

RESEARCH ARTICLE

# Adaptive multi-view multi-label learning for identifying disease-associated candidate miRNAs

Cheng Liang<sup>1\*</sup>, Shengpeng Yu<sup>1</sup>, Jiawei Luo<sup>2\*</sup>

**1** School of Information Science and Engineering, Shandong Normal University, Jinan, China, **2** College of Computer Science and Electronic Engineering, Hunan University, Changsha, China

\* [alcs417@sdnu.edu.cn](mailto:alcs417@sdnu.edu.cn) (CL); [luojiawei@hnu.edu.cn](mailto:luojiawei@hnu.edu.cn) (JL)



## Abstract

Increasing evidence has indicated that microRNAs(miRNAs) play vital roles in various pathological processes and thus are closely related with many complex human diseases. The identification of potential disease-related miRNAs offers new opportunities to understand disease etiology and pathogenesis. Although there have been numerous computational methods proposed to predict reliable miRNA-disease associations, they suffer from various limitations that affect the prediction accuracy and their applicability. In this study, we develop a novel method to discover disease-related candidate miRNAs based on Adaptive Multi-View Multi-Label learning(AMVML). Specifically, considering the inherent noise existed in the current dataset, we propose to learn a new affinity graph adaptively for both diseases and miRNAs from multiple similarity profiles. We then simultaneously update the miRNA-disease association predicted from both spaces based on multi-label learning. In particular, we prove the convergence of AMVML theoretically and the corresponding analysis indicates that it has a fast convergence rate. To comprehensively illustrate the prediction performance of our method, we compared AMVML with four state-of-the-art methods under different validation frameworks. As a result, our method achieved comparable performance under various evaluation metrics, which suggests that our method is capable of discovering greater number of true miRNA-disease associations. The case study conducted on thyroid neoplasms further identified a potential diagnostic biomarker. Together, the experimental results confirms the utility of our method and we anticipate that our method could serve as a reliable and efficient tool for uncovering novel disease-related miRNAs.

## OPEN ACCESS

**Citation:** Liang C, Yu S, Luo J (2019) Adaptive multi-view multi-label learning for identifying disease-associated candidate miRNAs. *PLoS Comput Biol* 15(4): e1006931. <https://doi.org/10.1371/journal.pcbi.1006931>

**Editor:** Ilya Ioshikhes, Ottawa University, CANADA

**Received:** January 3, 2019

**Accepted:** March 5, 2019

**Published:** April 1, 2019

**Copyright:** © 2019 Liang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files. The datasets and source code of the proposed method is also freely available at <https://github.com/alcs417/AMVML>.

**Funding:** CL was supported by National Nature Science Foundation of China (Grant No. 61602283) and Natural Science Foundation of Shandong (Grant No. ZR2016FB10). JWL was supported by National Nature Science Foundation of China (Grant No. 61572180, 61873089). The funders had no role in study design, data collection and

## Author summary

MiRNAs are a class of small non-coding RNAs that are associated with a variety of complex biological processes. Increasing studies have shown that miRNAs have close relationships with many human diseases. The prediction of the associations between miRNAs and diseases has thus become a hot topic. Although traditional experimental methods are reliable, they could only identify a limited number of associations as they are in general time-

analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

consuming and expensive. Consequently, great efforts have been made to effectively predict reliable disease-related miRNAs based on computational methods. In this study, we develop a novel method to discover potential miRNA-disease associations based on Adaptive Multi-View Multi-Label learning. Considering the inherent noise existed in the current dataset, we propose to learn a new affinity graph adaptively for both diseases and miRNAs from multiple biological data source, including miRNA sequence similarity, miRNA functional similarity and Gaussian interaction profile kernel similarity. Notably, our method is applicable to diseases without any known associated miRNAs and also obtains satisfactory results. The case study conducted on thyroid neoplasms further confirms the prediction reliability of the proposed method. Overall, results show that our method can predict the potential associations between miRNAs and diseases effectively.

## Introduction

MiRNAs are a group of short non-coding RNAs that mediate post-transcriptional gene silencing[1]. Accumulating evidence has proved that miRNAs play crucial roles in a variety of cancer-related pathways. Therefore, the identification of miRNA-disease associations can shed new light on understanding possible pathogenesis of diseases.

To compensate for the limitations of experiment-based approaches, a great number of computational models have been proposed to identify potential disease-related miRNAs in recent years[2]. Under the assumption that functionally similar miRNAs tend to be associated with phenotypically similar diseases, Jiang *et al.* prioritized the entire microRNAome for over a thousand diseases by constructing an integrated phenome-microRNAome network[3]. Chen *et al.* measured the global network similarity and inferred potential miRNA-disease interactions based on random walk with restart[4]. Shi *et al.* adopted a similar idea and further integrated the protein-protein interactions into the prediction process[5]. Chen *et al.* proposed a novel heterogeneous graph inference method by iteratively updating the association probability[6, 7]. Liu *et al.* constructed a heterogeneous network in which they integrated the miRNA-target gene and miRNA-lncRNA associations[8]. Specifically, the methods introduced above mainly predicted disease-related miRNAs by applying random walk algorithms to the reconstructed similarity networks[9]. Another family of prediction methods was generally based on network topological characteristics and also achieved remarkable performance. For instance, Zou *et al.* computed the similarity score based on walks of different lengths between the miRNA and disease nodes[10]. Sun *et al.* exploited the potential disease-related miRNAs based on known miRNA-disease network topological similarity[11]. You *et al.* proposed to measure the association score for a miRNA-disease pair by calculating the accumulative contributions from all paths between them[12]. Li *et al.* used DeepWalk to enhance the existing associations through a topology-based similarity measure[13]. Chen *et al.* computed the association possibility between a disease node and a miRNA node in the corresponding graphlet interaction isomers[14]. Although effective, these methods are sensitive to the change of the network topological structures, which might affect the prediction accuracy. Alternatively, prediction methods that were based on semi-supervised learning as well as supervised learning have been well developed. Xiao *et al.* introduced a graph regularized non-negative matrix factorization to effectively discover sparse miRNA-disease associations[15]. Both Chen *et al.* and Yu *et al.* adopted matrix completion to recover the potential missing miRNA-disease associations[16, 17]. Zeng *et al.* used a derivative algorithm structural perturbation method to estimate the link predictability with structural consistency as the indicator[18]. Chen *et al.* used an ensemble

model where a sequence of weak learners were trained to collectively obtain a predicted association score[19]. Recently, we reconstructed the miRNA and disease similarity matrices based on global linear neighborhoods and then applied label propagation to predict potential associations between diseases and miRNAs[20, 21]. Chen *et al.* extracted novel feature vectors for both miRNAs and diseases to train a random forest classifier for the prediction task[22].

Although great efforts have been made to efficiently uncover potential miRNA-disease associations, most existing computational approaches still suffer from several limitations. Specifically, the inherent noise in the current datasets resulted in incomplete and sparse similarity matrices and thus inevitably affected the prediction accuracies of these methods. Moreover, the integration of multiple biological data sources in calculating the similarity matrices for both miRNAs and diseases was generally performed by averaging the input similarity information, which might lead to suboptimal results. Lastly, the predicted association scores from miRNA space and disease space were often updated separately during the learning process. To solve these problems, in this paper, we propose a novel Adaptive Multi-View Multi-Label (AMVML) learning framework to infer disease-related miRNAs. In particular, our method adaptively learns a new affinity graph for miRNAs and diseases respectively from multiple data sources (i.e. miRNA sequence similarity, Gaussian interaction profile kernel similarity and so on). In addition, we unify the optimization process for both disease space and miRNA space based on multi-label learning. The experimental results under several different evaluation metrics clearly demonstrate the superior performance of our method over previous methods. We further carry out a case study on thyroid cancer to identify potential prognostic biomarkers.

## Materials and methods

### Human miRNA-disease associations

The known human miRNA-disease associations were retrieved from HMDD v2.0 database [23]. HMDD is a database for experimentally supported human miRNA and disease associations that were manually collected from all the miRNA-related publications in PubMed. Each entry in HMDD contains four items, i.e. miRNA name, disease name, experimental evidence for the miRNA-disease association and the publication PubMed ID. To keep consistent of data from different sources, we also downloaded the annotation information of 4796 human miRNAs released on March 2018 from miRBase[24]. We then downloaded the latest MeSH descriptors from the National Library of Medicine(<https://www.nlm.nih.gov/>) and only retained items from Category C for diseases, which resulted in 11572 unique items. After mapping the miRNA names and disease names involved in each association with miRBase records and MeSH descriptors, we finally obtained 6088 associations between 328 diseases and 550 miRNAs for subsequent analysis(S1 File). Specifically, we classified the 328 diseases based on the Diseases Categories provided in MeSH. For diseases belonging to multiple categories, we increased the count by one for each category accordingly. As a result (Fig 1, S2 File), we can see that most diseases recorded in HMDD were cancers. For convenience, we used a binary matrix  $Y \in \mathbb{R}^{328 \times 550}$  to represent the miRNA-disease associations. For a given disease  $i$  and miRNA  $j$ ,  $Y_{ij} = 1$  if  $i$  is related to  $j$ , and  $Y_{ij} = 0$  otherwise.

### Disease semantic similarity

As described in [25], the disease semantic similarity can be calculated based on Directed Acyclic Graphs (DAGs). Specifically, for a given disease  $d$ , its DAG is composed of three items, i.e.  $DAG = (d, T(d), E(d))$ , where  $T(d)$  represents  $d$  itself together with all its ancestor nodes, and  $E(d)$  contains all direct links connecting the parent nodes to child nodes. The contribution  $D_d(t)$

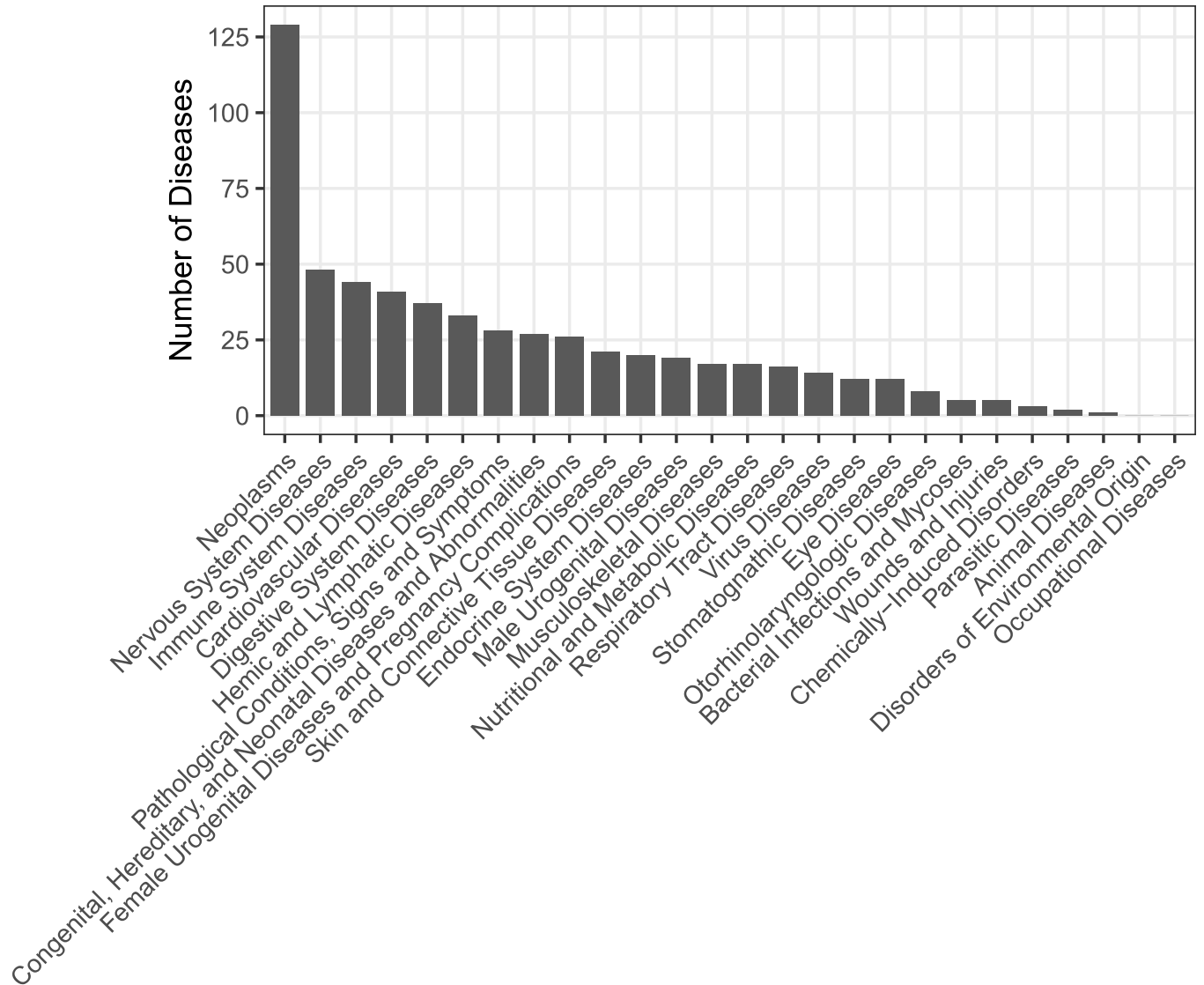


Fig 1. The type and distribution of the 328 diseases recorded in HMDD v2.0.

<https://doi.org/10.1371/journal.pcbi.1006931.g001>

of a disease  $t$  in a  $DAG_d$  to the semantics of disease  $d$  was defined as follows:

$$\begin{cases} D_d(D) = 1 \\ D_d(t) = \max\{0.5 * D_d(t') | t' \in \text{children of } t\} \text{ if } t \neq d \end{cases} \quad (1)$$

The semantic similarity score between two diseases  $i$  and  $j$  can then be calculated by:

$$S(i, j) = \frac{\sum_{t \in T(i) \cap T(j)} (D_i(t) + D_j(t))}{\sum_{t \in T(i)} D_i(t) + \sum_{t \in T(j)} D_j(t)} \quad (2)$$

Moreover, the similarity between a given disease  $d$  and a group of diseases  $D_t = \{d_{t1}, d_{t2}, \dots, d_{tk}\}$  was defined by:

$$S(d, D_t) = \max_{1 \leq i \leq k} (S(d, d_{ti})) \tag{3}$$

Finally, we obtained the semantic similarities for each disease pair according to (Eq 2). We denoted the semantic similarity matrix as  $AD^{(1)} \in \mathbb{R}^{328 \times 328}$  where  $AD_{ij}^{(1)}$  represents the semantic similarity between disease  $i$  and disease  $j$  (S3 File).

### MiRNA similarity measures

In this subsection, to comprehensively characterize similarities between miRNAs, we adopt three measures using different biological data sources for subsequent predictions[26].

**MiRNA sequence similarity.** The sequence information of all miRNAs were downloaded from miRBase[24]. We then used the "pairwiseAlignment" function in R package Biostrings to calculate a similarity score for each miRNA pair based on their entire mature sequences with a gap opening penalty of 5 and a gap extension penalty of 2. Moreover, to generate a substitution matrix for sequence alignment, we set the match score to 1 and the mismatch score to -1. Finally, the sequence similarity score obtained for each miRNA pair was further normalized to the range [0, 1] by the following equation:

$$Score(i, j) = \frac{Score(i, j) - Score_{min}}{Score_{max} - Score_{min}} \tag{4}$$

where  $Score_{min}$  and  $Score_{max}$  represent the minimum and maximum similarity score of all miRNA pairs. For simplicity, we use  $AM^{(1)} \in \mathbb{R}^{328 \times 550}$  to denote the sequence similarity matrix where  $AM_{ij}^{(1)}$  represents the sequence similarity between miRNA  $i$  and miRNA  $j$  (S4 File).

**MiRNA functional similarity.** To measure the functional similarity for each miRNA pair, we followed the same pipeline presented in [25]. Let  $DT_i$  and  $DT_j$  denote the disease sets related to miRNA  $i$  and  $j$ , respectively, the functional similarity is calculated as follows:

$$MFS(i, j) = \frac{\sum_{1 \leq p \leq |DT_i|} S(dt_p, DT_j) + \sum_{1 \leq q \leq |DT_j|} S(dt_q, DT_i)}{|DT_i| + |DT_j|} \tag{5}$$

where  $S(dt, DT)$  is the same as that defined in (Eq 3). We use  $AM^{(2)} \in \mathbb{R}^{550 \times 550}$  to denote the miRNA functional similarity matrix and  $AM_{ij}^{(2)}$  represents the functional similarity between miRNA  $i$  and miRNA  $j$  (S5 File).

**MiRNA semantic similarity.** As stated in previous section, the miRNA functional similarities could be obtained based on the overlap of miRNA-related diseases[15]. However, it is not applicable to miRNAs without any known associated diseases. Therefore, we here propose to use miRNA target information and the Gene Ontology (GO) annotations to better describe the miRNA semantic similarities. To this end, we first downloaded the experimentally-verified miRNA-gene interactions from mirTarBase[27], which contains 380639 interactions between 2599 miRNAs and 15064 genes. For each miRNA pair in our analysis, we maintained their target gene lists and then calculated the semantic similarity between the two corresponding gene groups by using the "clusterSim" function in the R package GOsemSim[28]. Specifically, the GO annotations were retrieved from the Bioconductor package "org.Hs.eg.db" and "BMA" method was used for combining semantic similarity scores of multiple GO terms[29]. Similarly, we used  $AM^{(3)} \in \mathbb{R}^{550 \times 550}$  to denote the miRNA semantic similarity matrix where  $AM_{ij}^{(3)}$  represents the semantic similarity between miRNA  $i$  and miRNA  $j$  (S6 File).

### Gaussian interaction profile kernel similarity

Gaussian interaction profile kernel similarity has been widely used in previous studies and proved effective in measuring both miRNA and disease similarities. For a given miRNA  $i$  or disease  $j$ , its interaction profile  $IP(m_i)$  or  $IP(d_j)$  was a binary vector extracted from the  $i$ -th row or the  $j$ -th column of the association matrix  $Y$ . The kernel similarity between two miRNAs or two diseases could then be computed by:

$$KM(m_i, m_j) = \exp\left(-\beta_m \left\|IP(m_i) - IP(m_j)\right\|^2\right) \tag{6}$$

$$KD(d_i, d_j) = \exp\left(-\beta_d \left\|IP(d_i) - IP(d_j)\right\|^2\right) \tag{7}$$

where  $\beta_m$  and  $\beta_d$  are defined as follows:

$$\beta_m = \beta'_m / \left(\frac{1}{550} \sum_{i=1}^{550} \left\|IP(m_i)\right\|^2\right) \tag{8}$$

$$\beta_d = \beta'_d / \left(\frac{1}{328} \sum_{i=1}^{328} \left\|IP(d_i)\right\|^2\right) \tag{9}$$

where  $\beta'_m$  and  $\beta'_d$  are two parameters controlling the kernel bandwidth. As a result, we used  $AM^{(4)} \in \mathbb{R}^{550 \times 550}$  and  $AD^{(2)} \in \mathbb{R}^{328 \times 328}$  to represent the obtained Gaussian interaction profile similarity matrices for miRNAs and diseases, respectively.

### Adaptive multi-view multi-label learning for miRNA-disease association prediction

We summarize the notations used throughout this paper. Given a matrix  $M$ ,  $M_{ij}$  and  $M_i$  represent its  $ij$ -th element and  $i$ -th row, respectively. The transpose of  $M$  is denoted by  $M^T$ .  $Tr(M)$  denotes the trace of  $M$  and the Frobenius norm of  $M$  is represented as  $\|M\|_F$ . For a similarity matrix  $S$ , its Laplacian matrix  $L_S$  is defined as  $L_S = D_S - \frac{S^T+S}{2}$ , where  $D_S$  is a diagonal matrix with its  $i$ -th diagonal element equal to  $\sum_j(S_{ij} + S_{ji})/2$ .

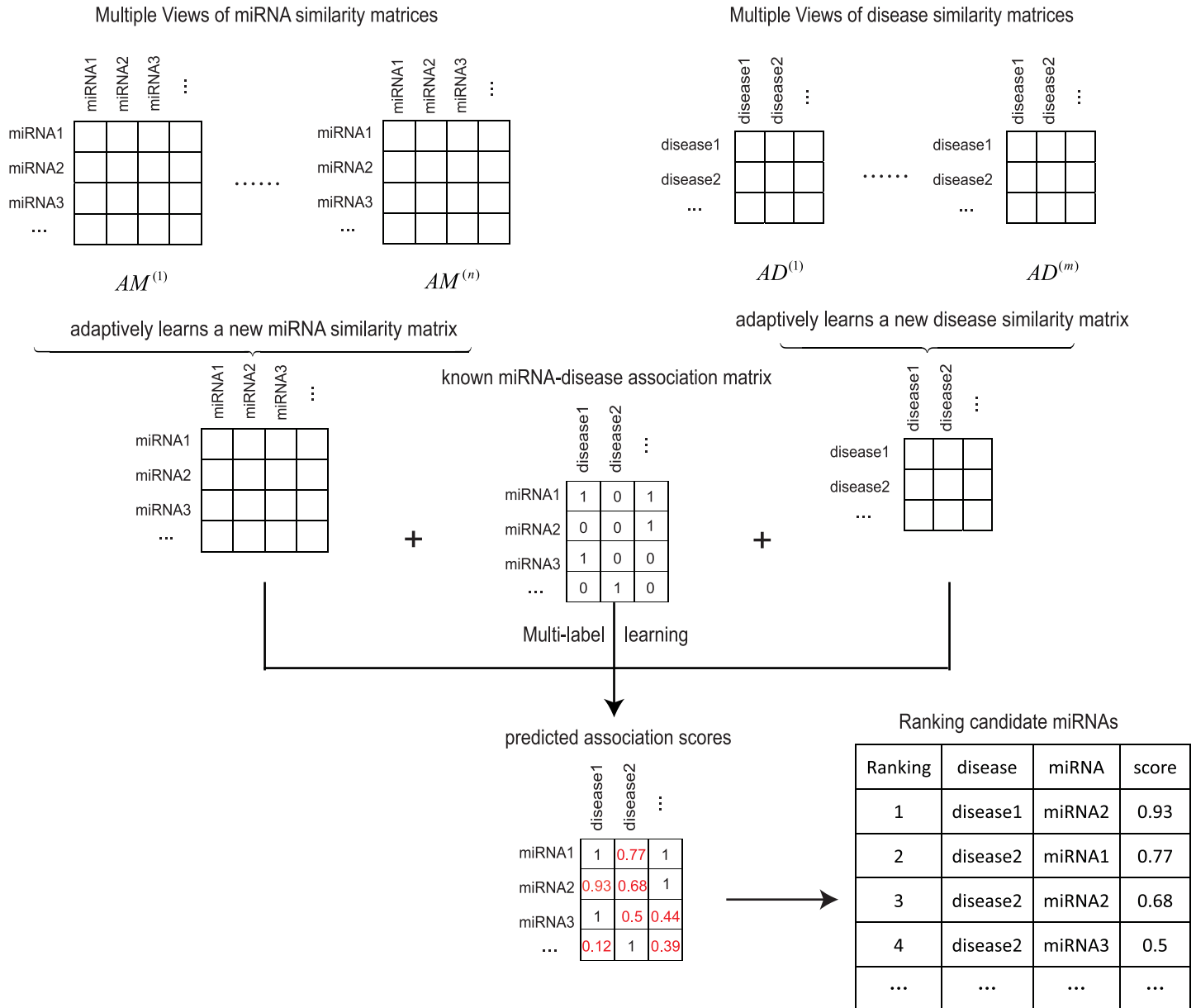
**Graph-based multi-label learning.** Multi-label learning refers to the problems where an instance can be assigned to more than one category[30]. The graph-based multi-label learning framework is characterized by simultaneously exploiting the inherent correlations among multiple labels and the label consistency over the graph[31]. As a matter of fact, since each miRNA or disease could be associated with multiple diseases or miRNAs, this learning framework can be directly applied to solve the miRNA-disease association prediction problem by defining its objective function as follows:

$$\min_F Tr((F - Y)^T(F - Y)) + \mu Tr(F^T L F) + \nu Tr(F C F^T) \tag{10}$$

where  $L$  and  $C$  are the normalized Laplacian matrices corresponding to the similarity matrices of miRNAs and diseases, respectively.  $\mu$  and  $\nu$  are two non-negative trade-off parameters. By differentiating the objective function with respect to  $F$ , (Eq 10) can be efficiently solved by a Sylvester equation.

**AMVML.** The graph-based multi-label learning provides us a unified framework to collaboratively update the prediction results from miRNA space and disease space, which





**Fig 2. The overall workflow of AMVML.** Our method first learns two new affinity graphs for miRNAs and diseases from multiple sources of biological datasets, respectively. It then updates the miRNA-disease association information simultaneously based on multi-label learning framework. It repeats this process until the algorithm converges and finally outputs the prediction results.

<https://doi.org/10.1371/journal.pcbi.1006931.g002>

successfully solves the last issue mentioned above. However, we still suffer from the inherent noises in the existing datasets as well as the lack of appropriate methods to integrate datasets from multiple biological data sources. To conquer these limitations, we propose a new objective function which can adaptively learn new affinity graphs for miRNAs and diseases given similarity information obtained from multiple views, respectively (Fig 2). Moreover, instead of explicitly assigning weights for each view, our method can also perform self-conducted weight learning during the optimization process [32]. Specifically, let  $n$  be the number of views in miRNA space and  $AM^{(1)}, AM^{(2)}, \dots, AM^{(n)}$  be the corresponding miRNA similarity matrix of each view, where  $AM^{(u)} \in \mathbb{R}^{p \times p}$  ( $1 \leq u \leq n$ ). Similarly, let  $m$  be the number of views in disease

space and  $AD^{(1)}, AD^{(2)}, \dots, AD^{(m)}$  be the corresponding disease similarity matrix of each view, where  $AD^{(v)} \in \mathbb{R}^{q \times q}$  ( $1 \leq v \leq m$ ).  $p$  and  $q$  are the number of miRNAs and diseases, respectively. Our objective is to obtain the predicted association matrix  $F$  as well as two optimal similarity matrices  $SD$  and  $SM$  by considering multiple input views in both disease and miRNA spaces simultaneously, which is formulated as follows:

$$\begin{aligned} \min_{SD, SM, F} & \sum_{v=1}^m w_D^{(v)} \left\| SD - AD^{(v)} \right\|_F^2 + 2\alpha(F^T L_{SD} F) + \\ & \sum_{u=1}^n w_M^{(u)} \left\| SM - AM^{(u)} \right\|_F^2 + 2\beta(F L_{SM} F^T) + \left\| F - Y \right\|_F^2 \end{aligned} \tag{11}$$

*s.t.*  $SD_i \mathbf{1} = 1, 0 \leq SD_{ij} \leq 1, SM_i \mathbf{1} = 1, 0 \leq SM_{ij} \leq 1, F \in \mathbb{R}^{q \times p}$

where  $L_{SD}$  and  $L_{SM}$  are the corresponding Laplacian matrices for  $SD$  and  $SM$ , respectively.  $\mathbf{1}$  is a column vector with all elements equal to 1.  $w_D^{(v)}$  and  $w_M^{(u)}$  are weight parameters for the  $v$ -th view of disease similarity and  $u$ -th view of miRNA similarity defined by:

$$w_D^{(v)} = 1 / \left( 2 \left\| SD - AD^{(v)} \right\|_F \right) \tag{12}$$

$$w_M^{(u)} = 1 / \left( 2 \left\| SM - AM^{(u)} \right\|_F \right) \tag{13}$$

As can be seen from Eqs (12) and (13), the two weight factors  $w_D^{(v)}$  and  $w_M^{(u)}$  are seamlessly coupled with  $\left\| SD - AD^{(v)} \right\|_F$  and  $\left\| SM - AM^{(u)} \right\|_F$ , respectively. Intuitively, if the  $v$ -th or  $u$ -th view is good, then  $\left\| SD - AD^{(v)} \right\|_F$  or  $\left\| SM - AM^{(u)} \right\|_F$  should be small and thus the learnt  $w_D^{(v)}$  or  $w_M^{(u)}$  will be assigned a larger weight accordingly. As a result,  $w_D^{(v)}$  and  $w_M^{(u)}$  are updated adaptively in terms of the quality of the corresponding view during each iteration, which essentially makes the optimization of our objective function a self-weighted learning process. In the next part, we propose an efficient algorithm to solve (Eq 11).

**Optimization.** It is difficult to directly solve (Eq 11) as it involves three variables. Therefore, we iteratively optimize one variable by fixing the others[33].

1. Solving  $SD$ . When  $SM$  and  $F$  are fixed, (Eq 11) becomes:

$$\sum_{v=1}^m w_D^{(v)} \left\| SD - AD^{(v)} \right\|_F^2 + 2\alpha(F^T L_{SD} F), \text{ s.t. } SD_i \mathbf{1} = 1, 0 \leq SD_{ij} \leq 1 \tag{14}$$

(Eq 14) can be further transformed into:

$$\min_{0 < SD_{ij} < 1, SD_i \mathbf{1} = 1} \sum_{v=1}^m w_D^{(v)} \sum_{i,j=1}^q (SD_{ij} - AD_{ij}^{(v)})^2 + \alpha \sum_{i,j=1}^q \left\| f_i - f_j \right\|_2^2 SD_{ij} \tag{15}$$

Since (Eq 15) is independent for different  $i$ , we can optimize each row separately:

$$\min_{0 < SD_{ij} < 1, SD_i \mathbf{1} = 1} \sum_{j=1}^q \sum_{v=1}^m w_D^{(v)} (SD_{ij} - AD_{ij}^{(v)})^2 + \alpha \sum_{j=1}^q \left\| f_i - f_j \right\|_2^2 SD_{ij} \tag{16}$$



Denoting  $z_i$  as a vector with its  $j$ -th element  $z_{ij} = \|f_i - f_j\|_2^2$  and similarly for  $SD_i$  and  $AD_i^{(v)}$ , (Eq 16) could be rewritten as:

$$\min_{SD_i > 0_p^T, SD_i \mathbf{1} = 1} \left\| SD_i - \left( \sum_{v=1}^m w_D^{(v)} AD_i^{(v)} - \frac{\alpha}{2} z_i \right) / \sum_{v=1}^m w_D^{(v)} \right\|_2^2 \quad (17)$$

Problem (17) can be solved efficiently by an iterative algorithm[34].

2. Solving  $SM$ . Similarly, when  $SD$  and  $F$  are fixed, (Eq 11) becomes:

$$\sum_{u=1}^n w_M^{(u)} \left\| SM - AM^{(u)} \right\|_F^2 + 2\beta (FL_{SM} F^T), \text{ s.t. } SM_i \mathbf{1} = 1, 0 \leq SM_{ij} \leq 1 \quad (18)$$

By following the same optimization process for  $SD$ , we can derive the solution for problem (18) as follows:

$$\min_{SM_i > 0_p^T, SM_i \mathbf{1} = 1} \left\| SM_i - \left( \sum_{u=1}^n w_M^{(u)} AM_i^{(u)} - \frac{\beta}{2} z_i \right) / \sum_{u=1}^n w_M^{(u)} \right\|_2^2 \quad (19)$$

3. Solving  $F$ . When  $SD$  and  $SM$  are fixed, (Eq 11) degenerates to:

$$\min_F 2\alpha (F^T L_{SD} F) + 2\beta (FL_{SM} F^T) + \left\| F - Y \right\|_F^2 \quad (20)$$

Taking the derivative of (Eq 20) with respect to  $F$  and setting it to zero, we could obtain:

$$(\beta L_{SM} + I)F + \alpha FL_{SD} = Y \quad (21)$$

(Eq 21) is a Sylvester equation and could be solved directly. We summarized the overall procedure in **Algorithm 1**. Besides, the datasets and source code of AMVML are freely available at <https://github.com/alcs417/AMVML>.

**Algorithm 1: AMVML**

**Input:** miRNA similarity matrices of  $n$  views  $\{AM^{(1)}, AM^{(2)}, \dots, AM^{(n)}\}$ , disease similarity matrices of  $m$  views  $\{AD^{(1)}, AD^{(2)}, \dots, AD^{(m)}\}$ , known association matrix  $Y \in \mathbb{R}^{q \times p}$ , parameter  $\alpha$  and  $\beta$ .

**Output:** Predicted association matrix  $F$ .

1. Initialize the weights of each view for both miRNAs and diseases by  $w_D^{(v)} = \frac{1}{m}, w_M^{(u)} = \frac{1}{n}$ .
2. Repeat:
3.   Repeat:
4.     Update  $SD$  by solving problem (Eq 17).
5.     Update  $SM$  by solving problem (Eq 19).
6.     Update  $F$  with (Eq 21).
7.   Until convergence
8. Update  $w_D^{(v)}, w_M^{(u)}$  according to Eqs (12) and (13)
9. Until convergence
10. Return  $SD, SM$  and  $F$

**Theoretical convergence analysis.** We prove the convergence of **Algorithm 1** by the following theorem.

**Theorem 1.** The iterative optimization process in **Algorithm 1** can monotonically decrease the objective function value of (Eq 11) until convergence.

*Proof.* By fixing  $SM$  and  $F$ , the optimization for  $SD$  in (Eq 17) is a quadratic programming problem[35]. Specifically, the Hessian matrix of the Lagrange function of (Eq 17) is positive definite, i.e.  $2I \in \mathbb{R}^{q \times q}$ . Therefore, we arrive at:

$$\Omega(SM, SD^t, F) > \Omega(SM, SD^{t+1}, F) \tag{22}$$

Similarly, we could prove that the optimization for  $SM$  by fixing the other variables is also convex and we could obtain that:

$$\Omega(SM^t, SD, F) > \Omega(SM^{t+1}, SD, F) \tag{23}$$

By fixing  $SD$  and  $SM$ , the optimization function for updating  $F$  is also convex[36]. Moreover, since  $\beta L_{SM} + I$  is positive definite and  $\beta L_{SD}$  is positive semi-definite, the eigenvalues  $\sigma_1, \sigma_2, \dots, \sigma_l$  of  $\beta L_{SM} + I$  and  $\xi_1, \xi_2, \dots, \xi_k$  of  $\beta L_{SD}$  satisfy the inequality  $\sigma_i + \xi_j > 0$  ( $i = 1, \dots, l; j = 1, \dots, k$ ), which guarantees that there is a unique solution to (Eq 21). As a result, we could obtain:

$$\Omega(SM, SD, F^t) \geq \Omega(SM, SD, F^{t+1}) \tag{24}$$

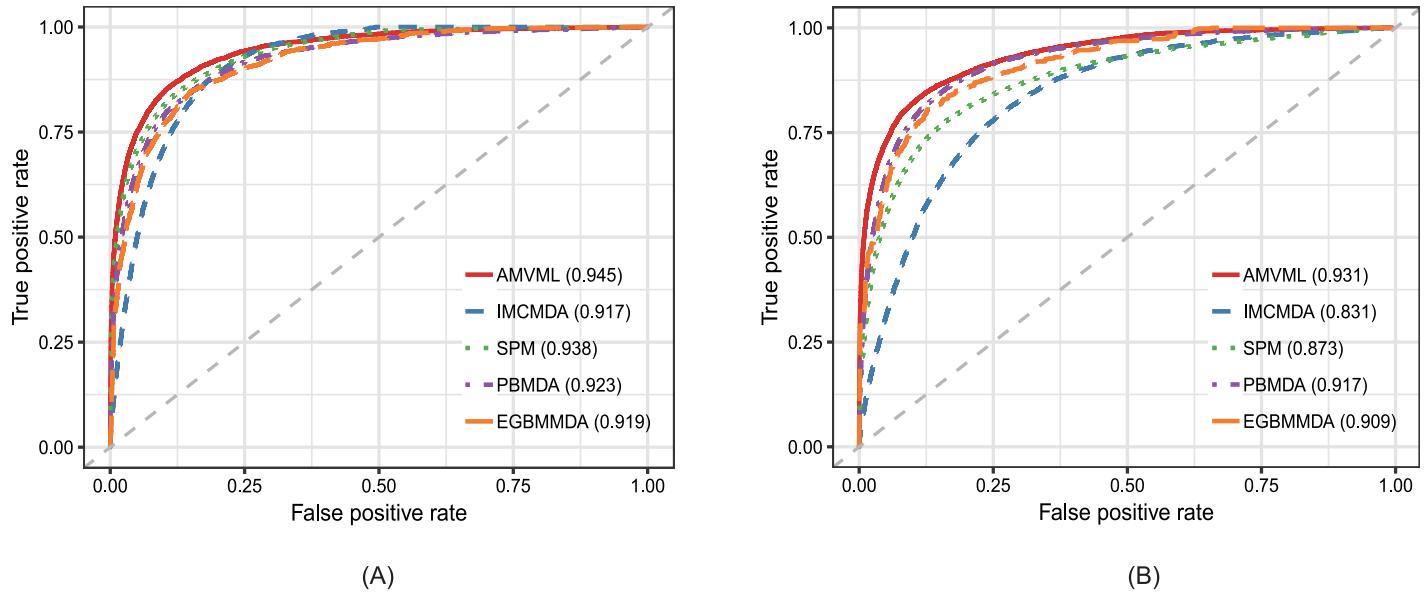
As demonstrated in the above analysis, **Algorithm 1** can monotonically decrease the objective function value of (Eq 11) in each iteration until it converges.

## Results

### Performance evaluation

To systematically evaluate the performance of our method and illustrate its superiority over existing alternatives, we compared AMVML with fourstate-of-the-art methods, i.e. IMCMDA [37], SPMMDA[38], PBMMDA[12] and EGBMMDA[19] under several evaluation metrics. All these methods have been proved effective in predicting reliable disease-associated miRNAs. First of all, we adopted the global Leave-One-Out Cross-Validation(LOOCV) and five-fold cross-validation to test the general prediction performance. Specifically, in the framework of global LOOCV, each known miRNA-disease association was selected as a test sample while the remaining associations were considered as training samples. For five-fold cross-validation, all known miRNA-disease associations were randomly divided into five subsets and each subset was chosen as the test samples. Besides, the five-fold cross-validation was repeated 10 times to eliminate the potential bias caused by the sample division. The prediction performance was illustrated by Receiver Operating Characteristic(ROC) curve and the accuracy was quantified by the Area Under the ROC Curve(AUC). As shown in Fig 3, AMVML achieved the highest accuracy among all methods in both global LOOCV and five-fold cross-validation.

Next, we employed another evaluation metric called Leave-One-Disease-Out Cross-Validation(LODOCV) to verify the prediction performance when no prior information is available. Specifically, for each disease  $d$ , we removed all known miRNAs associated with  $d$  and carried out predictions based on miRNA association information of the other diseases. Since there are no known associations for each tested disease in advance, LODOCV is more difficult than global LOOCV and five-fold cross-validation. We calculated an AUC value for each disease in LODOCV and thus obtained a vector consisting of 328 AUC values for each method. We then demonstrated the comparison results by density plots (Fig 4A). As a result, the AUC values obtained by our method mainly concentrated over the interval [0.9, 1], indicating a better

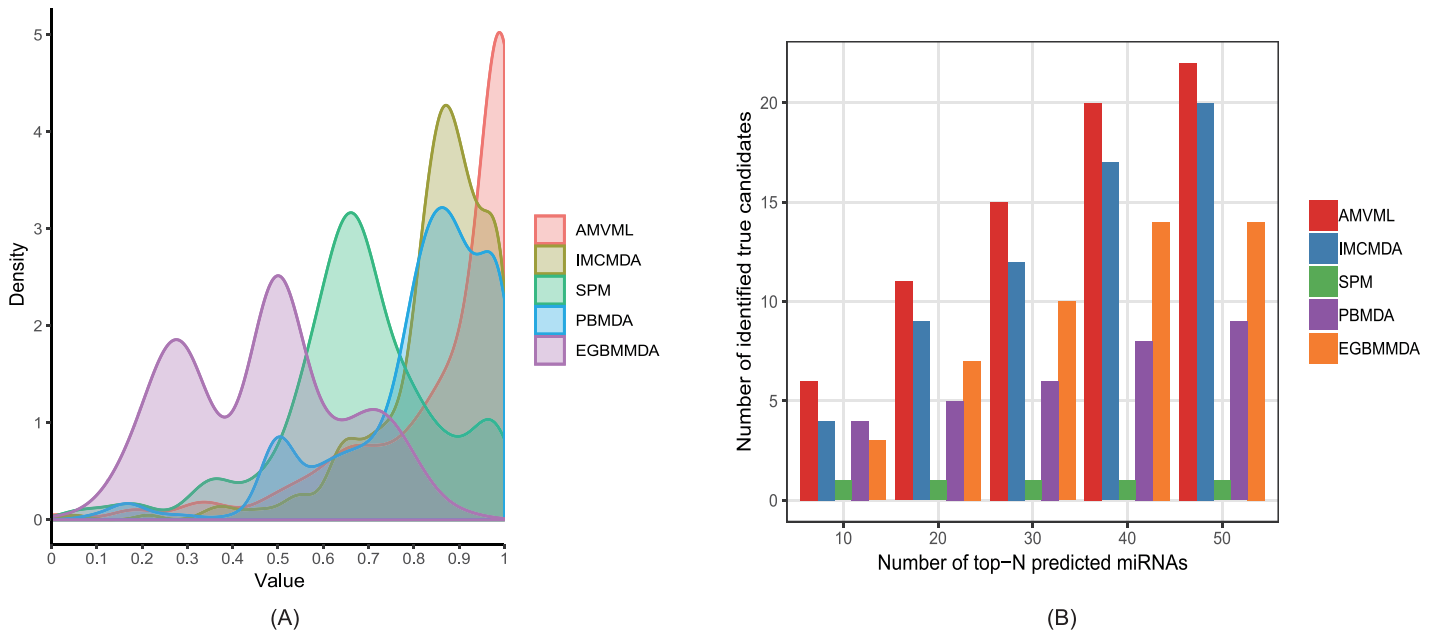


**Fig 3.** The comparison results between our method and the other four methods in terms of (A) global LOOCV; (B) five-fold cross-validation.

<https://doi.org/10.1371/journal.pcbi.1006931.g003>

performance than that of the other methods in terms of LODOCV. Wilcoxon signed-rank test further confirmed the statistical significance of the comparison results (Table 1).

Lastly, we conducted experiments on real datasets to further demonstrate the prediction ability of our method. To this end, we first downloaded the older version of HMDD (v1.0) which contains 1474 known associations between 129 diseases and 280 miRNAs after filtering (S7 File). Compared to HMDD v1.0, there were 4614 (i.e. 6088–1474) new miRNA-disease



**Fig 4.** (A) The comparison results between our method and the other four methods in terms of LODOCV. As shown in the density plot, AMVML lies in the rightmost position of the figure, indicating a superior performance over the other methods; (B) The number of identified true positives in the latest HMDD by all methods.

<https://doi.org/10.1371/journal.pcbi.1006931.g004>

**Table 1. Statistical significance of differences in performance between AMVML and the other four methods in terms of LOOCV.** *P-values* were calculated by Wilcoxon signed rank test.

	IMCMDA	SPM	PBMMA	EGBMMA
<i>p-value</i>	2.67e-10	1.36e-41	2.86e-13	2.18e-21

<https://doi.org/10.1371/journal.pcbi.1006931.t001>

associations, 199 (i.e. 328–129) new diseases and 270 (i.e. 550–280) new miRNAs involved in HMDD v2.0. In particular, among the 4614 newly recorded associations in HMDD v2.0, 2445 associations were related with miRNAs and diseases already existed in HMDD v1.0, while 2169 associations were related with either new miRNAs or new diseases only contained in HMDD v2.0. Moreover, the degree distribution of miRNAs as well as that of diseases for the 4614 associations indicating that only a minority of these associations were related with highly connected miRNAs and diseases (S1 Fig). We then applied each method on HMDD v1.0 and validated the prediction results by the 4614 associations newly added in HMDD v2.0. Therefore, for each method, the greater the number of true positives predicted, the better the performance. Specifically, we compared the number of true positives in the top-*N* miRNAs predicted by each method with *N* ranging from 10 to 50 and an interval of 10. As exhibited in Fig 4B, AMVML obtained greater number of validated disease-associated miRNAs than the other methods. Similar results were also obtained with increased *N* and larger intervals (S2 Fig). Taken together, the experimental results under various evaluation metrics proved the effectiveness of our method.

### Parameter analysis

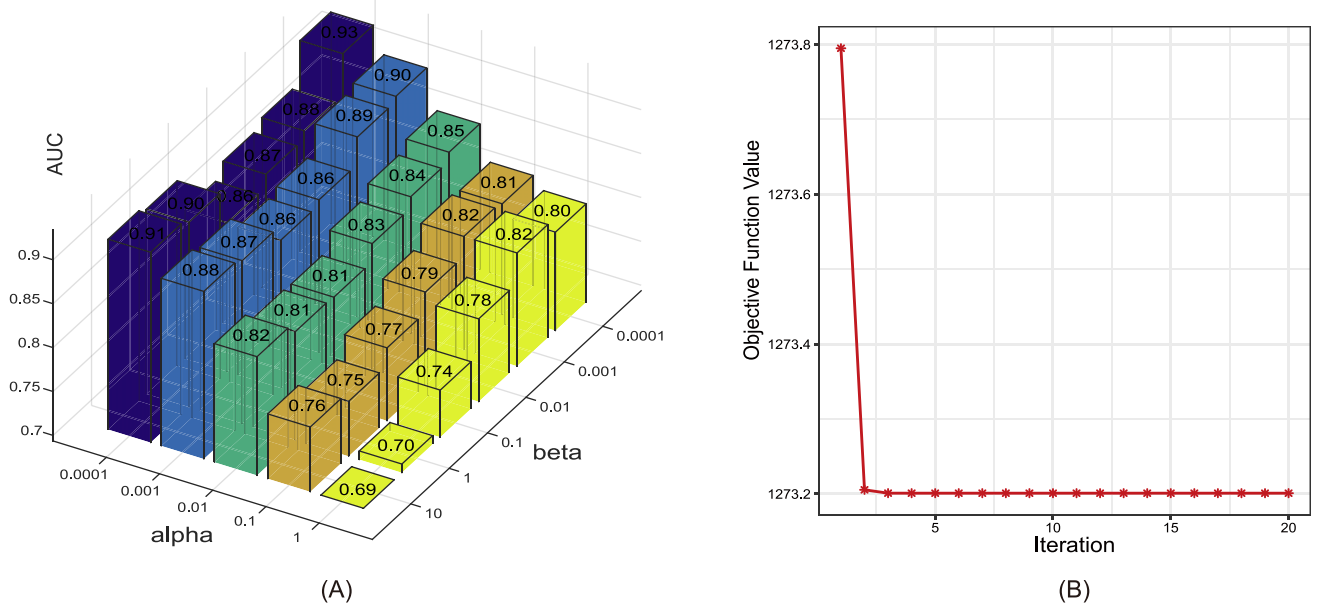
There were two trade-off parameters  $\alpha$  and  $\beta$  in our method which balance the learned similarity matrices and the predicted association matrix. Generally, since our objective function is a minimization problem, setting a large value to  $\alpha$  or  $\beta$  indicates a large impact of the label consistency between diseases or miRNAs on the learned disease or miRNA similarity matrix. To show a reasonable searching range of these two parameters as well as a general trend of the prediction performance affected by varying their values, in this subsection, we analyzed their influences on the prediction accuracy in terms of five-fold cross-validation (Fig 5A). Similar trends were also observed in global LOOCV. In particular, when  $\beta$  was fixed, the smaller the  $\alpha$ , the better the performance. In contrast, when  $\alpha$  was fixed, the performance varied in a "U" shape with the change of  $\beta$ . We can see that the proposed method reached the best performance when both  $\alpha$  and  $\beta$  were equal to  $1e-4$ .

### Convergence speed in practice

As described in previous section, we have theoretically proved the convergence of our algorithm. Here we investigated the convergence rate of our method by analyzing the variations of the objective function value in (Eq 11) with respect to the number of iterations. As demonstrated in Fig 5B, the objective function value reached a steady state within 5 iterations, indicating a fast convergence rate of our method.

### Case study

In this section, we conducted a case study on thyroid neoplasms to identify potential miRNA biomarkers for this disease. The overall prediction results and the differential expression analysis for several other diseases were also provided on Github (<https://github.com/alcs417/AMVML>). Thyroid cancer is the most common endocrine cancer and its incidence rate has increased rapidly over recent years[39]. We first downloaded the miRNA expression profiles



**Fig 5.** (A) The influence of the two parameters alpha and beta on the prediction accuracy of five-fold cross-validation; (B) The convergence rate of AMVML.

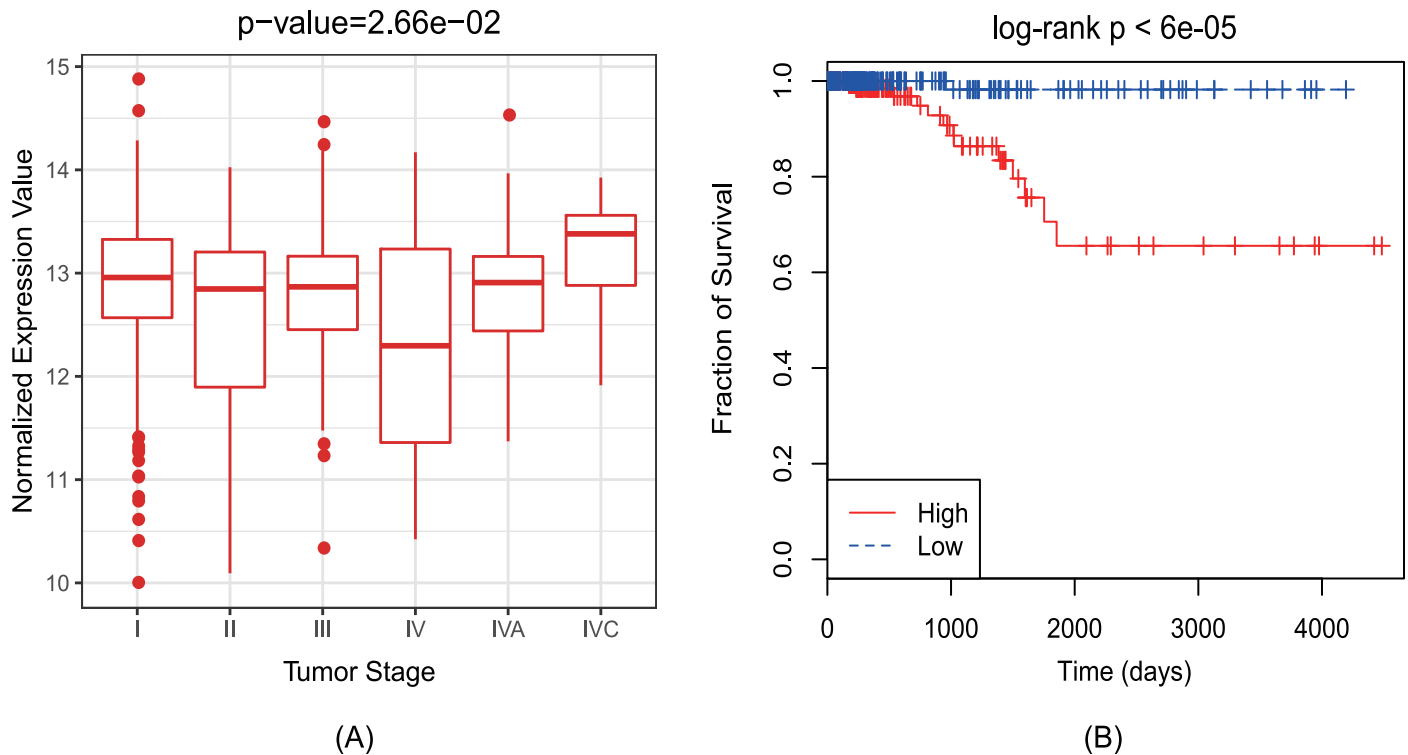
<https://doi.org/10.1371/journal.pcbi.1006931.g005>

together with the clinical information of thyroid carcinoma from GDC data portal (<https://portal.gdc.cancer.gov/projects/TCGA-THCA>). Concretely, the downloaded data contained 506 tumors samples and 59 normal samples and each sample measured the expression level of 1881 miRNAs. We then applied our method on the given disease to obtain the top-10 predicted miRNAs (Table 2). Specifically, we evaluated the classification power of these miRNAs in differentiating tumor samples from normal samples according to their expression profile and the results of five-fold cross-validation illustrated that they could achieve a mean classification accuracy of 0.983 (S3 Fig). Next, we calculated for each miRNA the fold-change as well as the statistical significance of differential expression using the R package edgeR (Table 2) [40]. Besides, we searched in another two databases dbDEMC and miR2Disease to see if the predicted miRNAs were also recorded in them [41, 42]. dbDEMC is an integrated database that designed to store and display differentially expressed miRNAs in human cancers detected by high-throughput methods while miR2Disease is a manually curated database providing

**Table 2.** Top-10 miRNAs predicted to be associated with thyroid neoplasms by AMVML.

Ranking	miRNA	logFC	FDR	Evidence
1	hsa-mir-181a-2	1.47	1.55e-41	dbDEMC; miR2Disease
2	hsa-mir-205	2.01	2.76e-06	dbDEMC
3	hsa-mir-143	-0.26	1.51e-01	dbDEMC
4	hsa-mir-145	-0.4	1.52e-03	dbDEMC
5	hsa-mir-101-2	-0.044	1.00	unconfirmed
6	hsa-mir-34a	2.23	1.86e-59	dbDEMC; miR2Disease
7	hsa-mir-133a-1	-0.62	2.37e-03	dbDEMC
8	hsa-mir-99a	-0.42	2.24e-06	dbDEMC
9	hsa-mir-218-1	-0.61	2.43e-08	dbDEMC
10	hsa-mir-34c	-0.14	7.20e-01	unconfirmed

<https://doi.org/10.1371/journal.pcbi.1006931.t002>



**Fig 6.** (A) The expression profiles of hsa-mir-181a-2 at different tumor stages. P-value was calculated by one-way ANOVA test; (B) Kaplan-Meier survival analysis using hsa-mir-181a-2 in thyroid cancer. The patients were divided into two groups 'High' and 'Low' according to their expression level of hsa-mir-181a-2 against the sample mean. As observed, patients with lower expression levels were at lower risk level.

<https://doi.org/10.1371/journal.pcbi.1006931.g006>

information about miRNA deregulation in various human diseases. As a result, the expression level of the top predicted miRNA hsa-mir-181a-2 was significantly altered between tumor samples and normal samples ( $\log_2$  fold-change > 1 and adjusted  $p$ -value < 0.05), which is consistent with the records in both db2DEMC and miR2Disease. Therefore, we further checked whether this miRNA could serve as a potential biomarker for thyroid cancer. Specifically, we carried out one-way ANOVA test to validate whether its expression level at different tumor stages also significantly altered. The tumor stages of all patients were retrieved from the clinical information and there were six pathologic stages after filtering. As expected, the expression level of hsa-mir-181a-2 varied significantly among different stages (Fig 6A). Furthermore, the Kaplan-Meier survival analysis confirmed that the survival rates of patients were also significantly related with its expression level (Fig 6B)[43]. Taken together, our results provided new evidence for the functional role of hsa-mir-181a-2 in the development of thyroid cancer.

## Discussion

Identification of disease-associated miRNAs has drawn much attention during the past decade and it still remains a challenging task. In this study, we proposed a novel computational framework to effectively uncover the potential links between miRNAs and diseases. Our method integrated datasets from multiple sources and adaptively learned two new similarity graphs. Specifically, instead of assigning predetermined weight values to each input similarity matrix, the proposed method automatically updated the view weights according to the reliability of each view. It is also worth mentioning that our method could be easily extended if there are new data sources available. Besides, our method could simultaneously update the prediction

results from both disease space and miRNA space. The convergence of our method has been proved both theoretically and experimentally. To demonstrate the utility of our method, we compared AMVML with five state-of-the-art methods and the experimental results confirmed the superiority of our method. We then applied our method on thyroid cancer and found that hsa-mir-181a-2 could be a potential prognostic biomarker. Notably, our method is not limited to discover miRNAs for which an association is already known between its paralogous miRNA and the same disease. In essence, as a semi-supervised learning model, our method could fully take advantage of the limited number of known miRNA-disease associations together with multiple sources of biological datasets to reliably predict novel associations. Therefore, we anticipate that our method could serve as an effective tool for miRNA-disease association prediction.

The superior performance of our model can be largely attributed to the following two reasons. First, the consensus similarity matrices obtained from multiple biological datasets based on multi-view learning for both miRNAs and diseases are more robust to outliers and noises compared to existing methods. Second, the graph-based multi-label learning unified the two prediction spaces into one optimization framework, which enhances the inherent correlations between miRNAs and diseases from the label-consistency point of view. Nevertheless, our method still has some limitations. Specifically, there are two parameters  $\alpha$  and  $\beta$  in the objective function that need to be tuned in advance, and it is a non-trivial task to find the best combination of the two parameters. In addition, although our method can adaptively learn a new affinity graph from different data sources, the integration of unreliable similarity matrices might weaken the overall prediction accuracy.

## Supporting information

**S1 Fig. The degree distribution of (A): Diseases and (B): miRNAs.**

(EPS)

**S2 Fig. The number of identified true positives in HMDD v2.0 by all methods with N ranging from (A) 200 to 1000; and (B) 2000 to 10000.**

(EPS)

**S3 Fig. Five-fold cross-validation results obtained by taking the top 10 predicted miRNAs as features.**

(EPS)

**S1 File. The 6088 miRNA-disease associations obtained from HMDD v2.0.**

(XLSX)

**S2 File. The number of diseases in each disease type contained in HMDD v2.0.**

(XLSX)

**S3 File. Disease semantic similarity matrix.**

(XLSX)

**S4 File. MiRNA sequence similarity matrix.**

(XLSX)

**S5 File. MiRNA functional similarity matrix.**

(XLSX)

**S6 File. MiRNA semantic similarity matrix.**

(XLSX)



**S7 File. The 1474 miRNA-disease associations obtained from HMDD v1.0.**  
(XLSX)

## Author Contributions

**Conceptualization:** Cheng Liang, Jiawei Luo.

**Data curation:** Cheng Liang.

**Formal analysis:** Cheng Liang, Shengpeng Yu.

**Funding acquisition:** Cheng Liang, Jiawei Luo.

**Investigation:** Cheng Liang.

**Methodology:** Cheng Liang.

**Project administration:** Cheng Liang.

**Supervision:** Jiawei Luo.

**Validation:** Shengpeng Yu.

**Visualization:** Cheng Liang, Shengpeng Yu.

**Writing – original draft:** Cheng Liang.

**Writing – review & editing:** Cheng Liang, Jiawei Luo.

## References

1. Lin SB, Gregory RI. MicroRNA biogenesis pathways in cancer. *Nat Rev Cancer*. 2015; 15(6):321–33. <https://doi.org/10.1038/nrc3932> PMID: 25998712
2. Zeng X, Zhang X, Zou Q. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Brief Bioinform*. 2016; 17(2):193–203. <https://doi.org/10.1093/bib/bbv033> PMID: 26059461
3. Jiang Q, Hao Y, Wang G, Juan L, Zhang T, Teng M, et al. Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst Biol*. 2010; 4 Suppl 1:S2.
4. Chen X, Liu MX, Yan GY. RWRMDA: predicting novel human microRNA-disease associations. *Mol Biosyst*. 2012; 8(10):2792–8. <https://doi.org/10.1039/c2mb25180a> PMID: 22875290
5. Shi H, Xu J, Zhang G, Xu L, Li C, Wang L, et al. Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes. *BMC Syst Biol*. 2013; 7:101. <https://doi.org/10.1186/1752-0509-7-101> PMID: 24103777
6. Chen X, Yan CC, Zhang X, You ZH, Huang YA, Yan GY. HGIMDA: Heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget*. 2016; 7(40):65257–69. <https://doi.org/10.18632/oncotarget.11251> PMID: 27533456
7. Chen X, Yin J, Qu J, Huang L. MDHGI: Matrix Decomposition and Heterogeneous Graph Inference for miRNA-disease association prediction. *PLoS Comput Biol*. 2018; 14(8):e1006418. <https://doi.org/10.1371/journal.pcbi.1006418> PMID: 30142158
8. Liu YS, Zeng XX, He ZY, Zou Q. Inferring MicroRNA-Disease Associations by Random Walk on a Heterogeneous Network with Multiple Data Sources. *IEEE-ACM Trans Comput Biol Bioinform*. 2017; 14(4):905–15.
9. Xuan P, Han K, Guo Y, Li J, Li X, Zhong Y, et al. Prediction of potential disease-associated microRNAs based on random walk. *Bioinformatics*. 2015; 31(11):1805–15. <https://doi.org/10.1093/bioinformatics/btv039> PMID: 25618864
10. Zou Q, Li J, Hong Q, Lin Z, Wu Y, Shi H, et al. Prediction of MicroRNA-Disease Associations Based on Social Network Analysis Methods. *Biomed Res Int*. 2015; 2015:810514. <https://doi.org/10.1155/2015/810514> PMID: 26273645
11. Sun D, Li A, Feng H, Wang M. NTSMDA: prediction of miRNA-disease associations by integrating network topological similarity. *Mol Biosyst*. 2016; 12(7):2224–32. <https://doi.org/10.1039/c6mb00049e> PMID: 27153230

12. You ZH, Huang ZA, Zhu Z, Yan GY, Li ZW, Wen Z, et al. PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. *PLoS Comput Biol*. 2017; 13(3): e1005455. <https://doi.org/10.1371/journal.pcbi.1005455> PMID: 28339468
13. Li GH, Luo JW, Xiao Q, Liang C, Ding PJ, Cao BW. Predicting MicroRNA-Disease Associations Using Network Topological Similarity Based on DeepWalk. *Ieee Access*. 2017; 5:24032–9.
14. Chen X, Guan NN, Li JQ, Yan GY. GIMDA: Graphlet interaction-based MiRNA-disease association prediction. *J Cell Mol Med*. 2018; 22(3):1548–61. <https://doi.org/10.1111/jcmm.13429> PMID: 29272076
15. Xiao Q, Luo JW, Liang C, Cai J, Ding PJ. A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics*. 2018; 34(2):239–48.
16. Chen X, Wang L, Qu J, Guan NN, Li JQ. Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics*. 2018; 34(24):4256–65. <https://doi.org/10.1093/bioinformatics/bty503> PMID: 29939227
17. Yu SP, Liang C, Xiao Q, Li GH, Ding PJ, Luo JW. MCLPMDA: A novel method for miRNA-disease association prediction based on matrix completion and label propagation. *J Cell Mol Med*. 2018.
18. Zeng X, Liu L, Lu L, Zou Q. Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics*. 2018; 34(14):2425–32. <https://doi.org/10.1093/bioinformatics/bty112> PMID: 29490018
19. Chen X, Huang L, Xie D, Zhao Q. EGBMMDA: Extreme Gradient Boosting Machine for MiRNA-Disease Association prediction. *Cell Death Dis*. 2018; 9(1):3. <https://doi.org/10.1038/s41419-017-0003-x> PMID: 29305594
20. Yu SP, Liang C, Xiao Q, Li GH, Ding PJ, Luo JW. GLNMDA: a novel method for miRNA-disease association prediction based on global linear neighborhoods. *RNA Biol*. 2018; 15(9):1215–27. <https://doi.org/10.1080/15476286.2018.1521210> PMID: 30244645
21. Qu Y, Zhang HX, Lyu C, Liang C. LLCMDA: A Novel Method for Predicting miRNA Gene and Disease Relationship Based on Locality-Constrained Linear Coding. *Frontiers in genetics*. 2018; 9.
22. Chen X, Wang CC, Yin J, You ZH. Novel Human miRNA-Disease Association Inference Based on Random Forest. *Mol Ther Nucleic Acids*. 2018; 13:568–79. <https://doi.org/10.1016/j.omtn.2018.10.005> PMID: 30439645
23. Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, et al. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res*. 2014; 42(Database issue):D1070–4. <https://doi.org/10.1093/nar/gkt1023> PMID: 24194601
24. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. 2014; 42(Database issue):D68–73. <https://doi.org/10.1093/nar/gkt1181> PMID: 24275495
25. Wang D, Wang J, Lu M, Song F, Cui Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics*. 2010; 26(13):1644–50. <https://doi.org/10.1093/bioinformatics/btq241> PMID: 20439255
26. Zou Q, Li J, Song L, Zeng X, Wang G. Similarity computation strategies in the microRNA-disease network: a survey. *Briefings in functional genomics*. 2016; 15(1):55–64. <https://doi.org/10.1093/bfpg/elt024> PMID: 26134276
27. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res*. 2018; 46(D1): D296–D302. <https://doi.org/10.1093/nar/gkx1067> PMID: 29126174
28. Yu GC, Li F, Qin YD, Bo XC, Wu YB, Wang SQ. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*. 2010; 26(7):976–8. <https://doi.org/10.1093/bioinformatics/btq064> PMID: 20179076
29. Wang JZ, Du Z, Payattakool R, Yu PS, Chen CF. A new method to measure the semantic similarity of GO terms. *Bioinformatics*. 2007; 23(10):1274–81. <https://doi.org/10.1093/bioinformatics/btm087> PMID: 17344234
30. Chen G, Song Y, Wang F, Zhang C, editors. Semi-supervised Multi-label Learning by Solving a Sylvester Equation. *Siam International Conference on Data Mining, SDM 2008*, April 24–26, 2008, Atlanta, Georgia, Usa; 2008.
31. Zha ZJ, Mei T, Wang J, Wang Z, Hua XS. Graph-based semi-supervised learning with multiple labels. *J Vis Commun Image Represent*. 2009; 20(2):97–103.
32. Nie FP, Cai GH, Li J, Li XL. Auto-Weighted Multi-View Learning for Image Clustering and Semi-Supervised Classification. *Ieee Transactions on Image Processing*. 2018; 27(3):1501–11. <https://doi.org/10.1109/TIP.2017.2754939> PMID: 28945592
33. Nie F, Li J, Li X. Self-weighted Multiview Clustering with Multiple Graphs. *The Twenty-Sixth International Joint Conference on Artificial Intelligence*; Melbourne, Australia 2017. p. 2564–70.

34. Huang J, Nie FP, Huang H, editors. A new simplex sparse learning model to measure data similarity for clustering. The twenty-fourth International Conference on Artificial Intelligence; 2015; Buenos Aires, Argentina.
35. Zhu L, Huang Z, Li ZH, Xie L, Shen HT. Exploring Auxiliary Context: Discrete Semantic Transfer Hashing for Scalable Image Retrieval. *Ieee Transactions on Neural Networks and Learning Systems*. 2018; 29(11):5264–76. <https://doi.org/10.1109/TNNLS.2018.2797248> PMID: 29994644
36. Zhang HX, Cao LL, Gao S. A locality correlation preserving support vector machine. *Pattern Recogn*. 2014; 47(9):3168–78.
37. Chen X, Wang L, Qu J, Guan NN, Li JQ. Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics*. 2018.
38. Zeng XX, Liu L, Lu LY, Zou Q. Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics*. 2018; 34(14):2425–32. <https://doi.org/10.1093/bioinformatics/bty112> PMID: 29490018
39. Wang TS, Sosa JA. Thyroid surgery for differentiated thyroid cancer—recent advances and future directions. *Nature Reviews Endocrinology*. 2018; 14(11):670–83. <https://doi.org/10.1038/s41574-018-0080-7>
40. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616> PMID: 19910308
41. Yang Z, Wu L, Wang A, Tang W, Zhao Y, Zhao H, et al. dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res*. 2017; 45(D1):D812–D8. <https://doi.org/10.1093/nar/gkw1079> PMID: 27899556
42. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res*. 2009; 37(Database issue):D98–104. <https://doi.org/10.1093/nar/gkn714> PMID: 18927107
43. Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res*. 2010; 1(4):274–8. <https://doi.org/10.4103/0974-7788.76794> PMID: 21455458