Predicting small molecule-miRNA associations using matrix enhancement and collaborative double matrix completion

Tiyao Liu[a], Shudong Wang[a,*], Yuanyuan Zhang[b], Chuanru Ren[c], Yunyin Li[a], Xiaodong Tan[a] and Shanchen Pang[a]

[a] *College of Computer Science and Technology, Qingdao Institute of Software, China University of Petroleum, Qingdao 266580, China*
[b] *School of Information and Control Engineering, Qingdao University of Technology, Qingdao 266525, China*
[c] *College of Computer Science and Technology, Tongji University, Shanghai 201804, China*

ARTICLE INFO

Keywords:
MMAs
matrix enhancement
truncated schatten p-norm
truncated matrix factorization
matrix completion
association prediction

ABSTRACT

An increasing body of research underscores the significant role of noncoding RNAs as viable targets for numerous existing small molecule (SM) drugs. Matrix completion is a feasible approach, however existing methods have limitations such as high computational complexity and not optimal solutions. In addition, similarity matrices may be noisy, and their topological information not fully exploited. In this work, we designed a matrix enhancement and collaborative double matrix completion framework (MECDMC) for small molecule-miRNA association prediction. We use Gaussian Radial Basis Function (GRBF) to perform matrix enhancement of the integrated SM/miRNA similarity to obtain the final refined SM and miRNA similarity. This step takes into account the structural information of the integrated similarity matrix, which improves the accuracy of the similarity measure. Second, we develop a Collaborative Double Matrix Completion (CDMC) framework that skillfully combines two complementary matrix completion techniques. We first use truncated schatten p-paradigm minimization to complement the association matrix, which solves the problem of overly sparse association matrix. We then use truncated matrix factorization on the updated correlation matrix to obtain the final prediction results. Notably, the truncated matrix factorization greatly improves the speed of the model by truncating the number of singular values. The experimental outcomes demonstrate that MECDMC obtains an AUC of 0.9976±0.0002 (0.9156±0.0016) and 0.9981 (0.9237) under 5-fold Cross-Validation (CV) and global LOOCV using dataset 1 (dataset 2), respectively, which surpasses the other six advanced approaches. Notably, our application of MECDMC to miRNA-disease association prediction yielded excellent performance results in this specific area.

1. Introduction

MicroRNAs (miRNAs) are evolutionarily conserved non-coding RNA molecules present in various organisms, typically spanning 21 to 25 nucleotides. They are critical in regulating a variety of biological processes [1, 2]. The expression pattern of miRNAs is histospecific, and any aberrant expression tends to impact the cellular state [3]. Extensive clinical and experimental evidence has revealed the involvement of miRNAs in a wide range of complex human diseases, including cardiovascular conditions, metabolic diseases and inflammatory disorders [4, 5]. For instance, the hsa-miR-125a-3p expression exhibited a notable decrease in breast cancer cells [6]. In squamous cell lung cancer tissues exhibit systematic changes in the expression pattern of miRNAs, and an elevated level of miR-21 is linked to a reduced survival time [7]. Consequently, miRNAs hold potential as valuable clinical and prognostic biomarkers. An expanding body of research indicates that small molecule (SM) drugs can effectively intervene in specific miRNA expression and function, thereby achieving therapeutic effects [8]. Recently, a growing number of SM-miRNA associations (MMAs) have been identified. For example, miR-155 expression, upregulated in various cancers, including lung, colorectal, and breast cancers, can be downregulated by curcumin. This downregulation inhibits the protrusion and invasion of tumor cells [9]. The ability of SM to precisely target disease-specific miRNA pathways makes it an important tool for individualized medicine and targeted therapies.

Various SMs exhibit distinct mechanisms of action and efficacy by targeting different miRNAs. Therefore, finding specific associations between SMs and miRNAs is necessary. However, conventional biological experiments face challenges such as time, cost, and technological limitations, hindering their application in large-scale studies and comprehensive analyses. Accordingly, there is an imperative to develop computational models for predicting MMAs. In recent years, numerous computational models have been proposed to predict MMAs. These approaches fall into three primary categories: network inference-based approaches, machine learning-based approaches, and matrix completion-based approaches.

Network inference-based approaches leverage biological information to formulate heterogeneous networks and employ inference algorithms for predictions. Qu et al. [10] introduced the TLHNSMMA model, integrating SM, miRNA, and disease information to build a three-layer heterogeneous network. Unknown MMAs are then inferred using this network information. Yin et al. proposed the SLHGIS-MMA model [11], integrating SM/miRNA similarity and MMA information in a heterogeneous map. They utilized a sparse learning method for noise reduction in the MMA

[*] Corresponding author
20140017@upc.edu.cn (S. Wang)
ORCID(s):

matrix, thereby improving final prediction accuracy. GISMMA [12] utilized interactions among 28 isomers to the correlation strength between SMs and between miRNAs. The SM/miRNA similarity network was calculated to derive the association score. The SMiR-NBI was presented by Qu et al. [13], involving the construction of a heterogeneous network using drugs, miRNAs, and genetic information. Predictive scores were obtained by reasoning based on the information within the heterogeneous network. However, a potential drawback may be an over-reliance on web-based information, leading to inflexibility

Machine learning-based approaches predominantly concentrate on hidden feature extraction and classifiers [14]. DAESTB [15] created a multidimensional feature matrix by incorporating SM and miRNA-related information. Following this, a deep autoencoder was utilized to perform denoising and dimensionality reduction. Subsequently, we employed a scalable tree augmentation model to derive the prediction results. EKRRSMMA [16] curated a subset of SM and miRNA features, applying feature downscaling techniques to mitigate the impact of noisy data during the construction process. In the subsequent training phase, they leveraged an ensemble learning method, markedly augmenting the achievement of accurate prediction results. CLD-ISMMA [17] constructed a complex network consisting of SMs, diseases and miRNAs. Subsequently, the regularization model was used to infer unknown MMAs. RFSMMA [18] utilized the SM/miRNA similarity as features to denote MMAs. Subsequently, training was performed using the random forest technique to obtain the predicted scores. Despite their efficacy, these methods may struggle to capture the potential correlation of sparse matrices and lack interpretability.

Matrix-completion based approach to build and optimize the objective function for calculating the MMA scores. We can further divide matrix completion methods into two categories based on the method principle: rank approximation norms minimization methods and matrix factorization methods.

1) Rank approximation norms minimization: Chen et al. [19] introduced nuclear norm minimization method (BNNRSMMA). Initially, they built a heterogeneous network and defined a matrix to denote it. Subsequently, a prediction model was devised, employing nuclear norm minimization to fill in missing values within the matrix. Wang et al. [20] presented the AMCSMMA to predict MMAs. Their approach involved the integration of bioinformatics to create a heterogeneous network. The neighbor matrix of this network was processed as the goal matrix, and missing values were complemented via the truncated nuclear norm minimization. Wang et al. [21] presented a TSPN approach. The bioinformatic matrix was integrated into a heterogeneous network, treating neighboring matrices as targets. By minimizing the truncated schatten p-norm, they reconstructed missing values in the target matrix. An iterative algorithm was designed to solve the model, yielding correlation scores. Although these methods excel in feature selectivity and noise robustness, their computational approach, involving heterogeneous network adjacency matrices, introduces significant complexity.

2) Matrix factorization: Luo et al. [22] presented a SMANMF approach, using non-negative matrix factorization, to uncover MMAs. Zhao et al. [23] devised SNMF-SMMA, employing a symmetric non-negative matrix factorization model, solved through the Kronecker regularized least squares algorithm. Wang et al. [24] presented a DCMF approach. Initially, they utilized the WKNKN method for preprocessing. Subsequently, they constructed the objective function and iteratively updated the feature matrices for both SMs and miRNAs. Finally, the association score matrix was obtained by iteratively updating until convergence. While matrix decomposition proves highly computationally efficient and interpretive, it is sensitive to noise and may not perform optimally when dealing with sparse data.

Matrix completion is a feasible approach, however existing methods have limitations such as high computational complexity and not optimal solutions. In addition, similarity matrices may be noisy, and their topological information not fully exploited. Therefore, we designed a matrix enhancement and collaborative double matrix completion framework (MECDMC) for MMA prediction. We use Gaussian Radial Basis Function (GRBF) to perform matrix enhancement of the integrated SM/miRNA similarity to obtain the final refined SM and miRNA similarity. This step takes into account the structural information of the integrated similarity matrix, which improves the accuracy of the similarity measure. Second, we develop a Collaborative Double Matrix Completion (CDMC) framework that skillfully combines two complementary matrix completion techniques. We use truncated schatten p-norm minimization to complement the association matrix, which solves the problem of overly sparse association matrix. The truncated schatten p-norm considers the physical nature of the singular values, effectively complementing missing values better than other rank approximation norms. We then use truncated matrix decomposition on the updated correlation matrix to obtain the final prediction results. The truncated matrix factorization greatly improves the speed of the model by truncating the number of singular values. Specifically, the principle contributions in this work can be outlined as shown below:

(1) We use GRBF to perform matrix enhancement of the integrated SM/miRNA similarity to obtain the final refined SM and miRNA similarity. This step takes into account the structural information of the integrated similarity matrix, which improves the accuracy of the similarity measure.

(2) The truncated schatten p-norm considers the physical nature of the singular values, effectively complementing missing values better than other rank approximation norms. In addition, the truncated matrix factorization greatly improves the speed of the model by truncating the number of singular values.

(3) MECDMC skillfully combines two matrix completion methods, fully utilizing the feature selectivity and robustness of the former and the computationally efficient and

Table 1
Detailed information about each dataset

Datasets	SMs	miRNAs	Associations	Sparsity
dataset 1	831	541	664	1.477×10^{-3}
dataset 2	39	286	664	5.953×10^{-2}
new dataset	831	541	796	1.771×10^{-3}

interpretable nature of the latter, to exhibit comprehensive and flexible performance in MMA prediction.

(4) The experimental outcomes demonstrate that MECDMC obtains an AUC of 0.9976±0.0002 (0.9156±0.0016) and 0.9981 (0.9237) under 5-fold Cross-Validation (CV) and global LOOCV using dataset 1 (dataset 2), respectively, which surpasses the other six advanced approaches. Moreover, we created a novel dataset and employed 5-fold CV to evaluate the MECDMC's performance, utilizing multiple metrics for comprehensive evaluation. MECDMC still exerts a strong prediction ability. Notably, our application of MECDMC to miRNA-disease association prediction yielded excellent performance results in this specific area. In addition, a large number of miRNAs were predicted in both case studies demonstrating that MECDMC is a promising predictive model.

2. Methods

2.1. MMAs

This paper employs three datasets to assess the predictive performance of MECDMC. A total of 831 SMs were gathered from DrugBank [25], SM2miR v1.0 [26], and PubChem [27]. Additionally, 541 miRNAs were sourced from HMDD [28], miR2Disease [29], SM2miR v1.0 [26], and PhenomiR [30]. Consequently, dataset 1 comprises 831 SMs, 541 miRNAs, and 664 known MMAs. In dataset 2, SMs and miRNAs without known MMAs in dataset 1 were excluded, resulting in 39 SMs, 286 miRNAs, and 664 known MMAs. Both datasets share the same 664 known MMAs, sourced from SM2miR v1.0. It is unclear whether the 664 known MMAs that have been widely used in experiments are sensitive, so we constructed a new dataset,including 796 known MMAs, 831 SMs, and 541 miRNAs. Detailed information on the datasets employed in this research is succinctly presented in Table 1.

2.2. Integrated SM similarity

Inspired by the literature [31], we computed four types of SM similarity to derive the integrated SM similarity. These four similarities contain SM side-effect-based similarity [32], chemical-structure-based similarity [33], gene-functionally-consistent-based similarity [34], and indicated-phenotype-based similarity [32].

Four matrices, denoted as S_{SM}^{S}, S_{SM}^{C}, S_{SM}^{G}, and S_{SM}^{P}, each of size $ns \times ns$, were created to store the corresponding SM similarities. The elements $S_{SM}^{S}(i,j)$, $S_{SM}^{C}(i,j)$, $S_{SM}^{G}(i,j)$ and $S_{SM}^{P}(i,j)$ represent the different similarities between SMs s_i and s_j. To mitigate bias from individual similarities and achieve a balance among the four, we adopted a weighted combination strategy for integration. The resulting integrated SM similarity matrix, denoted as S_{sm}, is obtained through this weighted combination.

$$S_{sm} = \frac{\alpha_1 S_{sm}^{S} + \alpha_2 S_{sm}^{C} + \alpha_3 S_{sm}^{G} + \alpha_4 S_{sm}^{P}}{\sum_{i=1}^{4} \alpha_i} \tag{1}$$

where the coefficients α_1, α_2, α_3, and α_4 were uniformly set to 1, implying that each similarity was given equal weighting.

2.3. Integrated miRNA similarity

Inspired by the literature [31], we computed two types of miRNA similarity to derive the integrated miRNA similarity. These similarities include gene-functionally-consistent-based similarity [34], and disease-phenotypic-based similarity [32], respectively.

Subsequently, we utilized matrices S_m^G and S_m^D, both of size $nm \times nm$, to represent the two calculated miRNA similarities, respectively. The values in the coordinates (i,j) of each similarity matrix signify the similarity score between miRNAs m_i and m_j. To address potential bias in each similarity, we combined the two miRNA similarities. The integrated miRNA similarity matrix, indicated as S_m, is obtained through this integration process.

$$S_m = \frac{\beta_1 S_m^G + \beta_2 S_m^D}{\sum_{i=1}^{2} \beta_i} \tag{2}$$

where the coefficients β_1 and β_2 were uniformly set to 1.

2.4. MECDMC

In the study, we propose the MECDMC model, a matrix enhancement and collaborative double matrix completion framework for identifying MMAs. The flowchart of MECDMC is presented in Figure 1, which consists of the following three main steps: (1) We processed the integrated SM and miRNA similarities using GRBF to acquire the final refined SM/miRNA similarity. (2) The missing values of the MMA matrix were complemented by the truncated schatten p-norm minimization. (3) Finally, we used a truncated matrix factorization approach on the updated correlation matrix to acquire the prediction results.

2.4.1. Gaussian radial basis function

Existing computational methods often rely on widely used integrated SM/miRNA similarities, which may lead to errors during the collection and integration of multiple similarity metrics. Therefore, we use gaussian radial basis function to perform matrix augmentation of the integrated SM/miRNA similarity to obtain the final refined SM and miRNA similarity. To elaborate, the i-th row vector of the SM similarity, marked as S_{sm} encapsulates the similarity values between the i-th SM node and all SM nodes. This row vector effectively serves as the feature representation of

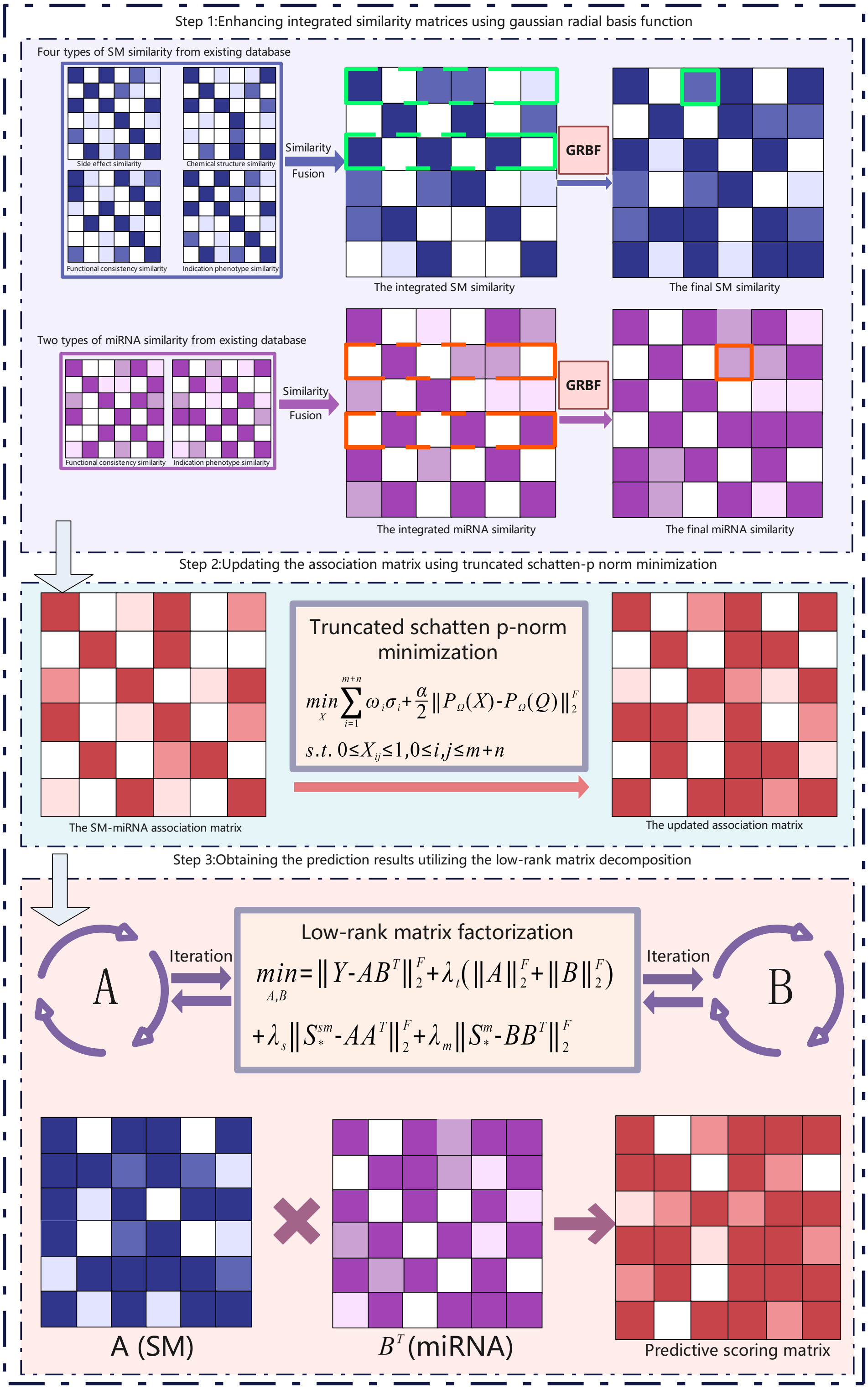

Step 1:Enhancing integrated similarity matrices using gaussian radial basis function
Four types of SM similarity from existing database
Side effect similarity
Chemical structure similarity
Functional consistency similarity
Indication phenotype similarity
Similarity Fusion
GRBF
The integrated SM similarity
The final SM similarity
Two types of miRNA similarity from existing database
Functional consistency similarity
Indication phenotype similarity
Similarity Fusion
GRBF
The integrated miRNA similarity
The final miRNA similarity
Step 2:Updating the association matrix using truncated schatten-p norm minimization
The SM-miRNA association matrix
Truncated schatten p-norm minimization
$\min_X \sum_{i=1}^{m+n} \omega_i \sigma_i + \frac{\alpha}{2} \|P_\Omega(X) - P_\Omega(Q)\|_2^F$
$s.t.\ 0 \le X_{ij} \le 1, 0 \le i,j \le m+n$
The updated association matrix
Step 3:Obtaining the prediction results utilizing the low-rank matrix decomposition
A
Iteration
Low-rank matrix factorization
$\min_{A,B} = \|Y - AB^T\|_2^F + \lambda_l (\|A\|_2^F + \|B\|_2^F)$
$+ \lambda_s \|S_*^{sm} - AA^T\|_2^F + \lambda_m \|S_*^m - BB^T\|_2^F$
Iteration
B
A (SM)
B^T(miRNA)
Predictive scoring matrix

Figure 1: The flowchart of MECDMC. Step1, Enhancing integrated similarity matrices using gaussian radial basis function; Step2, Updating the association matrix using truncated schatten-p norm minimization; Step3, Obtaining the prediction results utilizing the truncated matrix decomposition.

the i-th SM node. Mathematically, the SM Gaussian radial basis similarity can be formulated as follows:

$$S_{SM}(i,j) = exp(\frac{\left\|R_i - R_j\right\|}{-2\sigma^2}) \tag{3}$$

where R_i/R_j are the i-th/j-th row vectors of the matrix S_{sm}, respectively. The parameter σ is employed to regulate the function bandwidth. In this research, we consistently set the parameter σ to a value of 2. S_{SM} is the composite similarity matrix after GRBF processing. Similarly, by employing the miRNA similarity matrix S_m, we can calculate the miRNA Gaussian radial basis similarity, denoted by the corresponding similarity matrix S_M.

2.4.2. Truncated schatten p-norm minimization

Most of the computational methods still face a serious problem that the MMA matrix is too sparse. In this research, we used the truncated schatten p-norm minimization [21] to obtain a continuous value in the range of [0,1] to replace the missing data of the MMA matrix. Here, our objective matrix, denoted as $H \in R^{ns\times nm}$, and the matrix $X \in R^{ns\times nm}$ are to be recovered. Mathematically, it can be constructed as the following form:

$$\begin{aligned} &\min_X \operatorname{rank}(X) \\ &\text{s.t. } P_\Omega(X) = P_\Omega(H) \end{aligned} \tag{4}$$

where $rank(\cdot)$ is the rank function, $\Omega \subset \{1,\dots,ns\} \times \{1,\dots,nm\}$ is the set of positions corresponding to SM-miRNA pairs, and P_Ω is the orthogonal projection operator on Ω.

$$\left[P_\Omega(X)\right]_{\text{ij}} = \begin{cases} X_{ij} & \text{if (i, j)} \in \Omega \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Due to the discontinuity and nonconvexity of the rank function, there is currently no valid optimization algorithm available for its direct solution. Numerous studies suggest that the truncated schatten p-norm can serve as an optimal alternative to the rank function. We can transform equation (4) to a model utilizing the truncated schatten p-norm as shown below:

$$\begin{aligned} &\min_X \|X\|_r^p \\ &\text{s.t. } P_\Omega(X) = P_\Omega(H) \end{aligned} \tag{6}$$

Next, the important lemma of truncated schatten p-norm is introduced to facilitate the solution.

Lemma 1(See [35] and [36]) Consider a matrix $X \in R^{ns\times nm}$ with a rank $s(s \le min(ns, nm))$, and its singular value decomposition (SVD) as $X = U \triangle V^T$, where $U \in R^{ns\times ns}$, $\triangle \in R^{ns\times nm}$, $V \in R^{nm\times nm)}$. When $A \in R^{r\times ns}$, $B \in R^{r\times nm}$ and $0 < p \le 1$, the optimization problem has optimal solution. The specific formula is shown below:

$$\begin{aligned} \|X\|_{\text{r}}^p = \min_{A,B} \sum_{i=1}^{min(ns,nm)} \left(1 - \sigma_i\left(B^T A\right)\right)\left(\sigma_i(X)\right)^p \\ \text{s.t.} AA^T = I_{r\times r}, BB^T = I_{r\times r} \end{aligned} \tag{7}$$

Thanks to Lemma 1, we enhanced the initial model for minimizing the rank function (equation (4)) and developed a new model:

$$\begin{aligned} \min_X \sum_{i=1}^{min(ns,nm)} \left(1 - \sigma_i\left(B^\top A\right)\right)\left(\sigma_i(X)\right)^p \text{ s.t. } P_\Omega(X) = P_\Omega(H) \\ \text{s.t. } A \in R^{r\times ns},\ B \in R^{r\times nm}, \\ AA^\top = I_{r\times r},\ BB^\top = I_{r\times r}, \text{ and } \quad 0 < p \le 1 \end{aligned} \tag{8}$$

Equation (8) is nonconvex, rendering it more proficient in approximating rank function. However, its solution poses a challenge, as conventional methods are inadequate for addressing this non-convexity. For this reason, we first transformed the model (equation (8)). The detailed proof process is shown in Appendix I.

However, solving models with inequality constraints presents numerous challenges. Therefore, it is a widely adopted approach to replace the constrained module using a regularized counterpart. The incorporation of soft regularization not only allows for the accommodation of unforeseen noise but also significantly enhances the efficiency of our problem-solving procedures. Furthermore, we applied a constraint within the range of [0, 1] to all matrix values to ensure their practical significance [12, 37]. In conclusion, we constructed the following model:

$$\begin{aligned} \min_X \sum_{i=1}^{min(ns,nm)} \omega_i\sigma_i + \frac{\lambda}{2}\left\|P_\Omega(X) - P_\Omega(H)\right\|_F^2 \\ \text{s.t. } 0 \le X_{ij} \le 1 (0 \le i \le ns, 0 \le j \le nm) \end{aligned} \tag{9}$$

where λ is a equilibrium coefficient and $0 \le X_{i,j} \le 1$ (where $0 \le i \le ns, 0 \le j \le nm$) signifies that all elements in X are within the [0, 1] interval.

We formulated a framework utilizing the alternating direction multiplier approach (ADMM) [38] to handle the optimization problem as shown below.

Step 1: Initialize $X_1 = H$ and calculate the $(l+1)$-th iteration of $X_l = U_l \triangle_l V_l^T$. Next, proceed to derive A_l and B_l from U_l and V_l. Experimental validation shows that $l \in [1, 4]$ gives the optimal result.

Step 2: Calculate the k-th iteration of $W = \left\{\omega_i\right\}_1^{min(ns,nm)}$. Following this, the ADMM-based framework is employed for solving equation (9). Experimental validation shows that $k = 1$ produces the best result.

To facilitate the computation, we introduce an auxiliary matrix T for subsequent solution.

$$\begin{aligned} \min_X \sum_{i=1}^{min(ns,nm)} \varpi_i\sigma_i + \frac{\lambda}{2}\left\|P_\Omega(X) - P_\Omega(H)\right\|_F^2 \\ \text{s.t. X = T, } 0 \le X_{ij} \le 1 (0 \le i \le ns, 0 \le j \le nm) \end{aligned} \tag{10}$$

The augmented Lagrangian form of equation (10) is represented below:

$$\ell(T, X, E, \lambda, \eta) = \sum_{i=1}^{min(ns,nm)} \varpi_{\text{i}}\sigma_i + \frac{\lambda}{2}\left\|P_\Omega(T) - P_\Omega(H)\right\|_F^2$$

$$+\operatorname{Tr}\left(\mathrm{E}^{\mathrm{T}}(X-T)\right)+\frac{\eta}{2}\|X-T\|_F^2 \tag{11}$$

where E denotes the Lagrange multiplier, η denotes the penalty parameter. The minimization of equation (11) is an iterative computation process. During the k-th iterations, T_{k+1}, X_{k+1}, and E_{k+1} are calculated serially. The following is the detailed procedure for the iterative algorithm's solution process.

$$T_{k+1} = \underset{0 \le T_{ij} \le 1}{argmin} L(T, X_k, E_k, \lambda, \eta) \tag{12}$$

$$X_{k+1} = \underset{X}{argmin} L(T_{k+1}, X, E_k, \lambda, \eta) \tag{13}$$

$$E_{k+1} = \underset{X}{argmin} L(T_{k+1}, X_{k+1}, E, \lambda, \eta) \tag{14}$$

The optimized T_{k+1}, X_{k+1} and E_{k+1} are given by Equations (12-14). Thus, we obtain T_{k+1}, X_{k+1} and E_{k+1} as follows.

We first derive it and then according to reference [39], a closed-form solution for T_{k+1} is obtained as follows:

$$\begin{aligned}\overline{T}_{k+1} &= (\frac{1}{\eta}Z_k + \frac{\lambda}{\eta}P_\Omega(H) + X_k)\\ &-\frac{\lambda}{\lambda+\eta}P_\Omega(\frac{1}{\eta}Z_k + \frac{\lambda}{\eta}P_\Omega(H) + X_k)\end{aligned} \tag{15}$$

Applying an interval [0,1] range constraint to the values in $\overline{T}_{k+1}$, we can obtain the final T_{k+1} as follows:

$$\left[T_{\mathrm{k+1}}\right]_{ij} = \begin{cases} 0 & \text{if } \overline{T}_{k+1_{ij}} < 0 \\ \overline{T}_{k+1} & \text{if } 0 \le \overline{T}_{k+1_{ij}} \le 1 \\ 1 & \text{if } \overline{T}_{k+1_{ij}} > 1 \end{cases} \tag{16}$$

Based on the singular value shrinkage operator [40], X_{k+1} are represented as follows:

$$X_{k+1} = \underset{X}{argmin} \sum_{i=1}^{min(ns,nm)} \omega_i\sigma_i + \frac{\eta}{2}\left\|X - (T_{k+1} - \frac{1}{\eta}E_k)\right\|_F^2 \tag{17}$$

$$X_{k+1} = Soft_{\omega,\frac{1}{\eta}}(T_{k+1} - \frac{1}{\eta}E_k) \tag{18}$$

Then, update T_{k+1} and X_{k+1} by varying the Lagrange multiplier E_{k+1}. Here, We use the gradient ascent algorithm to compute E_{k+1}.

$$E_{k+1} = E_k + \eta(X_{k+1} - T_{k+1}) \tag{19}$$

Algorithm 1 Truncated schatten p-norm minimization

```
Require: $H, S_{sm}^S, S_{sm}^C, S_{sm}^G, S_{sm}^P, S_m^G, S_m^D, \lambda, \eta, p, r, \varepsilon_1, \varepsilon_2$
Ensure: $H^*$
$S_{sm} \leftarrow Matrix\_Fusion(S_{sm}^S, S_{sm}^C, S_{sm}^G, S_{sm}^P)$
$S_m \leftarrow Matrix\_Fusion(S_m^G, S_m^D)$
$S_{SM} \leftarrow GRBF_{SM}(S_{sm}),\ S_M \leftarrow GRBF_{miRNA}(S_m)$
$X_1 = P_\Omega(H),\ T_1 = X_1,\ E_1 = X_1,\ l = 0,\ k = 0$
outer-loop:
    $l \leftarrow l + 1$
    $X_l$: $[U_l, \triangle_l, V_l] = SVD(X_l)$, where $U_l = (\mu_1, \dots, \mu_{min(ns,nm)}) \in R^{ns\times ns)}$, $V_l = (\nu_1, \dots, \nu_{min(ns,nm)}) \in R^{nm\times nm}$.
    Compute $A_l = (\mu_1, \dots, \mu_r)^T$ and $B_l = (\nu_1, \dots, \nu_r)^T$.
    Calculate weight $W = \left\{p(1-\sigma_i(B^TA))(\sigma_i(X_1))^{p-1}\right\}_1^{min(ns,nm)}$
    inner-loop:
        $k \leftarrow k + 1$
        $\overline{T}_{k+1} = (\frac{1}{\eta}E_k + \frac{\lambda}{\eta}P_\Omega(H) + X_k) - \frac{\lambda}{\lambda+\eta}P_\Omega(\frac{1}{\eta}E_k + \frac{\lambda}{\eta}P_\Omega(H) + X_k)$
        $T_{k+1} = \begin{cases} 0 & if\ \ \overline{T}_{k+1} \le 0 \\ \overline{T}_{k+1} & if\ \ 0 < \overline{T}_{k+1} < 1 \\ 1 & if\ \ \overline{T}_{k+1} \ge 1 \end{cases}$
        $X_{k+1} = S_{\omega,\frac{1}{\eta}}(T_{k+1} - \frac{1}{\eta}E_k)$
        $E_{k+1} = E_k + \eta(X_{k+1} - T_{k+1})$
    Until: $S1_{k+1} = \frac{\|X_{k+1}-X_k\|_F}{\|X_k\|_F} \le \varepsilon_1$
           $S2_{k+1} = \frac{\|S1_{k+1}-S1_k\|_F}{max\{1,|S1_k|\}} \le \varepsilon_2$
Until: Iteration number $l$=2
return $H^*$
```

Finally, we obtain the complemented MMA matrix until the convergence conditions $\mathrm{S1}_{k+1} = \frac{\|X_{k+1}-X_k\|_F}{\|X_k\|_F} \le \varepsilon_1$ and $\mathrm{S2}_{k+1} = \frac{|d1_{k+1}-d_{1k}|}{\max\{|d1_k|,1\}} \le \varepsilon_2$ are satisfied. The values ε_1 and ε_2 are specified according to the work of Yang et al. [41]. Finally, the updated association matrix H^* can be acquired. Therefore, the specific algorithm of Truncated schatten p-norm minimization is shown in Algorithm 1.

2.4.3. *Truncated matrix factorization*

In the next third step, on the basis of the former CMF [42], we present an enhanced CMF approach for identifying MMAs. The method has three main advantages: (i) The model incorporated comprehensive SM/miRNA similarities processed through GRBF. (ii) To address sparse correlation matrices, we employ the truncated schatten p-norm minimization method to fill in missing values. (iii) Flexibility is introduced by allowing modification of the singular value dimension, thereby reducing the rank of the prediction score matrix and enhancing the association probability. The primary objective of matrix factorization is to decompose the MMA matrix into two feature matrices, A and B, where $H^* \approx AB^T$. Subsequently, we formulate the objective function for truncated matrix factorization utilizing the matrices A and B as follows:

$$\min_{\text{A,B}} = \left\| H^* - AB^T \right\|_F^2 + \lambda_t \left(\|A\|_F^2 + \|B\|_F^2 \right)$$
$$+\lambda_s \left\| S_{SM} - AA^T \right\|_F^2 + \lambda_m \left\| S_M - BB^T \right\|_F^2 \qquad (20)$$

where $\|\cdot\|_F$ indicates Frobenius norm and λ_t, λ_s, λ_m are non-negative parameters. The first entry represents an approximate model of the matrix H^*, aiming to identify the latent feature matrices A and B. The second entry serves as Tikhonov regularization, minimizing the norm of A and B to prevent overfitting. The third and fourth entries signify regularization requirements ensuring that the latent feature vectors of similar SMs and miRNAs are analogous.

Furthermore, to acquire the original value of A and B, we utilize the SVD approach for the matrix H^* as follows:

$$\begin{gathered}[U,S,V] = SVD(H^*,k), A = US_k^{1/2}, B = VS_k^{1/2} \\ where\ H^* = U * S_k * V\end{gathered} \qquad (21)$$

where S_k contains the k largest singular values of H^* and $U \in R^{ns*k}/V \in R^{k*nm}$ contains the associated singular vectors. Through experimental verification, it has been established that k=6 yields optimal results.

We apply the alternating least squares approach to address the optimization problem of equation (20). First, equation (20) is represented as L. Subsequently, by setting $\partial L/\partial A = 0$ and $\partial L/\partial B = 0$, we iteratively update A and B until convergence. The iterative formulas for A and B are expressed as follows:

$$A = (H^*B + \lambda_s S_{SM} A)(B^T B + \lambda_t I_k + \lambda_s A^T A)^{-1} \qquad (22)$$

$$B = (H^{*T} A + \lambda_m S_M B)(A^T A + \lambda_t I_k + \lambda_m B^T B)^{-1} \qquad (23)$$

where I_k is a diagonal matrix with dimension k, having all diagonal elements set to 1. Finally, the final prediction score matrix $\overline{H} = AB^T$ can be acquired. Therefore, the specific algorithm of truncated matrix factorization is shown in Algorithm 2.

Algorithm 2 truncated matrix factorization

```
Require: $H^*$, $S_{SM}$, $S_M$, $\lambda_s$, $\lambda_m$, $\lambda_t$
Ensure: $\overline{H}$
[U, S, V] = $SVD(H^*, k), A = US_k^{1/2}, B = VS_k^{1/2}$
Repeat
    Update $A$: $A = (H^*B + \lambda_{sm} S_{sm}^* A)(B^T B + \lambda_t I_k + \lambda_{sm} A^T A)^{-1}$
    Update $B$: $B = (H^{*T} A + \lambda_m S_m^* B)(A^T A + \lambda_t I_k + \lambda_m B^T B)^{-1}$
Until convergence
return $\overline{H} = AB^T$
```

3. Results

3.1. Performance evaluation based on widely used datasets

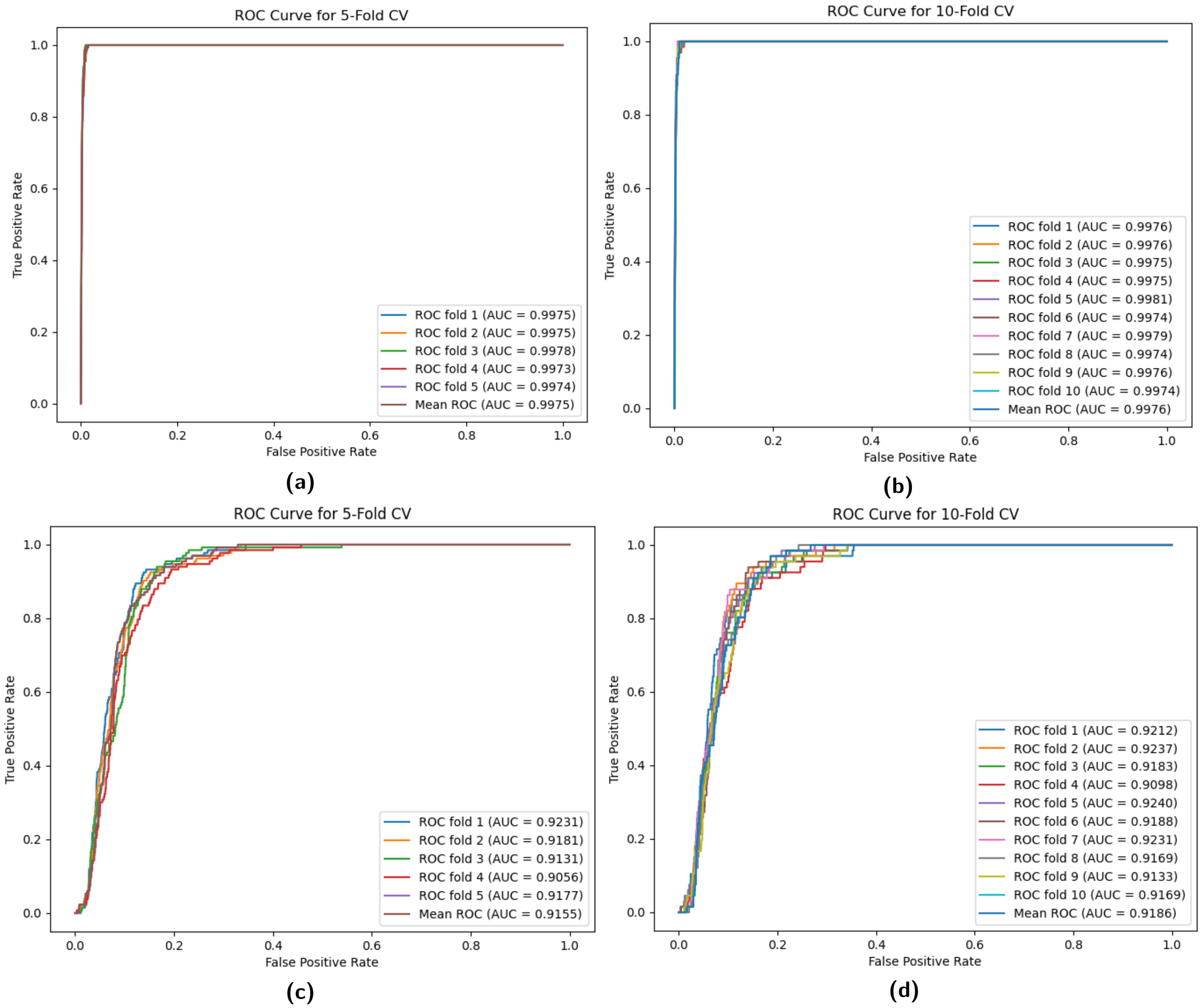
In this work, we conducted 5-fold CV and 10-fold CV on dataset 1 and dataset 2 to assess MECDMC's predictive capability. All predicted MMAs were ranked using their scores. Figure 2 illustrates that MECDMC achieved impressive AUC values, specifically attaining 0.9975 and 0.9976 for 5-fold and 10-fold CV, respectively, on dataset 1. Similarly, for dataset 2, the model yielded AUC of 0.9155 and 0.9186 for 5-fold and 10-fold CV. The AUC values of MECDMC are consistently high based on the CV scenarios of both datasets, which highlights the strong predictive capacity of MECDMC. Notably, the Receiver Operating Characteristic (ROC) graphs from the CV experiments further confirm the model's strong robustness.

3.2. Comparison with other methods

To further assess MECDMC's capability for identifying potential MMAs, we compared it with six alternative methods: TSPN [12], AMCSMMA [11], BNNRSMMA [10], DCMF [16], EKRRSMMA [22], and SLHGISMMA [18]. This comparison was carried out through 5-fold CV and global LOOCV on both dataset 1 and dataset 2, respectively. Figure 3 illustrates the experimental outcomes of MECDMC with the other six competitive approaches. We repeated the 5-fold CV 100 times and utilized its mean and standard deviation as the result. In the 5-fold CV of dataset 1/dataset2, the AUC values for MECDMC, TSPN, AMCSMMA, BNNRSMMA, DCMF, EKRRSMMA, and SLHGISMMA were 0.9976 ± 0.0002/0.9156 ± 0.0016, 0.9934 ± 0.0004/0.8834 ± 0.0038, 0.9910 ± 0.0004/0.8768 ± 0.0039, 0.9758 ± 0.0029/0.8759 ± 0.0041, 0.9836 ± 0.0012/0.8632 ± 0.0042, 0.9767 ± 0.0014/0.8560 ± 0.0027, and 0.9241 ± 0.0052/0.7724 ± 0.0032, respectively. Notably, MECDMC exhibited the highest AUC, surpassing the second-best model (TSPN) by 0.0042/0.0322. The standard deviation under 5-fold CV using dataset 1 and dataset 2 is only 0.0002 and 0.0016, which reflects the robustness of MECDMC. Furthermore, MECDMC achieved the highest AUC values in global LOOCV for both datasets. This comprehensive comparison clearly indicates that MECDMC excels in identifying potential MMAs, showcasing its superior performance over the evaluated methods.

3.3. Performance evaluation based on new dataset

It is unclear whether the MMAs used for performance comparisons are sensitive and the available data are significantly unbalanced between positive and negative samples. To address this, we curated a new dataset comprising 796 known MMAs involving 831 SMs and 541 miRNAs. To ensure a balanced representation of positive and negative samples, the same number of negative samples as positive samples were drawn from unknown MMAs. For the evaluation of MECDMC's performance in identifying MMAs, we employed 5-fold CV. Additionally, to comprehensively assess the overall impact of MECDMC, several standard



(a)

(b)

(c)

(d)

Figure 2: (a) MECDMC executes 5-fold CV on the dataset 1; (b) MECDMC executes 10-fold CV on the dataset 1; (c) MECDMC executes 5-fold CV on the dataset 2; (d) MECDMC executes 10-fold CV on the dataset 2.

evaluation metrics were employed, comprising Accuracy, Precision, Recall, F1-score, and MCC. The definitions of these indicators are detailed as shown below.

$$Precision = \frac{TP}{TP + FP} \tag{24}$$

$$Recall = \frac{TP}{TP + FN} \tag{25}$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{26}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{27}$$

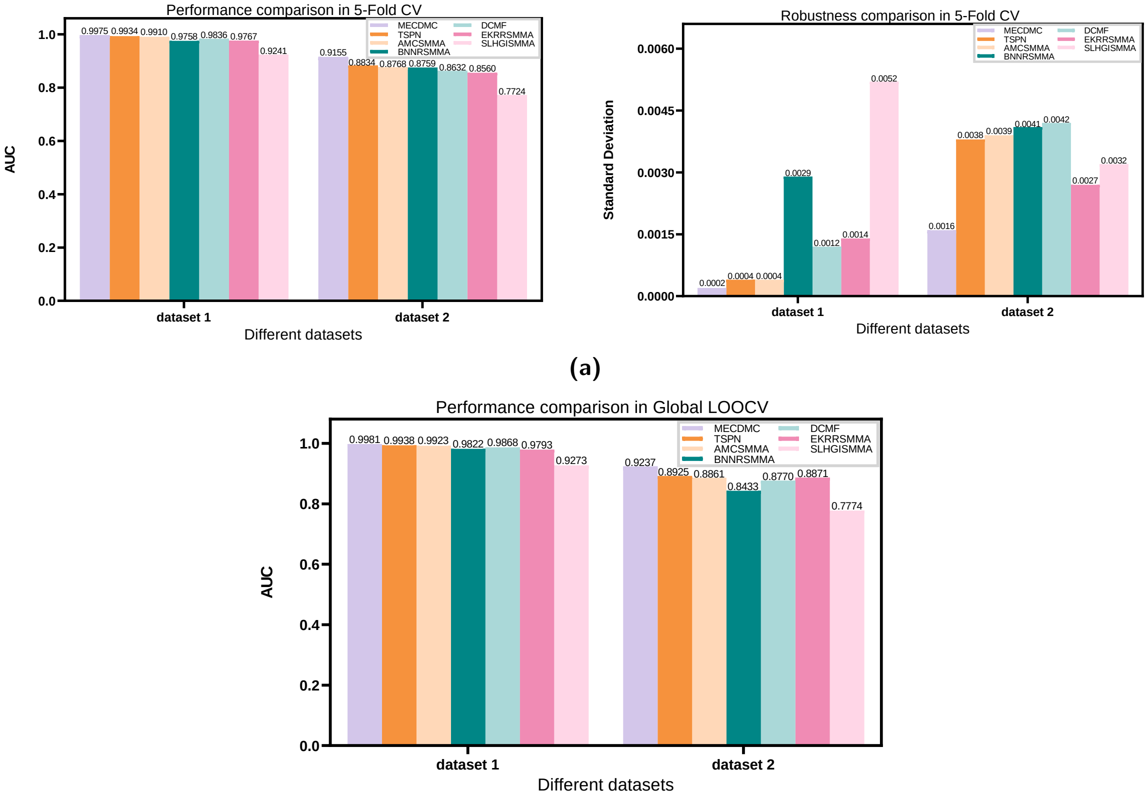
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{28}$$

where TP, TN, FP, and FN represent, respectively, the number of correctly predicted known MMAs, the number of correctly predicted unknown MMAs, the number of mispredicted known MMAs, and the number of mispredicted unknown MMAs.

Here, we applied the same parameter settings as in dataset 1 and dataset 2. The MECDMC's performance is presented in Table 2 and Figure 4. Table 2 highlights that MECDMC achieves an average accuracy, precision, recall, F1 score, and MCC of 0.9892, 0.9573, 0.9789, 0.9680, and 0.9616, respectively. Additionally, MECDMC's mean AUC and Area Under the Precision-Recall Curve (AUPR) reached 0.9978 and 0.9872, respectively. The associated ROC and Precision-Recall (PR) curves are depicted in Figure 4.

3.4. miRNA-disease association prediction

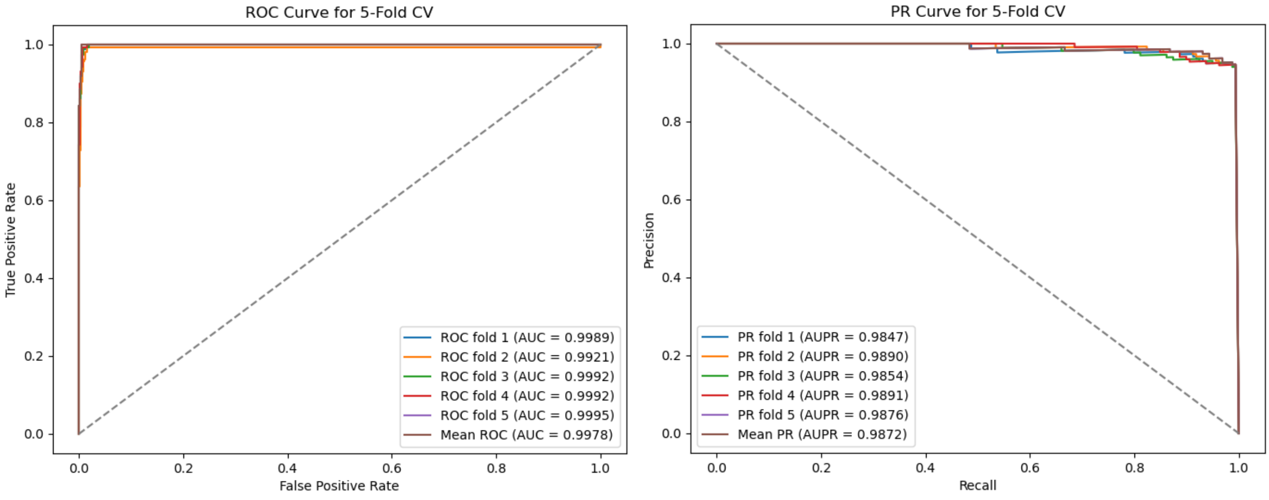
To showcase the versatility of MECDMC in association prediction beyond its original application, we employed



(a)

(b)

Figure 3: (a) Comparing the performance of various advanced models in terms of AUC and standard deviation (SD) through 5-fold CV on dataset 1 and dataset 2. (b) Comparing the performance of several advanced approaches in terms of AUC using global LOOCV on dataset 1 and dataset 2.



(a) **(b)**

Figure 4: (a) The ROC curves obtained by MECDMC on new dataset. (b) The PR curves obtained by MECDMC on new dataset.

our model to identify miRNA-disease associations. Leveraging a well-established dataset in the field [43], encompassing 495 miRNAs, 5430 known miRNA-disease associations, and 383 diseases, we conducted 5-fold CV. Our results were compared with several existing models, including the bounded kernel paradigm regularization model (BNNRMDA) [44], the weighted bipartite network projection model (WBNPMD) [45], the dichotomous network recommendation and the KATZ model (KATZBNRA) [46], the probability matrix decomposition model (PMFMDA) [47], and

Table 2
Outcomes of 5-fold CV acquired by MECDMC on new dataset

Fold	Accuracy	Precision	Recall	F1-score	MCC
1	0.9925	0.9635	0.9925	0.9778	0.9734
2	0.9912	0.9701	0.9774	0.9738	0.9685
3	0.9849	0.9481	0.9624	0.9552	0.9462
4	0.9875	0.9489	0.9774	0.9630	0.9556
5	0.9899	0.9559	0.9848	0.9701	0.9643
Average	0.9892	0.9573	0.9789	0.9680	0.9616

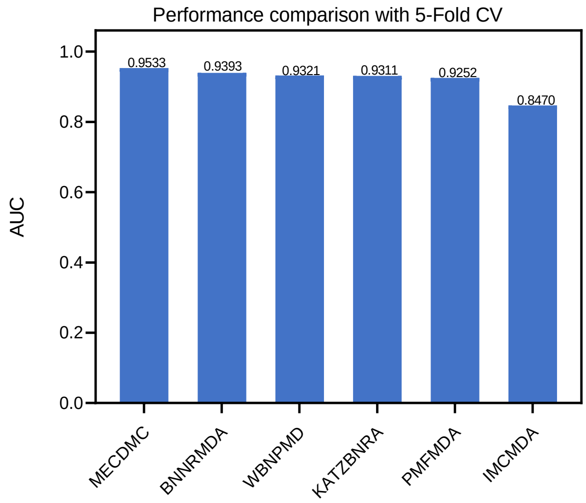


Figure 5: Comparison with other approaches using 5-fold CV on a miRNA-disease dataset.

the induced matrix completion model (IMCMDA) [48]. As shown in Figure 5, our model exhibits strong performance, further demonstrating the superior generalization ability of MECDMC in different association prediction domains.

3.5. Parameter sensitivity analysis

In the MECDMC method, we employ truncated matrix factorization to adjust the rank of the correlation matrix, facilitating dimensionality reduction for efficient extraction of key features. To clarify the effect of rank size on MECDMC's predictive performance, we conducted experiments setting the rank to 3, 6, 12, 24, and 48 on dataset 1. A comparative analysis using 5-fold CV was performed to identify the rank that yields the optimal model performance. As depicted in Figure 6, the model achieves its best performance during training when the rank is set to 6. Consequently, we establish the rank of the processed MMA matrix as 6.

3.6. Ablation experiments

To assess the importance of GRBF and CDMC, we performed the following five experiments.

• MECDMC-A: GRBF was omitted and extensive SM and miRNA similarity was used.

• MECDMC-B: first the truncated matrix factorization, then the truncated schatten p-norm minimization.

Table 3
Results of ablation experiments of MECDMC with all variants

Models	dataset 1	dataset 2
MECDMC	0.9975	0.9155
MECDMC-A	0.9807	0.8990
MECDMC-B	0.9972	0.9148
MECDMC-C	0.8447	0.6921
MECDMC-D	0.8124	0.5665

Table 4
Ablation experiments with different combinations of bimatrix complement in running time

Models	dataset 1	dataset 2
MECDMC	3.37s	0.34s
MECDMC-B	123.98s	3.16s

• MECDMC-C: truncated schatten p-norm minimization was omitted.

• MECDMC-D: truncated matrix factorization was omitted.

• MECDMC: the complete MECDMC algorithm was used.

Based on the realization of 5-fold CV on both datasets, we used AUC value as an assessment metric for ablation experiment. As shown in Table 3, all variants of MECDMC exhibited decreased performance, validating the prediction that all modules effectively contribute to miRNA-lncRNA interactions. Additionally, we compared MECDMC and MECDMC-B in terms of the time required to obtain the prediction matrices on the two datasets. In addition, we have the following findings: (1) MECDMC outperforms MECDMC-A, suggesting that matrix enhancement using Gaussian radial basis functions aids in predicting MMAs. (2) MECDMC and MECDMC-B outperform MECDMC-C and MECDMC-D, indicating that double-matrix complementation is superior to single-matrix complementation for predicting MMAs. (3) The AUC of MECDMC is better than that of MECDMC-B, and the running time of MECDMC is significantly less than that of MECDMC-B (as listed in Table 4), demonstrating the advantage of an architecture that performs truncated schatten p-norm minimization followed by truncated matrix factorization.

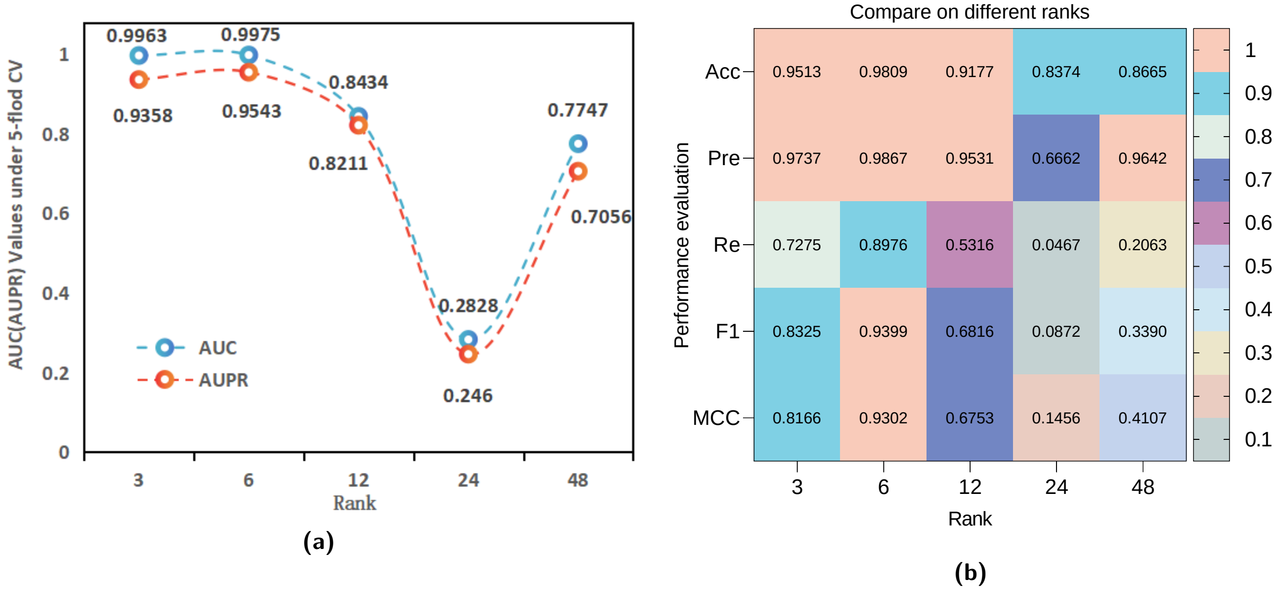


Figure 6: (a) AUC and AUPR values obtained by MECDMC based on the size of different ranks of the truncated matrix decomposition. (b) MECDMC was assessed using various ranks for multi-evaluation indicators.

3.7. Case studies

We conducted practical applications of the MECDMC model, testing its efficacy with two common SM (5-FU and Diethylstilbestrol). The model was trained using data from dataset 1. Specifically, for 5-FU and Diethylstilbestrol, we treated certain miRNAs associated with these SMs as unknown associations, essentially considering them as novel entities. For each SM under investigation, candidate miRNAs were ranked using their predictive correlation scores. The top 50 candidates were then subjected to validation through literature analysis on PubMed. In all case studies, a large number of miRNAs associated with specific SMs were experimentally validated, affirming the reliability of MECDMC in practical applications.

5-FU is an antimetabolic drug that blocks DNA and RNA synthesis in cancer cells. Among the top 50 pairs relevant to 5-FU, 36 connections have been validated, as detailed in Table 5. In particular, Huang et al. [49] discovered that inhibiting inflammatory and metastatic genes in colon cancer via miR-142-3p enhances the anti-tumor efficacy of 5-FU. Furthermore, Biswal et al. [50] demonstrated that miR-203a plays a role in augmenting the chemosensitivity of 5-FU in various malignant tumors through its interaction with target genes. Diethylstilbestrol is a synthetic nonsteroidal estrogen used in the treatment of certain cancers. Among the top 50 pairs associated with Diethylstilbestrol, 35 connections have been confirmed, as delineated in Table 6. Specifically, Li et al. [51] identified repressive chromatin marks on the down-regulated motif miR-9-3 in epithelial cells pre-exposed to hexestrol.

Table 5
Using dataset 1, we made predictions for the top 50 5-FU-related miRNAs.

miRNA(1-25)	Support	miRNA(26-50)	Support
hsa-mir-518c	unproven	hsa-mir-518d	unproven
hsa-mir-9-1	22614822	hsa-mir-451a	33255413
hsa-mir-9-2	22614822	hsa-mir-515-1	unproven
hsa-mir-126	26062749	hsa-mir-34a	24447928
hsa-mir-17	19956872	hsa-mir-200b	17702597
hsa-mir-9-3	22614822	hsa-mir-663a	26198104
hsa-mir-27a	26198104	hsa-mir-23a	26198104
hsa-mir-337	33628811	hsa-mir-502	37790850
hsa-mir-27b	17702597	hsa-mir-22	25449431
hsa-mir-409	unproven	hsa-mir-222	19956872
hsa-mir-371a	34844630	hsa-let-7b	25789066
hsa-mir-30c-1	unproven	hsa-mir-194-2	22614822
hsa-mir-514a-1	unproven	hsa-mir-671	21506117
hsa-mir-514a-2	unproven	hsa-mir-124-1	32382656
hsa-mir-1915	21506117	hsa-mir-124-2	32382656
hsa-mir-99b	unproven	hsa-mir-124-3	32382656
hsa-mir-24-1	26198104	hsa-mir-15b	22614822
hsa-mir-148a	23056401	hsa-mir-203a	37792370
hsa-mir-542	unproven	hsa-mir-345	unproven
hsa-mir-34b	37488217	hsa-mir-181d	35014676
hsa-mir-595	unproven	hsa-mir-219-a	unproven
hsa-mir-324	30103475	hsa-mir-142	31349708
hsa-mir-515-2	33536745	hsa-mir-26a-1	29296753
hsa-mir-92a-1	19956872	hsa-mir-520f	unproven
hsa-mir-106a	19956872	hsa-mir-431	unproven

Table 6
Using dataset 1, we made predictions for the top 50 Diethylstilbestrol-related miRNAs.

miRNA(1-25)	Support	miRNA(26-50)	Support
hsa-mir-194-2	19549897	hsa-mir-25	19549897
hsa-mir-22	19549897	hsa-mir-15a	19549897
hsa-mir-29a	unproven	hsa-mir-34a	19549897
hsa-mir-375	19549897	hsa-mir-671	19549897
hsa-mir-346	19549897	hsa-mir-26a-1	19549897
hsa-mir-194-1	19549897	hsa-mir-195	19549897
hsa-mir-32	unproven	hsa-mir-130b	19549897
hsa-mir-494	19549897	hsa-mir-218-2	19549897
hsa-mir-141	unproven	hsa-mir-335	19549897
hsa-mir-498	unproven	hsa-mir-497	19549897
hsa-mir-373	19549897	hsa-mir-146b	19549897
hsa-mir-152	19549897	hsa-mir-370	19549897
hsa-mir-630	unproven	hsa-mir-196a-1	19549897
hsa-mir-345	19549897	hsa-mir-196a-2	19549897
hsa-mir-320a	unproven	hsa-mir-26a-2	19549897
hsa-mir-16-1	unproven	hsa-mir-454	19549897
hsa-mir-16-2	unproven	hsa-mir-33a	unproven
hsa-mir-29b-1	unproven	hsa-mir-361	19549897
hsa-mir-29b-2	unproven	hsa-mir-638	unproven
hsa-mir-193a	unproven	hsa-mir-20b	19549897
hsa-mir-106a	19549897	hsa-mir-422a	19549897
hsa-mir-92a-1	19549897	hsa-mir-758	19549897
hsa-mir-150	unproven	hsa-mir-337	19549897
hsa-mir-9-3	19549897	hsa-mir-92b	19549897
hsa-mir-106b	19549897	hsa-mir-221	unproven

4. Discussion

The dysregulated expression of miRNAs is intricately linked to the onset and progression of various human diseases [52, 53]. SM have the ability to target miRNAs, thereby modulating their expression and function. Delving deep into the relationship between SMs and miRNAs enhances our understanding of both the properties of SMs and the functions of miRNAs. This exploration can contribute to optimizing the efficacy of existing drugs and minimizing adverse effects.

Matrix completion is a feasible approach, however existing methods have limitations such as high computational complexity and not optimal solutions. In addition, similarity matrices may be noisy, and their topological information not fully exploited. Therefore, designed Therefore, we designed a matrix enhancement and collaborative double matrix completion framework, MECDMC, to identify MMAs. MECDMC utilizes a Gaussian radial basis function to augment the similarity matrix, effectively mitigating errors inherent in data collection and integration processes. Second, we develop a CDMC framework that skillfully combines two complementary matrix completion techniques. We first use truncated schatten p-paradigm minimization to complement the association matrix, which solves the problem of overly sparse association matrix. We then use truncated matrix factorization on the updated correlation matrix to obtain the final prediction results. The MECDMC's effectiveness is due to three key factors: (1) We use GRBF to perform matrix enhancement of the integrated SM/miRNA similarity to obtain the final refined SM and miRNA similarity. This step takes into account the structural information of the integrated similarity matrix, which improves the accuracy of the similarity measure. (2) The truncated schatten p-norm considers the physical nature of the singular values, effectively complementing missing values better than other rank approximation norms. In addition, the truncated matrix factorization greatly improves the speed of the model by truncating the number of singular values. (3) MECDMC skillfully combines two types of matrix completion methods, taking full advantage of the feature selectivity and robustness of the former and the efficient computation and interpretability of the latter, and exhibits comprehensive and flexible performance in MMA prediction. (4) We compare MECDMC with six advanced approaches using 5-fold CV and 10-fold CV on dataset 1 and dataset 2, and the results show that MECDMC outperforms them. Noteworthy achievements include the exceptional performance of MECDMC in miRNA-disease association prediction, highlighting its promise as a predictive model. Moreover, the model successfully predicted a substantial number of miRNAs in both case studies, underscoring MECDMC's potential as an effective and versatile tool in this specific domain.

However, there is still potential for refinement in our proposed model. First, the setting of model parameters still relies on the grid search method based on CV experiments, which can be time-consuming and may not be optimal. Secondly, our model has yet to delve into the mechanisms underlying the potential changes in SM-regulated target miRNAs. Moving forward, we anticipate integrating miRNA expression regulation information into the model to facilitate a more comprehensive and nuanced prediction.

5. Conclusion

MECDMC first augments the similarity matrix using a Gaussian radial basis function, and then combines truncated schatten p-norm minimization and truncated matrix factorization methods to efficiently predict MMAs. Additionally, we curated a novel dataset by incorporating 132 newly identified associations alongside the existing 664 known associations, affirming MECDMC's sustained excellent predictive efficacy. Moreover, MECDMC demonstrates remarkable generalization capabilities across other association prediction domains.

6. Appendix I

We let $Q(\sigma(X)) = \sum_{i=1}^{ns+nm}(1 - \sigma_i(B^T A)(\sigma_i(X))^p$. Subsequently, we computed the derivative of the equation with regard to $\sigma(X)$.

$$\bigtriangledown Q(\sigma(X)) = \sum_{i=1}^{ns+nm} p(1 - \sigma_i(B^T A)(\sigma_i(X))^{p-1} \quad (29)$$

Then, the first-order Taylor expansion for $Q(\sigma(X))$ was attained as shown below:

$$\begin{aligned} Q(\sigma(X)) &= Q\left(\sigma\left(X_k\right)\right) + \left\langle \nabla Q\left(\sigma\left(X_k\right)\right), \sigma(X) - \sigma\left(X_k\right)\right\rangle \\ &= \nabla Q\left(\sigma\left(X_k\right)\right) \cdot \sigma(X) \\ &= \sum_{\mathrm{i}=1}^{ns+nm} \mathrm{p}\left(1-\sigma_{\mathrm{i}}\left(B^T A\right)\right)\left(\sigma_i\left(X_k\right)\right)^{p-1} \cdot \sigma_i(X) \end{aligned} \tag{30}$$

We let $\omega_i = p(1-\sigma_i\left(B^T A\right))\left(\sigma_i(X_k)\right)^{p-1}$. Then $Q\left(\sigma(x)\right) = \sum_{i=1}^{ns+nm} \omega_i \sigma_i(X)$, where $W := \left\{\omega_i\right\}_1^{ns+nm}$ is a weight sequence. After processing, we acquired the following solvable convex optimization model:

$$\min_{\mathrm{X}} \sum_{i=1}^{ns+nm} \omega_i \sigma_i \text{ s.t. } \mathrm{P}_\Omega(X) = \mathrm{P}_\Omega(H) \tag{31}$$

7. Data availability

The code for MECDMC and datasets are accessible on GitHub at https://github.com/Skyrocket-lty/MECDMC.git.

8. Declaration of competing interest

The authors state that they do not have any known competing interests.

9. Funding

This research received backing from the National Key Research and Development Project of China (2021YFA1000102, 2021YFA1000103) and the Natural Science Foundation of Shandong Province (ZR2023MF053).

References

[1] Kinga Nemeth, Recep Bayraktar, Manuela Ferracin, and George A Calin. Non-coding rnas in disease: from mechanisms to therapeutics. *Nature Reviews Genetics*, pages 1–22, 2023.

[2] Yirong Wang, Xiaolu Tang, and Jian Lu. Convergent and divergent evolution of microrna-mediated regulation in metazoans. *Biological Reviews*, 99(2):525–545, 2024.

[3] Giulia Alloisio, Chiara Ciaccio, Giovanni Francesco Fasciglione, Umberto Tarantino, Stefano Marini, Massimo Coletta, and Magda Gioia. Effects of extracellular osteoanabolic agents on the endogenous response of osteoblastic cells. *Cells*, 10(9):2383, 2021.

[4] Rajesha Rupaimoole and Frank J Slack. Microrna therapeutics: towards a new era for the management of cancer and other diseases. *Nature reviews Drug discovery*, 16(3):203–222, 2017.

[5] Pamela Agbu and Richard W Carthew. Microrna-mediated regulation of glucose and lipid metabolism. *Nature reviews Molecular cell biology*, 22(6):425–438, 2021.

[6] Xin Xu, Yong-gang Lv, Chang-you Yan, Jun Yi, and Rui Ling. Enforced expression of hsa-mir-125a-3p in breast cancer cells potentiates docetaxel sensitivity via modulation of brca1 signaling. *Biochemical and biophysical research communications*, 479(4):893–900, 2016.

[7] Mario Dioguardi, Francesca Spirito, Diego Sovereto, Mario Alovisi, Giuseppe Troiano, Riccardo Aiuto, Daniele Garcovich, Vito Crincoli, Luigi Laino, Angela Pia Cazzolla, et al. Microrna-21 expression as a prognostic biomarker in oral cancer: Systematic review and meta-analysis. *International journal of environmental research and public health*, 19(6):3396, 2022.

[8] Sarah Bajan and Gyorgy Hutvagner. Rna-based therapeutics: from antisense oligonucleotides to mirnas. *Cells*, 9(1):137, 2020.

[9] Priya Mondal, Jagadish Natesh, Dhanamjai Penta, and Syed Musthapa Meeran. Progress and promises of epigenetic drugs and epigenetic diets in cancer prevention and therapy: A clinical update. In *Seminars in Cancer Biology*, volume 83, pages 503–522. Elsevier, 2022.

[10] Jia Qu, Xing Chen, Ya-Zhou Sun, Jian-Qiang Li, and Zhong Ming. Inferring potential small molecule–mirna association based on triple layer heterogeneous network. *Journal of cheminformatics*, 10:1–14, 2018.

[11] Jun Yin, Xing Chen, Chun-Chun Wang, Yan Zhao, and Ya-Zhou Sun. Prediction of small molecule–microrna associations by sparse learning and heterogeneous graph inference. *Molecular pharmaceutics*, 16(7):3157–3166, 2019.

[12] Na-Na Guan, Ya-Zhou Sun, Zhong Ming, Jian-Qiang Li, and Xing Chen. Prediction of potential small molecule-associated micrornas using graphlet interaction. *Frontiers in pharmacology*, 9:1152, 2018.

[13] Jie Li, Kecheng Lei, Zengrui Wu, Weihua Li, Guixia Liu, Jianwen Liu, Feixiong Cheng, and Yun Tang. Network-based identification of micrornas as potential pharmacogenomic biomarkers for anticancer drugs. *Oncotarget*, 7(29):45584, 2016.

[14] Shubhrangshu Ghosh and Pralay Mitra. Matpip: A deep-learning architecture with explainable ai for sequence-driven, feature mixed protein-protein interaction prediction. *Computer Methods and Programs in Biomedicine*, 244:107955, 2024.

[15] Li Peng, Yuan Tu, Li Huang, Yang Li, Xiangzheng Fu, and Xiang Chen. Daestb: inferring associations of small molecule–mirna via a scalable tree boosting model based on deep autoencoder. *Briefings in Bioinformatics*, 23(6):bbac478, 2022.

[16] Chun-Chun Wang, Chi-Chi Zhu, and Xing Chen. Ensemble of kernel ridge regression-based small molecule–mirna association prediction in human disease. *Briefings in Bioinformatics*, 23(1):bbab431, 2022.

[17] Chun-Chun Wang and Xing Chen. A unified framework for the prediction of small molecule–microrna association based on cross-layer dependency inference on multilayered networks. *Journal of chemical information and modeling*, 59(12):5281–5293, 2019.

[18] Chun-Chun Wang, Xing Chen, Jia Qu, Ya-Zhou Sun, and Jian-Qiang Li. Rfsmma: a new computational model to identify and prioritize potential small molecule–mirna associations. *Journal of chemical information and modeling*, 59(4):1668–1679, 2019.

[19] Xing Chen, Chi Zhou, Chun-Chun Wang, and Yan Zhao. Predicting potential small molecule–mirna associations based on bounded nuclear norm regularization. *Briefings in Bioinformatics*, 22(6):bbab328, 2021.

[20] Shudong Wang, Chuanru Ren, Yulin Zhang, Shanchen Pang, Sibo Qiao, Wenhao Wu, and Boyang Lin. Amcsmma: Predicting small molecule–mirna potential associations based on accurate matrix completion. *Cells*, 12(8):1123, 2023.

[21] Shudong Wang, Tiyao Liu, Chuanru Ren, Wenhao Wu, Zhiyuan Zhao, Shanchen Pang, and Yuanyuan Zhang. Predicting potential small molecule–mirna associations utilizing truncated schatten p-norm. *Briefings in Bioinformatics*, 24(4):bbad234, 2023.

[22] Jiawei Luo, Cong Shen, Zihan Lai, Jie Cai, and Pingjian Ding. Incorporating clinical, chemical and biological information for predicting small molecule-microrna associations based on non-negative matrix factorization. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(6):2535–2545, 2020.

[23] Yan Zhao, Xing Chen, Jun Yin, and Jia Qu. Snmfsmma: using symmetric nonnegative matrix factorization and kronecker regularized least squares to predict potential small molecule-microrna association. *RNA biology*, 17(2):281–291, 2020.

[24] Shu-Hao Wang, Chun-Chun Wang, Li Huang, Lian-Ying Miao, and Xing Chen. Dual-network collaborative matrix factorization for predicting small molecule-mirna associations. *Briefings in Bioinformatics*, 23(1):bbab500, 2022.

[25] Craig Knox, Vivian Law, Timothy Jewison, Philip Liu, Son Ly, Alex Frolkis, Allison Pon, Kelly Banco, Christine Mak, Vanessa Neveu, et al. Drugbank 3.0: a comprehensive resource for omics research on drugs. *Nucleic acids research*, 39(suppl_1):D1035–D1041, 2010.

[26] Xinyi Liu, Shuyuan Wang, Fanlin Meng, Jizhe Wang, Yan Zhang, Enyu Dai, Xuexin Yu, Xia Li, and Wei Jiang. Sm2mir: a database of the experimentally validated small molecules effects on microrna expression. *Bioinformatics*, 29(3):409–411, 2013.

[27] Yanli Wang, Jewen Xiao, Tugba O Suzek, Jian Zhang, Jiyao Wang, and Stephen H Bryant. Pubchem: a public information system for analyzing bioactivities of small molecules. *Nucleic acids research*, 37(suppl_2):W623–W633, 2009.

[28] Ming Lu, Qipeng Zhang, Min Deng, Jing Miao, Yanhong Guo, Wei Gao, and Qinghua Cui. An analysis of human microrna and disease associations. *PloS one*, 3(10):e3420, 2008.

[29] Qinghua Jiang, Yadong Wang, Yangyang Hao, Liran Juan, Mingxiang Teng, Xinjun Zhang, Meimei Li, Guohua Wang, and Yunlong Liu. mir2disease: a manually curated database for microrna deregulation in human disease. *Nucleic acids research*, 37(suppl_1):D98–D104, 2009.

[30] Andreas Ruepp, Andreas Kowarsch, Daniel Schmidl, Felix Buggenthin, Barbara Brauner, Irmtraud Dunger, Gisela Fobo, Goar Frishman, Corinna Montrone, and Fabian J Theis. Phenomir: a knowledgebase for microrna expression in diseases and biological processes. *Genome biology*, 11:1–11, 2010.

[31] Yingli Lv, Shuyuan Wang, Fanlin Meng, Lei Yang, Zhifeng Wang, Jing Wang, Xiaowen Chen, Wei Jiang, Yixue Li, and Xia Li. Identifying novel associations between small molecules and mirnas based on integrated molecular networks. *Bioinformatics*, 31(22):3638–3644, 2015.

[32] Assaf Gottlieb, Gideon Y Stein, Eytan Ruppin, and Roded Sharan. Predict: a method for inferring novel drug indications with application to personalized medicine. *Molecular systems biology*, 7(1):496, 2011.

[33] Masahiro Hattori, Yasushi Okuno, Susumu Goto, and Minoru Kanehisa. Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *Journal of the American Chemical Society*, 125(39):11853–11865, 2003.

[34] Sali Lv, Yan Li, Qianghu Wang, Shangwei Ning, Teng Huang, Peng Wang, Jie Sun, Yan Zheng, Weisha Liu, Jing Ai, et al. A novel method to quantify gene set functional association based on gene ontology. *Journal of The Royal Society Interface*, 9(70):1063–1072, 2012.

[35] Beijia Chen, Huaijiang Sun, Guiyu Xia, Lei Feng, and Bin Li. Human motion recovery utilizing truncated schatten p-norm and kinematic constraints. *Information Sciences*, 450:89–108, 2018.

[36] Lei Feng, Huaijiang Sun, Quansen Sun, and Guiyu Xia. Image compressive sensing via truncated schatten-p norm regularization. *Signal Processing: Image Communication*, 47:28–41, 2016.

[37] Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, and Yuling Yan. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM journal on optimization*, 30(4):3098–3121, 2020.

[38] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[39] Junfeng Yang and Xiaoming Yuan. Linearized augmented lagrangian and alternating direction methods for nuclear norm minimization. *Mathematics of computation*, 82(281):301–329, 2013.

[40] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on optimization*, 20(4):1956–1982, 2010.

[41] Mengyun Yang, Huimin Luo, Yaohang Li, and Jianxin Wang. Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics*, 35(14):i455–i463, 2019.

[42] Zhen Shen, You-Hua Zhang, Kyungsook Han, Asoke K Nandi, Barry Honig, De-Shuang Huang, et al. mirna-disease association prediction with collaborative matrix factorization. *Complexity*, 2017, 2017.

[43] Shudong Wang, Fuyu Wang, Sibo Qiao, Yu Zhuang, Kuijie Zhang, Shanchen Pang, Robert Nowak, and Zhihan Lv. Mshganmda: Meta-subgraphs heterogeneous graph attention network for mirna-disease association prediction. *IEEE journal of biomedical and health informatics*, 2022.

[44] Yidong Rao, Minzhu Xie, and Hao Wang. Predict potential mirna-disease associations based on bounded nuclear norm regularization. *Frontiers in Genetics*, 13:978975, 2022.

[45] Guobo Xie, Zhiliang Fan, Yuping Sun, Cuiming Wu, and Lei Ma. Wbnpmd: weighted bipartite network projection for microrna-disease association prediction. *Journal of translational medicine*, 17(1):1–11, 2019.

[46] Shiru Li, Minzhu Xie, and Xinqiu Liu. A novel approach based on bipartite network recommendation and katz model to predict potential micro-disease associations. *Frontiers in Genetics*, 10:1147, 2019.

[47] Junlin Xu, Lijun Cai, Bo Liao, Wen Zhu, Peng Wang, Yajie Meng, Jidong Lang, Geng Tian, and Jialiang Yang. Identifying potential mirnas-disease associations with probability matrix factorization. *Frontiers in genetics*, 10:1234, 2019.

[48] Xing Chen, Lei Wang, Jia Qu, Na-Na Guan, and Jian-Qiang Li. Predicting mirna–disease association based on inductive matrix completion. *Bioinformatics*, 34(24):4256–4265, 2018.

[49] Yan-Jiun Huang, Vijesh Kumar Yadav, Prateeti Srivastava, Alexander TH Wu, Thanh-Tuan Huynh, Po-Li Wei, Chi-Ying F Huang, and Tse-Hung Huang. Antrodia cinnamomea enhances chemo-sensitivity of 5-fu and suppresses colon tumorigenesis and cancer stemness via up-regulation of tumor suppressor mir-142-3p. *Biomolecules*, 9(8):306, 2019.

[50] Priyajit Biswal, Anthony Lalruatfela, Subham Kumar Behera, Sruti Biswal, and Bibekanand Mallick. mir-203aa multifaceted regulator modulating cancer hallmarks and therapy response. *IUBMB life*, 2023.

[51] Gerwin Heller, Marlene Weinzierl, Christian Noll, Valerie Babinsky, Barbara Ziegler, Corinna Altenberger, Christoph Minichsdorfer, György Lang, Balazs Döme, Adelheid End-Pfützenreuter, et al. Genome-wide mirna expression profiling identifies mir-9-3 and mir-193a as targets for dna methylation in non–small cell lung cancers. *Clinical cancer research*, 18(6):1619–1629, 2012.

[52] Boxue He, Zhenyu Zhao, Qidong Cai, Yuqian Zhang, Pengfei Zhang, Shuai Shi, Hui Xie, Xiong Peng, Wei Yin, Yongguang Tao, et al. mirna-based biomarkers, therapies, and resistance in cancer. *International journal of biological sciences*, 16(14):2628, 2020.

[53] Xing Liu, Yongguang Zhang, Peng Jiang, Jiachen Cai, Qiuhong Fu, Xiaolei Li, and Zhou Li. Ultrasonic cardiogram and mirna-21 analysis of cardiac dysfunction in patients with cardiac arrest following cardiopulmonary resuscitation. *Computer Methods and Programs in Biomedicine*, 190:105284, 2020.

Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.