, L., Liu, Y., et al. (2016a). 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., You, Z. H., Huang, Y. A., and Yan, G. Y. (2016b). HGIMDA: heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget* 7, 65257–65269. doi: 10.18632/oncotarget.11251

Chen, X., and Yan, G. Y. (2014). Semi-supervised learning for potential human microRNA-disease associations inference. *Sci. Rep.* 4:5501. doi: 10.1038/srep05501

Cheng, A. M., Byrom, M. W., Shelton, J., and Ford, L. P. (2005). Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res.* 33, 1290–1297. doi: 10.1093/nar/gki200

Das, S. S., Saha, P., and Chakravorty, N. (2018). miRwayDB: a database for experimentally validated microRNA-pathway associations in pathophysiological conditions. *Database.* doi: 10.1093/database/bay023

Fu, L., and Peng, Q. (2017). A deep ensemble model to predict miRNA-disease association. *Sci. Rep.* 7:14482. doi: 10.1038/s41598-017-15235-6

Huang, T., Li, B. Q., and Cai, Y. D. (2016). The integrative network of gene expression, microrna, methylation and copy number variation in colon and rectal cancer. *Curr. Bioinformat.* 11, 59–65. doi: 10.2174/1574893611666151119215823

Jiang, Q., Hao, Y., Wang, G., Juan, L., Zhang, T., Teng, M., et al. (2010). Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst. Biol.* 4 (Suppl. 1):S2. doi: 10.1186/1752-0509-4-S1-S2

Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., et al. (2009). miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* 37, D98–104. doi: 10.1093/nar/gkn714

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854. doi: 10.1016/0092-8674(93)90529-Y

Li, G. H., Luo, J. W., Xiao, Q., Liang, C., and Ding, P. J. (2018). Prediction of microRNA-disease associations with a Kronecker kernel matrix dimension reduction model. *RSC Adv.* 8, 4377–4385. doi: 10.1039/C7RA12491K

Li, J. Q., Rong, Z. H., Chen, X., Yan, G. Y., and You, Z. H. (2017). MCMDA: Matrix completion for MiRNA-disease association prediction. *Oncotarget* 8, 21187–21199. doi: 10.18632/oncotarget.15061

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

Liao, Z. J., Li, D. P., Wang, X. R., Li, L. S., and Zou, Q. (2018). Cancer diagnosis through isomir expression with machine learning method. *Curr. Bioinf.* 13, 57–63. doi: 10.2174/1574893611666160609081155

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. T Comput. Biol.* 14, 905–915. doi: 10.1109/TCBB.2016.2550432

Luo, J. W., and Xiao, Q. (2017). A novel approach for predicting microRNA-disease associations by unbalanced bi-random walk on heterogeneous network. *J. Biomed. Inform.* 66, 194–203. doi: 10.1016/j.jbi.2017.01.008

Mei, Q. L., Zhang, H. X., and Liang, C. (2016). A discriminative feature extraction approach for tumor classification using gene expression data. *Curr. Bioinf.* 11, 561–570. doi: 10.2174/1574893611666160728114747

Miska, E. A. (2005). How microRNAs control cell division, differentiation and death. *Curr. Opin. Genet. Dev.* 15, 563–568. doi: 10.1016/j.gde.2005.08.005

Mohammadi-Yeganeh, S., Paryan, M., Samiee, S. M., Soleimani, M., Arefian, E., Azadmanesh, K., et al. (2013). Development of a robust, low cost stem-loop real-time quantification PCR technique for miRNA expression analysis. *Mol. Biol. Rep.* 40, 3665–3674. doi: 10.1007/s11033-012-2442-x

Mørk, S., Pletscher-Frankild, S., Palleja Caro, A., Gorodkin, J., and Jensen, L. J. (2014). Protein-driven inference of miRNA-disease associations, *Bioinformatics* 30, 392–397. doi: 10.1093/bioinformatics/btt677

Qu, Y., Zhang, H., Liang, C., Ding, P., and Luo, J. (2018a). SNMDA: a novel method for predicting microRNA-disease associations based on sparse neighbourhood. *J. Cell. Mol. Med.* 22, 5109–5120. doi: 10.1111/jcmm.13799

Qu, Y., Zhang, H. X., Liang, C., and Dong, X. (2018b). KATZMDA: prediction of miRNA-disease associations based on KATZ Model. *IEEE Access* 6, 3943–3950. doi: 10.1109/ACCESS.2017.2754409

Reid, G., Pel, M. E., Kirschner, M. B., Cheng, Y. Y., Mugridge, N., Weiss, J., et al. (2013). Restoring expression of miR-16: a novel approach to therapy for malignant pleural mesothelioma. *Ann. Oncol.* 24, 3128–3135. doi: 10.1093/annonc/mdt412

Ren, Z., Tong, H. W., Chen, L., Yao, Y. F., Huang, S. C., Zhu, F. J., et al. (2016). miR-211 and miR-429 are involved in Emodin's anti-proliferative effects on lung cancer. *Int. J. Clin. Exp. Med.* 9, 2085–2093.

Saffari, S. A., and Ebrahimi-Moghadam, A. (2015). Label propagation based on local information with adaptive determination of number and degree of neighbor's similarity. *Neurocomputing* 153, 41–53. doi: 10.1016/j.neucom.2014.11.053

Shi, H., Xu, J., Zhang, G., Xu, L., Li, C., Wang, L., et al. (2013). Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes. *BMC Syst. Biol.* 7:101. doi: 10.1186/1752-0509-7-101

Tang, W., Wan, S. X., Yang, Z., Teschendorff, A. E., and Zou, Q. (2018). Tumor origin detection with tissue-specific miRNA and DNA methylation markers. *Bioinformatics* 34, 398–406. doi: 10.1093/bioinformatics/btx622

Thomson, J. M., Parker, J. S., and Hammond, S. M. (2007). Microarray analysis of miRNA gene expression. *Methods Enzymol.* 427, 107–122. doi: 10.1016/S0076-6879(07)27006-5

Wang, D., Wang, J., Lu, M., Song, F., and Cui, Q. (2010a). Inferring the human microRNA functional similarity and functional network

based on microRNA-associated diseases. *Bioinformatics* 26, 1644–1650. doi: 10.1093/bioinformatics/btq241

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., and Gong, Y. (2010b). "Locality-constrained Linear Coding for image classification", in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition IEEE Computer Society Conference on CVPRW* (San Francisco, CA), 3360–3367.

Xiao, Q., Luo, J. W., Liang, C., Cai, J., and Ding, P. J. (2018). A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics* 34, 239–248. doi: 10.1093/bioinformatics/btx545

Xu, P., Guo, M., and Hay, B. A. (2004). MicroRNAs and the regulation of cell death. *Trends Genet.* 20, 617–624. doi: 10.1016/j.tig.2004.09.010

Xuan, P., Han, K., Guo, M., Guo, Y., Li, J., Ding, J., et al. (2013). Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLoS ONE* 8:e70204. doi: 10.1371/journal.pone.0070204

Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., et al. (2006). Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell.* 9, 189–198. doi: 10.1016/j.ccr.2006.01.025

Yang, Z., Wu, L., Wang, A., Tang, W., Zhao, Y., Zhao, H., et al. (2017). dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res.* 45, D812–D818. doi: 10.1093/nar/gkw1079

You, Z. H., Huang, Z. A., Zhu, Z., Yan, G. Y., Li, Z. W., Wen, Z., et al. (2017). 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

Zang, F., and Zhang, J. S. (2012). Label propagation through sparse neighborhood and its applications. *Neurocomputing* 97, 267–277. doi: 10.1016/j.neucom.2012.03.017

Zeng, X., Liu, L., Lu, L., and Zou, Q. (2018). Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics* 34, 2425–2432. doi: 10.1093/bioinformatics/bty112

Zeng, X., Zhang, X., and Zou, Q. (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, W., Chen, Y. L., and Li, D. F. (2017). Drug-target interaction prediction through label propagation with linear neighborhood information. *Molecules* 22:E2056. doi: 10.3390/molecules22122056

Zhang, W., Qu, Q. L., Zhang, Y. Q., and Wang, W. (2018). The linear neighborhood propagation method for predicting long non-coding RNA-protein interactions. *Neurocomputing* 273, 526–534. doi: 10.1016/j.neucom.2017.07.065

Zhao, Y., Chen, X., and Yin, J. (2018). A novel computational method for the identification of potential miRNA-disease association based on symmetric non-negative matrix factorization and kronecker regularized least square. *Front. Genet.* 9:324. doi: 10.3389/fgene.2018.00324

Zhou, D., Bousquet, O., Lal, T. N., and Weston, J. (2003). "Learning with local and global consistency," in *NIPS'03 Proceedings of the 16th International Conference on Neural Information Processing Systems* (Whistler, BC), 321–328.

Zhu, L., Huang, Z., Li, Z., Xie, L., and Shen, H. T. (2018). Exploring auxiliary context: discrete semantic transfer hashing for scalable image retrieval. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 5264–5276. doi: 10.1109/TNNLS.2018.2797248

Zou, Q., Chen, L., Huang, T., Zhang, Z., and Xu, Y. (2017). Machine learning and graph analytics in computational biomedicine. *Arti. Intell. Med.* 83:1. doi: 10.1016/j.artmed.2017.09.003

Zou, Q., Li, J., Hong, Q., Lin, Z., Wu, Y., Shi, H., et al. (2015). Prediction of MicroRNA-disease associations based on social network analysis methods. *Biomed. Res. Int.* 2015:810514. doi: 10.1155/2015/810514

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