frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 23 November 2022
DOI 10.3389/fmolb.2022.1009099

OPEN ACCESS

EDITED BY
Irina Sousa Moreira,
University of Coimbra, Portugal

REVIEWED BY
Panagiotis Alexiou,
Central European Institute of Technology (CEITEC), Czechia
Congshan Jiang,
Xi'an Children's Hospital, China

*CORRESPONDENCE
Jiuzhen Liang,
jzliang@cczu.edu.cn

SPECIALTY SECTION
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 05 August 2022
ACCEPTED 03 November 2022
PUBLISHED 23 November 2022

CITATION
Ni J, Cheng XL, Ni TG and Liang JZ (2022), Identifying SM-miRNA associations based on layer attention graph convolutional network and matrix decomposition.
Front. Mol. Biosci. 9:1009099.
doi: 10.3389/fmolb.2022.1009099

Identifying SM-miRNA associations based on layer attention graph convolutional network and matrix decomposition

Jie Ni, Xiaolong Cheng, Tongguang Ni and Jiuzhen Liang*

School of Computer Science and Artificial Intelligence and Aliyun School of Big Data and School of Software, Changzhou University, Changzhou, China

The accurate prediction of potential associations between microRNAs (miRNAs) and small molecule (SM) drugs can enhance our knowledge of how SM cures endogenous miRNA-related diseases. Given that traditional methods for predicting SM-miRNA associations are time-consuming and arduous, a number of computational models have been proposed to anticipate the potential SM–miRNA associations. However, several of these strategies failed to eliminate noise from the known SM-miRNA association information or failed to prioritize the most significant known SM-miRNA associations. Therefore, we proposed a model of Graph Convolutional Network with Layer Attention mechanism for SM-MiRNA Association prediction (GCNLASMMA). Firstly, we obtained the new SM-miRNA associations by matrix decomposition. The new SM-miRNA associations, as well as the integrated SM similarity and miRNA similarity were subsequently incorporated into a heterogeneous network. Finally, a graph convolutional network with an attention mechanism was used to compute the reconstructed SM-miRNA association matrix. Furthermore, four types of cross validations and two types of case studies were performed to assess the performance of GCNLASMMA. In cross validation, global Leave-One-Out Cross Validation (LOOCV), miRNA-fixed LOOCV, SM-fixed LOOCV and 5-fold cross-validation achieved excellent performance. Numerous hypothesized associations in case studies were confirmed by experimental literatures. All of these results confirmed that GCNLASMMA is a trustworthy association inference method.

KEYWORDS

microRNA, small molecule, deep learning, association prediction, matrix decomposition

1 Introduction

As a form of non-coding RNA (ncRNA), MicroRNA (miRNA), is roughly 22 nucleotides in length (Bartel, 2004; Hammond, 2015; Lu and Rothenberg, 2018). Lin-4 was the first human miRNA identified in1993 by Lee *et al.* in Caenorhabditis elegans (Lee et al., 1993; Wightman et al., 1993). With the advent of high-throughput sequencing technologies, an increasing number of miRNAs with important functions in human gene expression have been identified (Denzler et al., 2016; Tagliafierro et al., 2017; Thomou et al., 2017; Gam et al., 2018; Ghini et al., 2018; Liu et al., 2018). Specifically, miRNAs can attach to the 3′ UnTranslated Region (3' UTR) of target messenger RNAs (mRNAs) via base-pairing to control the degradation of target mRNAs and limit the translation of target mRNAs, hence regulating gene expression (Gorbea et al., 2017). In the control of target mRNA gene expression by miRNA, one miRNA may regulate many target mRNAs, or numerous miRNAs regulate one target mRNA (Saikia et al., 2020; Iwata et al., 2021; Zhong et al., 2021). Several studies demonstrated the role of miRNAs in the maturation of immune cells (Kumar Kingsley and Vishnu Bhat, 2017). Since the profound impact of miRNAs on biological development became apparent, numerous miRNA types have been identified to be involved in biological evolutionary processes (Rupaimoole and Slack, 2017; Cristino et al., 2019).

Small Molecule (SM) drugs are mostly composed of molecules with molecular weights typically fewer than 1,000 g/mol. More than 98 percent of today's drugs are SMs (Geng and Craig, 2021). The development of SMs that target miRNAs is a current trend in drug research (Dai and Tan, 2015; Yu et al., 2020). In previous drug development, protein enzymes and receptors were typically employed as therapeutic targets. Over 80 percent of drug development was intimately tied to protein enzymes and receptors (Deyle et al., 2017; Yekkirala et al., 2017; Nair et al., 2018; Lai-Kwon et al., 2021). In recent years, more scientific experiments have proven inextricable linkages between SMs and miRNAs (Healy et al., 2012; Monroig et al., 2015; Haniff et al., 2021). When miRNAs fail to regulate the gene expression of an organism, specific disorders such as cardiovascular diseases, neurological diseases and cancers may develop (Kumari et al., 2018; Xia et al., 2019; Dragomir et al., 2021). In addition, SMs are effective in regulating miRNA dysregulation to treat linked endogenous disorders, and numerous SMs have been created for clinical therapy of these diseases (Dragomir et al., 2021).

The development of novel SMs is facilitated by the accurate identification of miRNA-related SMs. Recent studies have focused on discovering possible associations between SMs and miRNAs (Chen et al., 2021; Li et al., 2021; Wang et al., 2021). Early identification approaches used high-throughput screening methods, such as mass spectrometry, fluorescence and reporter genes (Seth et al., 2005; Parsons et al., 2009; Carnevali et al., 2010; Chen et al., 2012). The most frequent method for discovering potential SM-miRNA associations is the reporter genes. On the basis of the reporter genes, a functional novel drug screening method capable of screening lead compounds was proposed. By substituting biomacromolecules with tiny organic compounds, the screening process for drugs could be expedited dramatically. The use of tiny organic compounds throughout the screening procedure could provide information on the functional responses of cells. (Wen et al., 2015). In drug screening research, luciferase reporter genes satisfy the requirements for high sensitivity, target specificity and high throughput (Thorne et al., 2010).

However, it was discovered that biological screening approaches are stochastic and time-consuming. With the proliferation of bioinformatics databases, the number of known SM-miRNA associations increased, as did the calculational methodologies for SM and miRNA similarity. Consequently, machine learning techniques obtained more precise prediction outcomes (Qu et al., 2019). Bioinformaticians have begun to employ machine learning techniques to predict probable SM-miRNA associations to circumvent time-consuming and labor-intensive biological investigations (Wang and Chen, 2019; Wang et al., 2019).

Among the previous methods for predicting probable SM-miRNA associations, (Qu et al., 2018), developed a model titled Triple Layer Heterogeneous Network based Small Molecule-MiRNA Association prediction (TLHNSMMA). TLHNSMMA first merged the known SM-miRNA associations, SM similarity and miRNA similarity into a three-layer heterogeneous network. The three-layer heterogeneous graph was then implemented with an iterative updating algorithm. Finally, the reconstructed SM-miRNA association matrix was obtained using an iterative propagation approach that made extensive use of global data. Based on the establishment of a three-layer SM-miRNA heterogeneous network, (Liu et al., 2020), suggested a novel model for potential SM-miRNA association prediction called Random Walk with Negative Samples (RWNS). Firstly, RWNS obtained integrated similarities of SM and miRNA. Then, Liu *et al.* devised a Credible Negative Sample extraction method (CNSMiRS) to extract plausible negative SM-miRNA samples under the premise that dissimilar SMs/miRNAs are unlikely to be associated with each other's related miRNAs/SMs. Finally, the reconstructed SM-miRNA association matrix was obtained by implementing a random walk algorithm on the constructed small molecule-disease-miRNA association network. However, the performance of TLHNSMMA and RWNS is dependent on the known SM-miRNA association adjacency matrix. Consequently, (Yin et al., 2019), suggested a model of Sparse Learning and Heterogeneous Graph Inference for Small Molecule-MiRNA Association prediction (SLHGISMMA). Yin *et al.* first used matrix decomposition on known SM-miRNA associations to obtain the new SM-miRNA associations. Then, the new SM-miRNA associations, integrated miRNA similarity and integrated SM similarity were incorporated into a heterogeneous network.

Finally, the reconstructed SM-miRNA association matrix was obtained using heterogeneous graph inference. Chen et al. (2021) recently proposed the Bounded Nuclear Norm Regularization for SM–miRNA Associations prediction (BNNRSMMA), which treated the problem of potential SM-miRNA association prediction as a matrix complementation problem. In addition, BNNRSMMA included a regularization term to remove the negative effects of data noise.

In recent years, improvements have been made to machine learning techniques, and deep learning has emerged as one of the brightest new stars (Wang et al., 2020). Deep learning has achieved exceptional results in traditional classification tasks, such as handwritten font recognition (Singh et al., 2021), computer vision (Borges Oliveira et al., 2021) and computational biology (Angermueller et al., 2016). In addition, deep learning has substantially affected the field of potential association prediction. For example, zeng *et al.* proposed a computational framework termed AOPEDF based on drug-target network and deep forest algorithm to predict potential drug-target associations (Zeng et al., 2020). AOPEDF attained excellent performance in identifying molecular targets among known drugs on two external validation datasets by comparison to other machine learning methods. Therefore, we proposed a model of Graph Convolutional Network with Layer Attention mechanism for SM-MiRNA Association prediction (GCNLASMMA). To evaluate the performance of GCNLASMMA, we used two types of cross validation, namely, 5-fold cross-validation and Leave-One-Out Cross Validation (LOOCV). Additionally, we also utilized two types of case studies to confirm the effectiveness of GCNLASMMA in identifying potential miRNAs for investigated SMs. The results showed that GCNLASMMA could accurately and effectively predict the SM-miRNA pairs most likely to be potentially associated.

2 Materials and methods

2.1 SM-miRNA associations

We named two datasets used in our work after dataset1 and dataset2. Eight hundred and thirty-one SMs in dataset1 were downloaded from three databases, namely SM2miR, DrugBank (Knox et al., 2011) and PubChem (Wang et al., 2009). Five hundred and forty-one miRNAs were downloaded from four databases, namely SM2miR, HMDD (Li et al., 2014), miR2Disease (Jiang et al., 2009) and PhenomiR (Ruepp et al., 2010). Six hundred and sixty-four known SM-miRNA associations were downloaded from a database, namely SM2miR V1.0 (Liu et al., 2013). On the basis of dataset1, we removed the SMs and miRNAs that did not constitute any known association. Then, we obtained dataset2 which included 286 different miRNAs, 39 different SMs and 664 known SM-miRNA association pairs. Specifically, the known SM-miRNA association A_{ij} between the i_{th} SM and the j_{th} miRNA was stored as follows.

2.2 Integration of SM similarities

The integrated SM similarity was calculated by (Lv et al., 2015). In his method, a total of four SM similarities were used, namely SM side effect similarity (Gottlieb et al., 2011), gene functional consistency-based similarity for SMs (Lv et al., 2012), SM chemical structure similarity (Hattori et al., 2003) and disease phenotype-based similarity for SMs (Gottlieb et al., 2011). In Lv's article, the side effect properties of SM were first downloaded from SIDe Effect Resource (SIDER) and calculated by Jaccard score to obtain SMs side effect similarities (Gottlieb et al., 2011). The calculation of gene functional consistency-based similarities for SMs was implemented on the target genes of SMs obtained from the DrugBank and Therapeutic Targets Database (TTD) (Liu et al., 2011). The Gene Set Functional Similarity (GSFS) method was given in the previous article (Lv et al., 2012). Specifically, we downloaded the SM chemical structure information. Then, a graph-based method, SIMilar COMPound (SIMCOMP) (Lv et al., 2012), was applied to obtain SMs' chemical structure similarities. Finally, the disease phenotype-based similarities for SMs were obtained by calculating the data downloaded from the DrugBank and TTD with the Jaccard score method.

After obtaining all four SM similarities, we named them after $SS1$, $SS2$, $SS3$ and $SS4$, respectively. Then, the scores of the four SM similarities were integrated by the following formula,

$$SSM = \frac{\sum_i \alpha_i SSi}{\sum_i \alpha_i}, (i = 1, 2, 3, 4) \tag{1}$$

where α represents the weights of SM similarities. All of the measures are important in terms of biology. Thus, we set the values of all α to 1, which means that each SM similarity made an equal contribution to constituting the integrated SM similarity (Li et al., 2004). Finally, the integrated SM similarity $SSM(s_i, s_j)$ between the i_{th} and j_{th} SMs was obtained after normalization as follows.

$$SSM\left(s_i, s_j\right) = \frac{SSM\left(s_i, s_j\right)}{\sqrt{\sum_{l=1}^{ns} SSM\left(s_i, s_l\right)}\sqrt{\sum_{l=1}^{ns} SSM\left(s_l, s_j\right)}} \tag{2}$$

2.3 Integration of miRNA similarities

Two miRNA similarities, gene function consistency-based similarity (Lv et al., 2012) and indication phenotype-based similarity (Gottlieb et al., 2011), were used to obtain

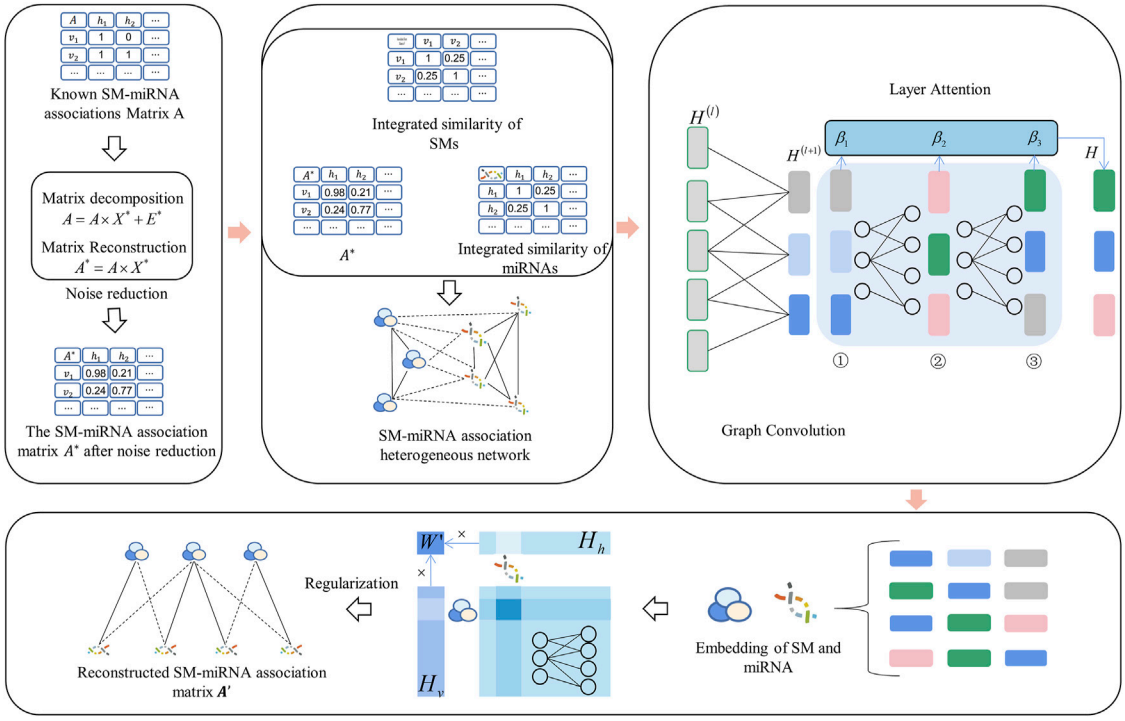

FIGURE 1
The flow chart of potential SM-miRNA association prediction based on GCNLASMMA. Firstly, the matrix decomposition is applied to obtain the new SM-miRNA associations. Then the new SM-miRNA associations, integrated SM similarity and integrated miRNA similarity are constructed into an SM-miRNA association heterogeneous network. Finally, a graph convolutional network with layer attention mechanism is applied to obtain the reconstructed SM-miRNA association matrix.

integrated miRNA similarity. Specifically, we downloaded the target scores of each miRNA from the database TargetScan (Agarwal et al., 2015) and obtained gene function consistency-based similarity using the GSFS method (Lv et al., 2012). The indication phenotype-based similarity was obtained from the Human MicroRNA Disease Database (HMDD) version 2.0 (v 2.0), miR2Disease and PhenomiR databases using the GSFS method. Then, we combined the gene function consistency-based similarity and the indication phenotype-based similarity using the Jaccard score. Then, we named the two kinds of miRNA similarities after $SM1$ and $SM2$, respectively. Moreover, the integrated miRNA similarity SMR was obtained by the following equation,

$$SMR = \frac{\sum_j \beta_j SMj}{\sum_j \beta_j}, (j = 1, 2) \tag{3}$$

where β_1 and β_2 represent the weights of miRNA similarities. Also, we set the values of β_1 and β_2 to 1, which means each miRNA similarity made an equal contribution to constituting the integrated miRNA similarity. Finally, the integrated miRNA similarity $SMR(m_i, m_j)$ between the i_{th} and j_{th} miRNAs was obtained after normalization as follows.

$$SMR(m_i, m_j) = \frac{SMR(m_i, m_j)}{\sqrt{\sum_{l=1}^{nm} SMR(m_i, m_l)} \sqrt{\sum_{l=1}^{nm} SMR(m_l, m_j)}} \tag{4}$$

2.4 GCNLASMMA

GCNLASMMA was separated into two steps. The known SM-miRNA association A was initially decomposed and reconstructed to obtain the new SM-miRNA association A^*. The reconstructed SM-miRNA association matrix A' was then obtained by calculating the new SM-miRNA association A^* using a graph convolutional network with an attention mechanism. More specifically, we obtained the new SM-miRNA associations by matrix decomposition. Then, the new SM-miRNA association matrix, integrated SM similarity and integrated miRNA similarity were constructed into a heterogeneous network. Finally, the graph convolutional network with layer attention mechanism was applied to obtain the reconstructed SM-miRNA association matrix. GCNLASMMA is a model of a neural network with more hidden layers than other networks. The multi-layer calculation thoroughly considered the known

TABLE 1 The illustration of the IALM algorithm.

Algorithm: Inexact augmented lagrange multipliers

Input: Known SM-miRNA association matrix A and $\alpha = 0.1$

Initialize:
$X = 0$, $E = 0$, $Y_1 = 0$, $Y_2 = 0$, $\mu = 10^{-4}$, $max_\mu = 10^{10}$, $\rho = 1.1$, $\varepsilon = 10^{10}$

While true

1. Fix others and $J = argmin\frac{1}{\mu}\|J\|_* + \frac{1}{2}\|J - (X + Y_2/\mu)\|_F^2$
2. Fix others and $X = (I + A^TA)(A^TA - A^TE + J + (A^TY_1 - Y_2)/\mu)$
3. Fix others and $E = argmin\frac{\alpha}{\mu}\|E\|_{2,1} + \frac{1}{2}\|E - (A - AX + Y_1/\mu)\|_F^2$
4. Update $Y_1 = Y_1 + \mu(A - AX - E)$; $Y_2 = Y_2 + \mu(X - J)$
5. Update $\mu = min(\rho\mu, max_\mu)$

If $\|A - AX - E\|_\infty < \varepsilon$ *and* $\|X - J\|_\infty < \varepsilon$

End while

Output: X^* and E^*

features and avoided overfitting. Moreover, the attention mechanism extracted significant information from each layer, thereby improving the accuracy of association prediction (Niu et al., 2021). The specific flow chart of GCNLASMMA is shown in Figure 1.

2.4.1 Matrix decomposition

The existence of noise in known SM-miRNA associations tends to reduce prediction accuracy. Prior research has demonstrated that hidden features with considerable value can be extracted by applying dimension-reduction and noise-reduction to the data (Vidal, 2011). A low-rank matrix is a tool for efficiently obtaining hidden features with significant values (Peng et al., 2012). Therefore, we used matrix decomposition to learn a low-rank matrix from the known SM-miRNA association A. The decomposition of A was performed as follows:

$$A = A \times X + E \tag{5}$$

Since the above equation contains an infinite number of solutions, we applied the constraint to turn it into:

$$\min_{X,E} \|X\|_* + \alpha\|E\|_{2,1} \; s.t. \; A = A \times X + E \tag{6}$$

where $\|X\|_* = \sum_i \sigma_i$, ($\sigma_i$ *is the singular value of matrix* X), $\|E\|_{2,1} = \sum_{j=1}^{n}\sqrt{\sum_{i=1}^{m}(E_{ij})^2}$. In Eq. 6, the nuclear norm and sparse norm were applied to constrain X and E, which allowed X and E to be low-rank and sparse matrices, respectively. The balance parameter of low-rank and sparse matrices α was set to 0.1. According to earlier research, if A in Eq. 6 is transformed into an identity matrix, then the model is degenerated to the Robust Principal Component Analysis (RPCA), a convex optimization problem with constraints (Chandrasekaran et al., 2009).

$$\min_{X,E,J} \|J\|_* + \alpha\|E\|_{2,1} \; s.t. \; A = A \times X + E, X = J \tag{7}$$

Based on the previous work (Meng et al., 2014), Eq. 7 can be converted into an unconstrained optimization problem. Therefore, the problem can be resolved using the Exact Augmented Lagrange Multipliers (EALM) algorithm.

$$L = \|J\|_* + \alpha\|E\|_{2,1} + tr\left(Y_1^T(A - A \times X - E)\right) + tr\left(Y_2^T(X - J)\right) + \frac{\delta}{2}\left(\|A - A \times X - E\|_F^2 + \|X - J\|_F^2\right) \tag{8}$$

In Eq. 8, the penalty parameter $\delta \geq 0$. According to the Inexact Augmented Lagrange Multipliers (IALM) algorithm (See Table 1), we fixed other variables and solved the minimum value of J, X and E by updating the Lagrange multipliers Y_1 and Y_2. Moreover, we defined X^* and E^* as the solution of Eq. 8. X^* represents the similarity matrix of miRNA or SM. E^* represents the noise matrix. Then, the new SM-miRNA association A^* was expressed as:

$$A^* = A \times X^* \tag{9}$$

2.4.2 SM-miRNA heterogeneous network

In this study, the new SM-miRNA association A^*, integrated SM similarity SSM and integrated miRNA similarity SMR were combined into a heterogeneous network. There would be a known association between the i_{th} SM and the j_{th} miRNA if element A^*_{ij} in A^* equaled 1. $SSM(i, j)$ represented the integrated similarity between the i_{th} SM and the j_{th} SM. $SMR(i, j)$ represented the integrated similarity between the i_{th} miRNA and the j_{th} miRNA. The specific equation of the heterogeneous network A_H construction is as follows:

$$A_H = \begin{bmatrix} \sim SMR & A^* \\ A^{*T} & \sim SSM \end{bmatrix} \tag{10}$$

where A^{*T} represents the transpose matrix of A^*. In Eq. 10, we normalized the similarity matrix of SM and miRNA by $\sim SSM = D_s^{-\frac{1}{2}}SSMD_s^{-\frac{1}{2}}$ and $\sim SMR = D_m^{-\frac{1}{2}}SMRD_m^{-\frac{1}{2}}$, respectively. Specifically, $D_s = diag(\sum_j SSM_{ij})$ and $D_m = diag(\sum_j SMR_{ij})$.

2.4.3 Graph convolutional network

As classic network models, Long-Short Term Memory (LSTM) and Convolution Neural Network (CNN) are only applicable to grid-structured data. Nevertheless, the Graph Convolutional Network (GCN) can manage data with generalized topological graph structures and deeply explore the features of the data (Habib and Qureshi, 2020). In this paper, we constructed GCNLASMMA, which is a model for graph convolution of biological information. Specifically, GCN was implemented on the SM-miRNA heterogeneous network A_H that was constructed by the known SM-miRNA associations, SM similarities and miRNA similarities. GCN is a neural network

TABLE 2 Validation of the random 50 SM-miRNAs associations. The first column records the random 1–25 associations. The second column records the random 26–50 associations.

SM	miRNA	Evidence	SM	miRNA	Evidence
CID 4116	hsa-mir-329-2	unconfirmed	CID 2662	hsa-mir-330	unconfirmed
CID 60726	hsa-mir-216b	unconfirmed	CID 7028	hsa-mir-592	unconfirmed
CID 4760	hsa-mir-520c	unconfirmed	CID 5656	hsa-mir-646	32083545
CID 3052	hsa-mir-193a	unconfirmed	CID 3520	hsa-mir-1266	unconfirmed
CID 444036	hsa-mir-199a-2	unconfirmed	CID 43008	hsa-mir-519a-1	unconfirmed
CID 3198	hsa-mir-216a	unconfirmed	CID 3343	hsa-mir-1469	unconfirmed
CID 157922	hsa-mir-1260a	unconfirmed	CID 5566	hsa-mir-548a-3	unconfirmed
CID 3698	hsa-mir-2110	unconfirmed	CID 5493444	hsa-mir-1285-2	unconfirmed
CID 4212	hsa-mir-219-2	unconfirmed	CID 60843	hsa-let-7d	unconfirmed
CID 8223	hsa-mir-98	unconfirmed	CID 110635	hsa-mir-216b	unconfirmed
CID 19861	hsa-mir-659	unconfirmed	CID 2801	hsa-mir-744	unconfirmed
CID 71329	hsa-mir-100	unconfirmed	CID 216239	hsa-mir-1273e	unconfirmed
CID 47641	hsa-mir-150	unconfirmed	CID 71398	hsa-mir-526a-1	unconfirmed
CID 443980	hsa-mir-760	unconfirmed	CID 4201	hsa-mir-153-2	unconfirmed
CID 5574	hsa-mir-512-2	unconfirmed	CID 5281040	hsa-mir-548a-2	unconfirmed
CID 8969	hsa-mir-543	unconfirmed	CID 444020	hsa-mir-320a	unconfirmed
CID 5282415	hsa-mir-619	unconfirmed	CID 3025	hsa-mir-24-1	unconfirmed
CID 65833	hsa-mir-760	unconfirmed	CID 3019	hsa-mir-1226	unconfirmed
CID 1775	hsa-mir-520f	unconfirmed	CID 1125	hsa-mir-27a	unconfirmed
CID 3749	hsa-mir-1285-2	unconfirmed	CID 1349907	hsa-mir-642a	unconfirmed
CID 2905	hsa-mir-96	unconfirmed	CID 656719	hsa-mir-611	unconfirmed
CID 3180	hsa-mir-148a	unconfirmed	CID 2795	hsa-mir-711	unconfirmed
CID 5566	hsa-mir-646	unconfirmed	CID 23994	hsa-mir-614	unconfirmed
CID 4212	hsa-mir-18a	31063487	CID 4099	hsa-mir-708	unconfirmed
CID 82146	hsa-mir-490	unconfirmed	CID 5281106	hsa-mir-1302-6	unconfirmed

structure consisting of an input layer, an output layer and many hidden layers that can represent nodes in a low-dimensional manner. Each hidden layer of GCN takes the output of the previous layer as input. The graph convolutional network propagation rule is as follows:

$$H^{(l+1)} = f\left(H^{(l)}, G\right) = \sigma\left(D^{-\frac{1}{2}}GD^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{11}$$

In Eq. 11, $H^{(l)}$ and $H^{(l+1)}$ denote the embeddings of nodes in the l_{th} and $(l+1)_{th}$ layers, respectively. $D = diag\left(\sum_j G_{ij}\right)$ is a diagonal matrix of input graph G, $W^{(l)}$ represents the trainable weight matrix with a layer-specific value, $\sigma(\cdot)$ denotes the nonlinear activation function.

In the encoder part, to learn low-dimensional representations of miRNAs and SMs, we combined the new SM-miRNA association, integrated SM similarity and integrated miRNA similarity into SM-miRNA association heterogeneous network A_H. Firstly, we set a penalty factor μ in the input graph G during the propagation process as follows:

$$G = \begin{bmatrix} \mu \sim SMR & A^{*} \\ A^{*T} & \mu \sim SSM \end{bmatrix} \tag{12}$$

Then, we initialized the input layer embeddings as:

$$H^{(0)} = \begin{bmatrix} 0 & A^{*} \\ A^{*T} & 0 \end{bmatrix} \tag{13}$$

In this way, we obtained the propagation formula for the first layer from Eqs 11, 13:

$$H^{(1)} = \sigma\left(D^{-\frac{1}{2}}GD^{-\frac{1}{2}}H^{(0)}W^{(0)}\right) \tag{14}$$

In Eq. 12, $W^{(0)}$ is a weight matrix that acts only between the input layer and the first hidden layer. $H^{(1)}$ is the first-layer embeddings of the heterogeneous network A_H, k is the dimension of the embeddings. Similarly, the propagation rules for the subsequent layers of the GCN encoder followed Eq. 11, where $l = 1, 2, \cdots, L$. After L iterations, we obtained L $k-$ *dimensional* embeddings from different graph convolution

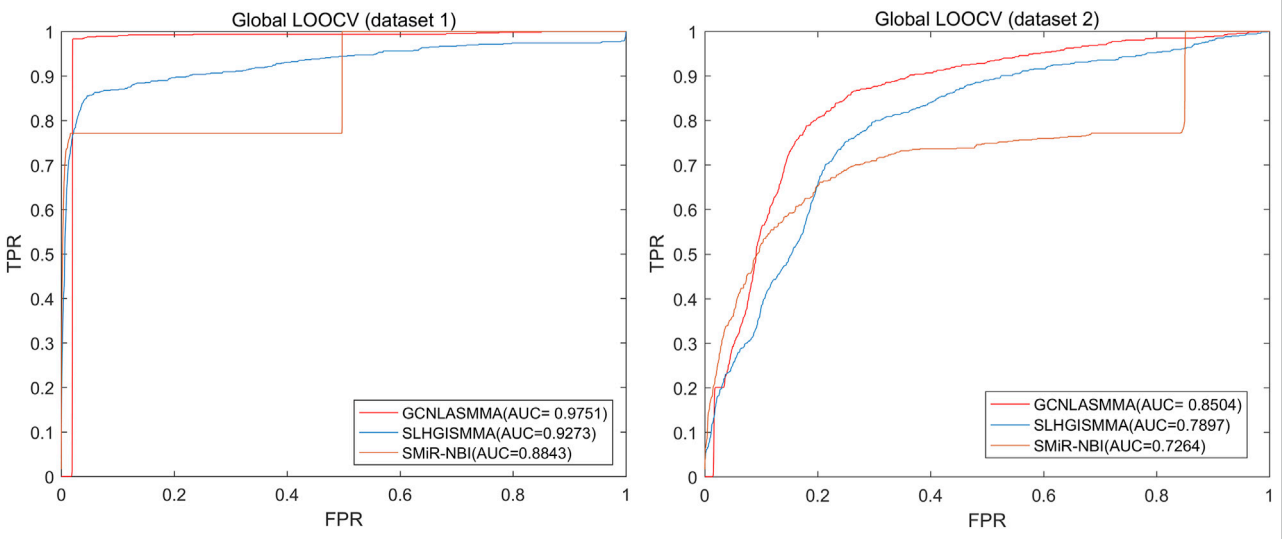

FIGURE 2
The left half of the figure shows the comparison of GCNLASMMA with two comparison algorithms under global LOOCV based on dataset1. As a result, GCNLASMMA, SLHGISMMA and SMiR-NBI achieve AUCs of 0.9751, 0.9273 and 0.8843, respectively. The right half of the figure shows the comparison of GCNLASMMA with two comparison algorithms under global LOOCV based on dataset2. As a result, GCNLASMMA, SLHGISMMA and SMiR-NBI achieve AUCs of 0.8504, 0.7897 and 0.7264, respectively.

layers. Exponential linear elements were used as nonlinear activation functions in the graph convolution layer, which sped up the learning process and significantly improved the generalization performance.

In addition, we tried several different combinations of parameters from the range $\alpha \in \{400, 600, 800, 1000\}$, $lr \in \{0.00700, 0.00725, 0.00750, 0.00775, 0.00800\}$. By adjusting the parameters empirically, we set the dimensions of embeddings $k = 64$, the number of layers $L = 3$, the initial learning rate of optimizer $lr = 0.00725$, the total training epochs $\alpha = 600$, the two dropout rates $\beta = 0.6$ and $\gamma = 0.4$, the penalty factor $\mu = 6$ on both dataset1 and dataset2.

2.4.4 Layer attention mechanism

In addition, the layer attention mechanism was added to this model by introducing an attention mechanism between each layer and storing the position information in A_H. As a help for the attention mechanism, we extracted the pertinent information straight from the source data when constructing the embeddings of each layer output during the decoding process. Through this mechanism, we obtained the final SM embeddings and final miRNA embeddings from the fully connected layer: $\begin{bmatrix} H_m \\ H_s \end{bmatrix} = \sum a_l H^l$, where H_m represents the final embeddings of miRNA, H_s is the final embeddings of SM. The neural network automatically adjusted the value of a_l by the initial input value $\frac{l}{(l+1)}, l = 1, 2, \cdots, L$. Finally, we obtained the reconstructed SM-miRNA association matrix A' by an activation function as follows,

$$A' = sigmoid\left(H_m W' H_s^T\right) \tag{15}$$

where W' is a trainable matrix. The corresponding element A'_{ij} is the potential correlation score between miRNA m_i and SM s_j.

3 Results

To evaluate the performance of GCNLASMMA, we used two types of cross validation, namely, 5-fold cross-validation and Leave-One-Out Cross Validation (LOOCV). The two different datasets include the same known 664 SM-miRNA associations. Specifically, dataset 1 has 831 SMs and 541 miRNAs. On the basis of dataset1, we removed the SMs and miRNAs that did not constitute any known association. Then, we obtained dataset2 which has only 286 different miRNAs, 39 different SMs. In this study, the Area Under the receiver operating characteristic Curves (AUCs) obtained under 5-fold cross-validation based on dataset1 and dataset2 were 0.9721 ± 0.0018 and 0.8393 ± 0.0047, respectively. The global AUC and local AUC obtained under LOOCV based on dataset1 were 0.9751 (global LOOCV), 0.9746 (miRNA-fixed LOOCV) and 0.5014 (SM-fixed LOOCV), respectively. Based on dataset2, the AUCs of GCNLASMMA were 0.8504 (global LOOCV), 0.8490 (miRNA-fixed LOOCV) and 0.6398 (SM-fixed LOOCV), respectively. Additionally, we utilized two types of case studies to confirm the effectiveness of GCNLASMMA in identifying

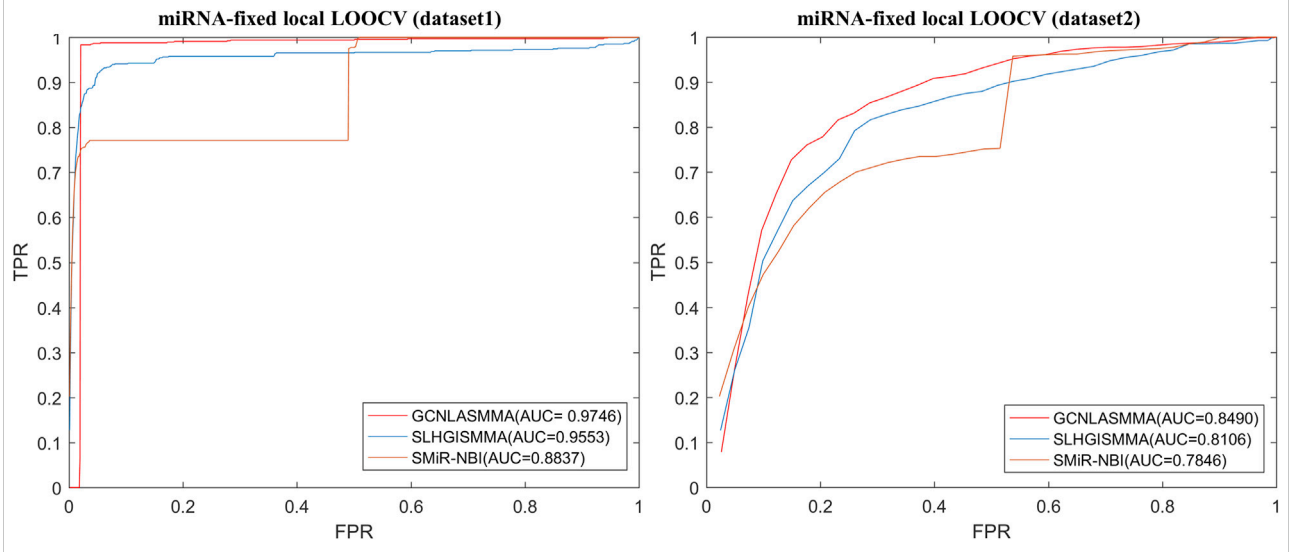

FIGURE 3
The left half of the figure shows the comparison of GCNLASMMA with two comparison algorithms under miRNA-fixed LOOCV based on dataset1. As a result, GCNLASMMA, SLHGISMMA and SMiR-NBI achieve AUCs of 0.9746, 0.9553 and 0.8837, respectively. The right half of the figure shows the comparison of GCNLASMMA with two comparison algorithms under miRNA-fixed LOOCV based on dataset2. As a result, GCNLASMMA, SLHGISMMA and SMiR-NBI achieve AUCs of 0.8490, 0.8106 and 0.7846, respectively.

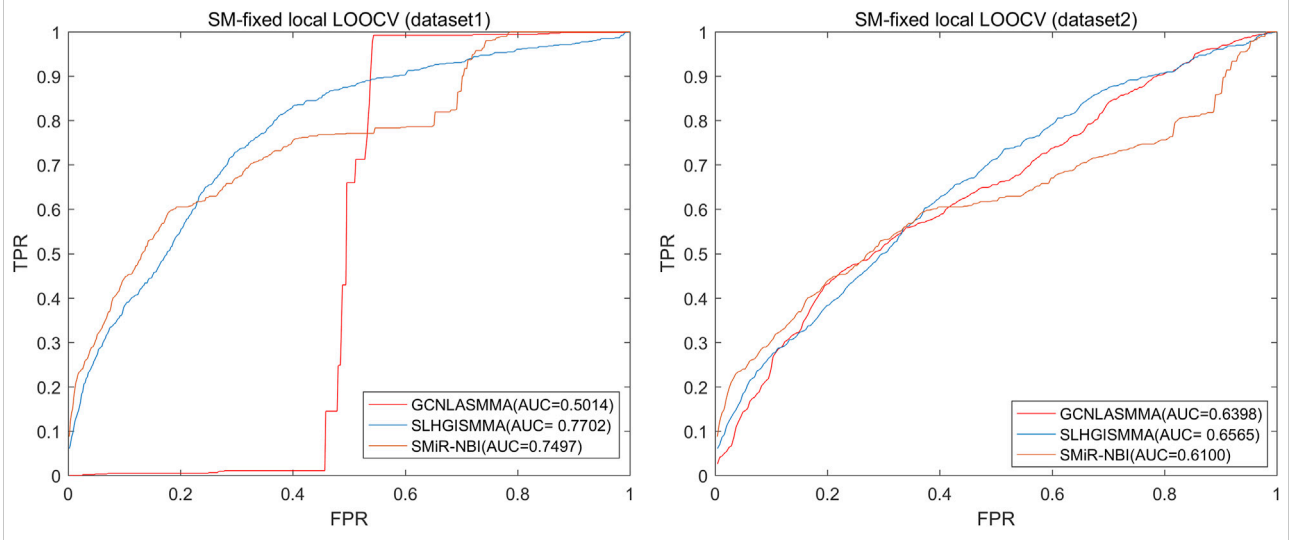

FIGURE 4
The left half of the figure shows the comparison of GCNLASMMA with two comparison algorithms under SM-fixed LOOCV based on dataset1. As a result, GCNLASMMA, SLHGISMMA and SMiR-NBI achieve AUCs of 0.5014, 0.7702 and 0.7497, respectively. The right half of the figure shows the comparison of GCNLASMMA with two comparison algorithms under SM-fixed LOOCV based on dataset2. As a result, GCNLASMMA, SLHGISMMA and SMiR-NBI achieve AUCs of 0.6398, 0.6565 and 0.6100, respectively.

potential miRNAs for investigated SMs. Specifically, GCNLASMMA has predicted the potential miRNAs associated with 5-Fluorouracil (5-Fu, CID: 3385), 5-Aza-2′-deoxycytidine (5-Aza-CdR, CID: 451668) and 17β-Estradiol (E2, CID: 5757). For 5-Fu, the results showed that 9, 16 and 39 out of the top 10, 20 and 50 potential related miRNAs in the first type of case studies, 8, 15 and 39 out of the top 10, 20 and 50 potential related miRNAs in the second type of case studies were validated in other

10.3389/fmolb.2022.1009099

TABLE 3 Validation of the top 50 miRNAs associated with 5-Fu in the first type of case studies. The first column records the top 1–25 related miRNAs. The second column records the top 26–50 related miRNAs.

SM	miRNA	Evidence	SM	miRNA	Evidence
CID 3385	hsa-miR-151a	23220571	CID 3385	hsa-miR-126	26062749
CID 3385	hsa-miR-195	21947305	CID 3385	hsa-miR-128-1	23220571
CID 3385	hsa-let-7d	23220571	CID 3385	hsa-miR-337	unconfirmed
CID 3385	hsa-miR-195	21947305	CID 3385	hsa-miR-181c	unconfirmed
CID 3385	hsa-miR-125a	23220571	CID 3385	hsa-miR-30c-1	unconfirmed
CID 3385	hsa-miR-345	unconfirmed	CID 3385	hsa-miR-27a	23220571
CID 3385	hsa-miR-16-1	26198104	CID 3385	hsa-let-7a-1	23220571
CID 3385	hsa-miR-24-1	26198104	CID 3385	hsa-miR-139	27173050
CID 3385	hsa-miR-23b	23220571	CID 3385	hsa-miR-302b	26457704
CID 3385	hsa-miR-1226	26198104	CID 3385	hsa-let-7b	25789066
CID 3385	hsa-miR-151a	23220571	CID 3385	hsa-miR-26b	23220571
CID 3385	hsa-miR-132	23220571	CID 3385	hsa-miR-221	27501171
CID 3385	hsa-125b-1	unconfirmed	CID 3385	hsa-miR-338	28928082
CID 3385	hsa-let-7e	23220571	CID 3385	hsa-miR-130a	unconfirmed
CID 3385	hsa-miR-19a	23220571	CID 3385	hsa-miR-10b	22322955
CID 3385	hsa-miR-181a-1	unconfirmed	CID 3385	hsa-miR-204	27095441
CID 3385	hsa-miR-181b-1	unconfirmed	CID 3385	hsa-miR-26a-1	unconfirmed
CID 3385	hsa-miR-25	23220571	CID 3385	hsa-miR-92a-1	23220571
CID 3385	hsa-miR-106a	23220571	CID 3385	hsa-miR-299	31786874
CID 3385	hsa-miR-200c	23220571	CID 3385	hsa-miR-107	26636340
CID 3385	hsa-miR-22	25449431	CID 3385	hsa-miR-181a-2	24462870
CID 3385	hsa-miR-20a	23220571	CID 3385	hsa-miR-205	24396484
CID 3385	hsa-let-7d	23220571	CID 3385	hsa-miR-23a	23220571
CID 3385	hsa-miR-34b	unconfirmed	CID 3385	hsa-miR-199b	unconfirmed
CID 3385	hsa-miR-205	24396484	CID 3385	hsa-miR-93	23220571

literature or databases, respectively. For 5-Aza-CdR, the results showed that 8, 13 and 26 out of the top 10, 20 and 50 potential related miRNAs in the first type of case studies, 8, 14 and 28 out of the top 10, 20 and 50 potential related miRNAs in the second type of case studies were validated in other literature or databases, respectively. For E2, the results showed that 6, 14 and 29 out of the top 10, 20 and 50 potential related miRNAs in the first type of case studies, 4, 11 and 29 out of the top 10, 20 and 50 potential related miRNAs in the second type of case studies were validated in other literature or databases, respectively.

3.1 Performance evaluation

In 5-fold cross-validation, all known SM-miRNA associations were randomly separated into five subsets of nearly comparable size. Then, each subset was in turn considered as the test sample, and the rest four subsets were treated as training samples. Moreover, all unknown SM-miRNA pairs were regarded as candidate samples. Subsequently, we obtained a predicted association score matrix by GCNLASMMA, and ranked the scores of each test sample against those of the candidate samples. This partition-prediction-ranking procedure was repeated 100 times to obtain a sound estimate of the mean and variance of GCNLASMMA's prediction accuracy. Finally, the prediction of a test sample was deemed successful if the sample's rank was higher than the given threshold. Therefore, we utilized the threshold to calculated the false positive rate (FPR, 1-specificity) and the true positive rate (TPR, sensitivity). The FPR and TPR represented the percentage of candidate samples that lower than the threshold and the percentage of test samples that higher than the threshold, respectively. Then, we regarded FPR and TPR as horizontal and vertical axis. The Receiver Operating Characteristic (ROC) curve were plotted. Finally, we attained the Area Under the ROC Curve (AUC) by computing the area under the ROC curves. In this investigation, GCNLASMMA achieved the AUCs of 0.9721 ± 0.0018 and 0.8393 ± 0.0047 under 5-fold cross-validation based on dataset1 and dataset2, respectively.

LOOCV was further classified as either global and local. Then, the local-LOOCV was subdivided into miRNA-fixed

TABLE 4 Validation of the top 50 miRNAs associated with 5-Fu in the second type of case studies. The first column records the top 1–25 related miRNAs. The second column records the top 26–50 related miRNAs.

SM	miRNA	Evidence	SM	miRNA	Evidence
CID 3385	hsa-miR-151a	23220571	CID 3385	hsa-miR-195	21947305
CID 3385	hsa-let-7d	23220571	CID 3385	hsa-miR-27a	23220571
CID 3385	hsa-miR-205	24396484	CID 3385	hsa-miR-204	27095441
CID 3385	hsa-miR-181a-2	24462870	CID 3385	hsa-miR-181a-1	unconfirmed
CID 3385	hsa-miR-23a	23220571	CID 3385	hsa-miR-25	23220571
CID 3385	hsa-miR-1226	26198104	CID 3385	hsa-miR-199b	unconfirmed
CID 3385	hsa-miR-181c	unconfirmed	CID 3385	hsa-miR-139	27173050
CID 3385	hsa-miR-151a	23220571	CID 3385	hsa-miR-195	21947305
CID 3385	hsa-miR-26a-1	unconfirmed	CID 3385	hsa-miR-132	23220571
CID 3385	hsa-miR-26b	23220571	CID 3385	hsa-miR-20a	23220571
CID 3385	hsa-miR-130a	unconfirmed	CID 3385	hsa-miR-126	26062749
CID 3385	hsa-miR-345	unconfirmed	CID 3385	hsa-125b-1	unconfirmed
CID 3385	hsa-miR-128-1	23220571	CID 3385	hsa-miR-200c	23220571
CID 3385	hsa-let-7d	23220571	CID 3385	hsa-miR-299	31786874
CID 3385	hsa-miR-181b-1	unconfirmed	CID 3385	hsa-miR-30c-1	unconfirmed
CID 3385	hsa-miR-205	24396484	CID 3385	hsa-miR-24-1	26198104
CID 3385	hsa-miR-125a	23220571	CID 3385	hsa-miR-93	23220571
CID 3385	hsa-miR-22	25449431	CID 3385	hsa-let-7e	23220571
CID 3385	hsa-miR-16-1	26198104	CID 3385	hsa-let-7b	25789066
CID 3385	hsa-miR-106a	23220571	CID 3385	hsa-miR-221	27501171
CID 3385	hsa-miR-23b	23220571	CID 3385	hsa-miR-19a	23220571
CID 3385	hsa-miR-338	28928082	CID 3385	hsa-miR-92a-1	23220571
CID 3385	hsa-miR-10b	22322955	CID 3385	hsa-miR-302b	26457704
CID 3385	hsa-let-7a-1	23220571	CID 3385	hsa-miR-107	26636340
CID 3385	hsa-miR-337	unconfirmed	CID 3385	hsa-miR-34b	unconfirmed

LOOCV and SM-fixed LOOCV. In LOOCV, each known SM-miRNA association was in turn considered to be the test sample and the others were treated as the training samples. Moreover, all unknown SM-miRNA pairs were treated as candidate samples. In miRNA-fixed LOOCV and SM-fixed LOOCV, test samples and training samples were chosen similarly. However, in SM-fixed LOOCV, only unknown SM-miRNA pairs containing the selected SM were regarded as candidate samples. Similarly, in miRNA-fixed LOOCV, candidate samples only included those involving the chosen miRNA. Then, we ranked the score of the test sample against those of the candidate samples. Finally, the prediction of a test sample was deemed successful if the rank of this test sample was higher than the given threshold. Based on dataset1, GCNLASMMA attained the AUCs of 0.9751, 0.9746 and 0.5014 under global LOOCV, miRNA-fixed LOOCV and SM-fixed LOOCV, respectively. Based on dataset2, GCNLASMMA attained the AUCs of 0.8504, 0.8490 and 0.6398 under global LOOCV, miRNA-fixed LOOCV and SM-fixed LOOCV, respectively.

The AUC comparison figures based on dataset1 (dataset2) were plotted to determine the differences between GCNLASMMA and other models' outcomes. AUC = 0.5 would suggest that the model was only capable of random prediction, whereas AUC = 1 would indicate that all test samples were accurately predicted. Figure 2 demonstrates that the results of GCNLASMMA under global LOOCV are significantly better than that of SMiR-NBI. Figures 3, 4 show that the results of GCNLASMMA under miRNA-fixed local LOOCV and SM-fixed local LOOCV were significantly better than those of SLHGISMMA and SMiR-NBI. Furthermore, the AUC of miRNA-fixed local LOOCV based on dataset1 is 0.9746, which means almost all potential SM-miRNA associations in dataset1 were predicted successfully.

3.2 Case studies

To further illustrate the GCNLASMMA's applicability to identify potential miRNAs, we conducted two types of case studies on three essential SMs, namely 5-Fluorouracil (5-Fu, CID: 3385), 5-Aza-2′-deoxycytidine (5-Aza-CdR, CID:

10.3389/fmolb.2022.1009099

TABLE 5 Validation of the top 50 miRNAs associated with 5-Aza-CdR in the first type of case studies. The first column records the top 1–25 related miRNAs. The second column records the top 26–50 related miRNAs.

SM	miRNA	Evidence	SM	miRNA	Evidence
CID 451668	hsa-miR-20a	23220571	CID 451668	hsa-miR-30a	unconfirmed
CID 451668	hsa-miR-320a	26198104	CID 451668	hsa-miR-107	23220571
CID 451668	hsa-miR-125a	23220571	CID 451668	hsa-miR-199b	24659709
CID 451668	hsa-miR-182	23220571	CID 451668	hsa-let-7a-1	unconfirmed
CID 451668	hsa-miR-204	unconfirmed	CID 451668	hsa-miR-92a-1	unconfirmed
CID 451668	hsa-miR-200b	23626803	CID 451668	hsa-miR-181a-1	23220571
CID 451668	hsa-miR-23a	unconfirmed	CID 451668	hsa-let-7e	22053057
CID 451668	hsa-let-7f-1	23220571	CID 451668	hsa-miR-26a-1	unconfirmed
CID 451668	hsa-let-7b	26708866	CID 451668	hsa-miR-1233-1	unconfirmed
CID 451668	hsa-miR-200c	23626803	CID 451668	hsa-miR-130a	23220571
CID 451668	hsa-miR-25	23220571	CID 451668	hsa-miR-30c-1	unconfirmed
CID 451668	hsa-miR-128-1	27705931	CID 451668	hsa-miR-22	23220571
CID 451668	hsa-miR-145	26198104	CID 451668	hsa-miR-301a	unconfirmed
CID 451668	hsa-miR-221	unconfirmed	CID 451668	hsa-let-7g	23220571
CID 451668	hsa-miR-19b-1	unconfirmed	CID 451668	hsa-miR-195	23333942
CID 451668	hsa-miR-197	unconfirmed	CID 451668	hsa-miR-302b	unconfirmed
CID 451668	hsa-let-7i	23220571	CID 451668	hsa-miR-26b	unconfirmed
CID 451668	hsa-miR-181b-1	unconfirmed	CID 451668	hsa-miR-205	unconfirmed
CID 451668	hsa-miR-338	unconfirmed	CID 451668	hsa-miR-218-1	unconfirmed
CID 451668	hsa-let-7d	26802971	CID 451668	hsa-miR-93	23220571
CID 451668	hsa-miR-139	unconfirmed	CID 451668	hsa-miR-124-1	unconfirmed
CID 451668	hsa-miR-328	unconfirmed	CID 451668	hsa-miR-15b	unconfirmed
CID 451668	hsa-miR-126	23220571	CID 451668	hsa-miR-10b	unconfirmed
CID 451668	hsa-miR-17	23220571	CID 451668	hsa-miR-128-2	unconfirmed
CID 451668	hsa-miR-19a	23220571	CID 451668	hsa-miR-27a	23220571

451668) and 17β-Estradiol (E2, CID: 5757). On the basis of all known SM-miRNA associations, the first type was applied to forecast potential miRNAs for investigated SMs. As the training set, we utilized the known SM-miRNA associations from dataset1. Then, for each investigated SM, we ranked all candidate miRNAs according on their predicted scores. The second type was used to forecast potential miRNAs for investigated SMs without any known SM-miRNA association. Therefore, we removed all verified associations related to the investigated SMs before the prediction and ranked them as the first type of case studies. After ranking all candidate miRNAs for each investigated SM based on their predicted scores, the top 50 predicted miRNAs were picked out and verified in other literature or databases. Moreover, we selected 10, 20 and 50 associations randomly from all potential associations to further demonstrate the validity of GCNLASMMA. The results show that only 0, 0 and 2 out of random 10, 20 and 50 associations are confirmed in other literature or databases (See Table 2), which significantly worse than the top 10, 20 and 50 miRNAs related to investigated SMs.

3.2.1 5-Fu

5-Fu, one of the earliest anticancer drugs, can be fully absorbed by tumor cells. Moreover, 5-Fu can decrease tumor cell proliferation by interfering with the formation of DeoxyriboNucleic Acid (DNA) and RiboNucleic Acid (RNA) in tumor cells. It has been demonstrated that 5-Fu has considerable inhibitory effects on various cancer cells. Therefore, 5-Fu is frequently used as a positive control in anticancer drug effect experiments and clinical adjuvant treatment of gastric cancer (Longley et al., 2003). The first type of case studies' results show that 9, 16 and 39 out of the top 10, 20 and 50 potential 5-Fu-associated miRNAs are confirmed in other literature or databases (See Table 3). The second type of case studies' results show that 8, 15 and 39 out of the top 10, 20 and 50 potential 5-Fu-associated miRNAs are confirmed in other literature or databases (See Table 4). For example, 5-Fu is the most common chemotherapeutic agent for colorectal cancer. On the one hand, over-expression of hsa-miR-23a causes the resistance to 5-Fu in microsatellite instability colorectal cancer, which results in a diminished effect of 5-Fu chemotherapy (Shang et al., 2014). On the other hand, Ectopic expression of hsa-miR-23a increased the viability and survival of microsatellite stability

TABLE 6 Validation of the top 50 miRNAs associated with 5-Aza-CdR in the second type of case studies. The first column records the top 1–25 related miRNAs. The second column records the top 26–50 related miRNAs.

SM	miRNA	Evidence	SM	miRNA	Evidence
CID 451668	hsa-miR-20a	23220571	CID 451668	hsa-miR-92a-1	unconfirmed
CID 451668	hsa-miR-181b-1	unconfirmed	CID 451668	hsa-miR-125a	23220571
CID 451668	hsa-miR-205	unconfirmed	CID 451668	hsa-let-7b	26708866
CID 451668	hsa-miR-19a	23220571	CID 451668	hsa-miR-302b	unconfirmed
CID 451668	hsa-miR-181a-1	23220571	CID 451668	hsa-miR-30a	unconfirmed
CID 451668	hsa-miR-130a	23220571	CID 451668	hsa-miR-23b	23220571
CID 451668	hsa-let-7g	23220571	CID 451668	hsa-miR-199b	24659709
CID 451668	hsa-miR-200b	23626803	CID 451668	hsa-miR-128-2	unconfirmed
CID 451668	hsa-miR-126	23220571	CID 451668	hsa-miR-15b	unconfirmed
CID 451668	hsa-miR-320a	26198104	CID 451668	hsa-miR-124-1	unconfirmed
CID 451668	hsa-miR-30c-1	unconfirmed	CID 451668	hsa-miR-26b	unconfirmed
CID 451668	hsa-miR-328	unconfirmed	CID 451668	hsa-miR-128-1	27705931
CID 451668	hsa-let-7e	22053057	CID 451668	hsa-let-7a-1	unconfirmed
CID 451668	hsa-miR-10b	unconfirmed	CID 451668	hsa-miR-218-1	unconfirmed
CID 451668	hsa-let-7f-1	23220571	CID 451668	hsa-miR-200c	23626803
CID 451668	hsa-miR-221	unconfirmed	CID 451668	hsa-miR-26a-1	unconfirmed
CID 451668	hsa-miR-182	23220571	CID 451668	hsa-miR-338	unconfirmed
CID 451668	hsa-let-7i	23220571	CID 451668	hsa-miR-93	23220571
CID 451668	hsa-miR-195	23333942	CID 451668	hsa-miR-139	unconfirmed
CID 451668	hsa-miR-27a	23220571	CID 451668	hsa-miR-145	26198104
CID 451668	hsa-miR-204	unconfirmed	CID 451668	hsa-miR-107	23220571
CID 451668	hsa-miR-25	23220571	CID 451668	hsa-let-7d	26802971
CID 451668	hsa-miR-23a	unconfirmed	CID 451668	hsa-miR-19b-1	unconfirmed
CID 451668	hsa-let-7f-1	23220571	CID 451668	hsa-miR-22	23220571
CID 451668	hsa-miR-17	23220571	CID 451668	hsa-miR-197	unconfirmed

colorectal cancer cells, thereby leading to the apoptosis of colorectal cancer cells (Li et al., 2015).

3.2.2 5-Aza-CdR

5- Aza-CdR can bind to DNA methyltransferases to reduce methylation levels, reducing the biological activity of methyltransferase inhibitors and regulating gene expression. In clinical usage, 5-Aza-CdR is frequently used in clinical settings to treat diseases caused by gene variants (Do Amaral et al., 2019). Additionally, 5-Aza-CdR can suppress tumor cell proliferation via demethylation, making it one of the most potent inhibitors currently available *in vitro* (Lemaire et al., 2008). Meanwhile, 5-Aza-CdR can enhance the sensitivity of targeted drugs in non-small cell lung cancer chemotherapy, inhibit cell proliferation, accelerate the apoptosis of cancer cells, induce cell differentiation and activate quiescent anticancer cells in the human body. The first type of case studies' results show that 8, 13 and 26 out of the top 10, 20 and 50 potential 5-Aza-CdR-associated miRNAs are confirmed in other literature or databases (See Table 5). The second type of case studies' results show that 8, 14 and 28 out of the top 10, 20 and 50 potential 5-Aza-CdR-associated miRNAs are confirmed in other literature or databases (See Table 6). For example, quantitative methylation-specific Polymerase Chain Reaction analysis showed hypermethylation of the choline phosphoglyceride island adjacent to hsa-let-7e, and demethylation treatment with 5-Aza-CdR or transfection of pYr-let-7e-shRNA plasmid containing unmethylated hsa-let-7e DNA sequence could restore hsa-let-7e expression and partly reduce the chemoresistance (Cai et al., 2013).

3.2.3 E2

In addition to stimulating the growth and maintenance of the reproductive system, E2 exerts protective effects on cardiovascular and other organs. Specifically, E2 can reduce blood cholesterol levels by decreasing Low-Density Lipoprotein (LDL), increasing High-Density Lipoprotein (HDL) and boosting apolipoprotein content (Oh et al., 2019). Moreover, researchers are paying more attention to the anti-inflammatory, antioxidant and anti-apoptotic properties of E2 on cardiovascular diseases such as coronary heart disease and atherosclerosis, are getting more attention from researchers (Tse

TABLE 7 Validation of the top 50 miRNAs associated with E2 in the first type of case studies. The first column records the top 1–25 related miRNAs. The second column records the top 26–50 related miRNAs.

SM	miRNA	Evidence	SM	miRNA	Evidence
CID 5757	hsa-miR-183	unconfirmed	CID 5757	hsa-miR-181b-1	unconfirmed
CID 5757	hsa-let-7g	23220571	CID 5757	hsa-miR-19b-1	unconfirmed
CID 5757	hsa-miR-181a-2	unconfirmed	CID 5757	hsa-miR-141	unconfirmed
CID 5757	hsa-miR-125a	21914226	CID 5757	hsa-miR-15a	unconfirmed
CID 5757	hsa-miR-107	23220571	CID 5757	hsa-miR-17	23220571
CID 5757	hsa-miR-26b	24735615	CID 5757	hsa-miR-10b	23220571
CID 5757	hsa-miR-19a	29416771	CID 5757	hsa-miR-30a	29331043
CID 5757	hsa-miR-195	unconfirmed	CID 5757	hsa-let-7f-1	23220571
CID 5757	hsa-miR-128-2	23220571	CID 5757	hsa-miR-302b	23220571
CID 5757	hsa-miR-181a-1	unconfirmed	CID 5757	hsa-miR-199b	unconfirmed
CID 5757	hsa-miR-128-1	23220571	CID 5757	hsa-miR-181c	unconfirmed
CID 5757	hsa-miR-130a	unconfirmed	CID 5757	hsa-miR-106b	28422740
CID 5757	hsa-miR-338	22996663	CID 5757	hsa-miR-23a	23220571
CID 5757	hsa-let-7e	23220571	CID 5757	hsa-miR-9-2	23220571
CID 5757	hsa-miR-20a	21914226	CID 5757	hsa-miR-182	28678802
CID 5757	hsa-miR-200c	23220571	CID 5757	hsa-miR-139	unconfirmed
CID 5757	hsa-miR-27a	23220571	CID 5757	hsa-let-7b	23220571
CID 5757	hsa-miR-200b	23220571	CID 5757	hsa-miR-25	unconfirmed
CID 5757	hsa-miR-221	21057537	CID 5757	hsa-miR-218-1	unconfirmed
CID 5757	hsa-miR-151a	unconfirmed	CID 5757	hsa-miR-22	24715036
CID 5757	hsa-miR-204	29789714	CID 5757	hsa-miR-15b	23220571
CID 5757	hsa-miR-106a	unconfirmed	CID 5757	hsa-miR-130a	unconfirmed
CID 5757	hsa-miR-205	unconfirmed	CID 5757	hsa-miR-23b	23220571
CID 5757	hsa-miR-92a-1	unconfirmed	CID 5757	hsa-miR-26a-1	unconfirmed
CID 5757	hsa-miR-130b	unconfirmed	CID 5757	hsa-miR-30c-1	23220571

et al., 1999; Rachoń et al., 2002). The first type of case studies' results show that 6, 14 and 29 out of the top 10, 20 and 50 potential E2-associated miRNAs are confirmed in other literature or databases (See Table 7). The second type of case studies' results show that 4, 11 and 29 out of the top 10, 20 and 50 potential E2-associated miRNAs are confirmed in other literature or databases (See Table 8). For example, hsa-miR-23a could be negatively regulated by E2 in both myocardium and cultured cardiomyocytes. Moreover, hsa-miR-23a could directly down-regulate peroxisome proliferator-activated receptor γ coactivator-alpha (PGC-1α) expression in cardiomyocytes via binding to its 3′-untranslated regions, which implied that hsa-miR-23a could be critical for the down-regulation of PGC-1α under E2 deficiency (Sun et al., 2014).

4 Discussion

Deep learning offers a wide range of applications in major areas of computer science, such as computer vision, natural language processing and machine translation. More effective models can be obtained by adding hidden layers to standard neural networks. Deep learning also contributes to medication development and precision medicine by predicting potential SM-miRNA associations. Furthermore, deep learning models have more hidden layer nodes than conventional neural networks. The number of hidden layers can even reach ten for extremely complex problems. After multiple layers of calculation, the results of deep learning-based algorithms are often closer to the actual situation than those of traditional machine learning-based algorithms. Initially, we utilized matrix decomposition to reduce noise from known SM-miRNA associations. Then, the layer attention mechanism was introduced to the deep learning model, which significantly improved the performance of our model by integrating the SM-miRNA association feature vectors used for calculation.

GCNLASMMA is a model of a neural network with numerous hidden layers. Multiple layers computations allowed the results to completely consider known features and avoid overfitting. The attention mechanism extracted vital information from each layer of the neural network. Besides, the matrix decomposition module reduced the noise of known SM-miRNA associations, significantly enhancing GCN's performance. GCNLASMMA was an attempt to identify

TABLE 8 Validation of the top 50 miRNAs associated with E2 in the second type of case studies. The first column records the top 1–25 related miRNAs. The second column records the top 26–50 related miRNAs.

SM	miRNA	Evidence	SM	miRNA	Evidence
CID 5757	hsa-miR-183	unconfirmed	CID 5757	hsa-miR-19a	29416771
CID 5757	hsa-miR-30c-1	23220571	CID 5757	hsa-miR-19b-1	unconfirmed
CID 5757	hsa-miR-15a	unconfirmed	CID 5757	hsa-miR-125a	21914226
CID 5757	hsa-miR-181a-1	unconfirmed	CID 5757	hsa-miR-15b	23220571
CID 5757	hsa-let-7f-1	23220571	CID 5757	hsa-miR-128-2	23220571
CID 5757	hsa-miR-181b-1	unconfirmed	CID 5757	hsa-miR-20a	21914226
CID 5757	hsa-miR-205	unconfirmed	CID 5757	hsa-miR-26b	24735615
CID 5757	hsa-miR-181a-2	unconfirmed	CID 5757	hsa-miR-10b	23220571
CID 5757	hsa-miR-9-2	23220571	CID 5757	hsa-miR-181c	unconfirmed
CID 5757	hsa-miR-23a	23220571	CID 5757	hsa-miR-22	24715036
CID 5757	hsa-miR-128-1	23220571	CID 5757	hsa-miR-139	unconfirmed
CID 5757	hsa-let-7e	23220571	CID 5757	hsa-miR-106a	unconfirmed
CID 5757	hsa-let-7b	23220571	CID 5757	hsa-miR-141	unconfirmed
CID 5757	hsa-miR-130a	unconfirmed	CID 5757	hsa-let-7g	23220571
CID 5757	hsa-miR-338	22996663	CID 5757	hsa-miR-107	23220571
CID 5757	hsa-miR-30a	29331043	CID 5757	hsa-miR-23b	23220571
CID 5757	hsa-miR-302b	23220571	CID 5757	hsa-miR-195	unconfirmed
CID 5757	hsa-miR-130b	unconfirmed	CID 5757	hsa-miR-27a	23220571
CID 5757	hsa-miR-106b	28422740	CID 5757	hsa-miR-25	unconfirmed
CID 5757	hsa-miR-199b	unconfirmed	CID 5757	hsa-miR-204	29789714
CID 5757	hsa-miR-200b	23220571	CID 5757	hsa-miR-221	21057537
CID 5757	hsa-miR-182	28678802	CID 5757	hsa-miR-151a	unconfirmed
CID 5757	hsa-miR-26a-1	unconfirmed	CID 5757	hsa-miR-218-1	unconfirmed
CID 5757	hsa-miR-17	23220571	CID 5757	hsa-miR-130a	unconfirmed
CID 5757	hsa-miR-200c	23220571	CID 5757	hsa-miR-92a-1	unconfirmed

potential SM-miRNA associations using deep learning. The advantages above enabled GCNLASMMA to accurately anticipate potential SM-miRNA associations.

Deep learning's spectacular performance is contingent on a vast number of known SM-miRNA associations. The number of known SM-miRNA associations utilized in this investigation was apparently insufficient to fulfill GCNLASMMA. Therefore, the performance of GCNLASMMA was still unsatisfactory. In addition, the parameters used in GCNLASMMA may not be ideal. Moreover, the construction of heterogeneous networks will yield better results if other biological information, such as long non-coding RNA or disease, is utilized. These factors will motivate researchers to develop more effective deep learning models to predict potential SM-miRNA associations using more trustworthy biological datasets.

Data availability statement

The Python code and datasets of GCNLASMMA are publicly available at https://github.com/1054366388/GCNLASMMA.

Author contributions

JZL led the project and supervised the writing. JN did the experiments, searched the literature and wrote the article. XLC and TGN did the subsequent revisions. All authors participated in the revisions and approved the final version of the manuscript.

Funding

This study was supported by the Postgraduate Research and Practice Innovation Program of Jiangsu Province (Grant Number KYCX21_2836).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Agarwal, V., Bell, G. W., Nam, J. W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4. doi:10.7554/eLife.05005

Angermueller, C., Pärnamaa, T., Parts, L., and Stegle, O. (2016). Deep learning for computational biology. *Mol. Syst. Biol.* 12 (7), 878. doi:10.15252/msb.20156651

Bartel, D. P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* 116 (2), 281–297. doi:10.1016/s0092-8674(04)00045-5

Borges Oliveira, D. A., Ribeiro Pereira, L. G., Bresolin, T., Pontes Ferreira, R. E., and Reboucas Dorea, J. R. (2021). A review of deep learning algorithms for computer vision systems in livestock. *Livest. Sci.* 253, 104700. doi:10.1016/j.livsci.2021.104700

Cai, J., Yang, C., Yang, Q., Ding, H., Jia, J., Guo, J., et al. (2013). Deregulation of let-7e in epithelial ovarian cancer promotes the development of resistance to cisplatin. *Oncogenesis* 2 (10), e75. doi:10.1038/oncsis.2013.39

Carnevali, M., Parsons, J., Wyles, D. L., and Hermann, T. (2010). A modular approach to synthetic RNA binders of the hepatitis C virus internal ribosome entry site. *ChemBioChem* 11 (10), 1364–1367. doi:10.1002/cbic.201000177

Chandrasekaran, V., Sanghavi, S., Parrilo, P. A., and Willsky, A. S. (2009). Rank-sparsity incoherence for matrix decomposition. *SIAM J. Optim.* 21 (2), 572–596. doi:10.1137/090761793

Chen, C. Z., Sobczak, K., Hoskins, J., Southall, N., Marugan, J. J., Zheng, W., et al. (2012). Two high-throughput screening assays for aberrant RNA–protein interactions in myotonic dystrophy type 1. *Anal. Bioanal. Chem.* 402 (5), 1889–1898. doi:10.1007/s00216-011-5604-0

Chen, X., Zhou, C., Wang, C. C., and Zhao, Y. (2021). Predicting potential small molecule-miRNA associations based on bounded nuclear norm regularization. *Brief. Bioinform.* 22 (6), bbab328. doi:10.1093/bib/bbab328

Cristino, A. S., Nourse, J., West, R. A., Sabdia, M. B., Law, S. C., Gunawardana, J., et al. (2019). EBV microRNA-BHRF1-2-5p targets the 3'UTR of immune checkpoint ligands PD-L1 and PD-L2. *Blood* 134 (25), 2261–2270. doi:10.1182/blood.2019000889

Dai, X., and Tan, C. (2015). Combination of microRNA therapeutics with small-molecule anticancer drugs: Mechanism of action and co-delivery nanocarriers. *Adv. Drug Deliv. Rev.* 81, 184–197. doi:10.1016/j.addr.2014.09.010

Denzler, R., McGeary, S. E., Title, A. C., Agarwal, V., Bartel, D. P., and Stoffel, M. (2016). Impact of MicroRNA levels, target-site complementarity, and cooperativity on competing endogenous RNA-regulated gene expression. *Mol. Cell* 64 (3), 565–579. doi:10.1016/j.molcel.2016.09.027

Deyle, K., Kong, X. D., and Heinis, C. (2017). Phage selection of cyclic peptides for application inresearch and drug development. *Acc. Chem. Res.* 50 (8), 1866–1874. doi:10.1021/acs.accounts.7b00184

Do Amaral, G., Planello, A. C., Borgato, G., de Lima, D. G., Guimarães, G. N., Marqueso, M. R., et al. (2019). 5-Aza-CdR promotes partial MGMT demethylation and modifies expression of different genes in oral squamous cell carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol* 127 (5), 424–432. doi:10.1016/j.oooo.2019.01.006

Dragomir, M. P., Knutsen, E., and Calin, G. A. (2021). Classical and noncanonical functions of miRNAs in cancers. *Trends Genet.* 38, 379–394. doi:10.1016/j.tig.2021.10.002

Gam, J. J., Babb, J., and Weiss, R. (2018). A mixed antagonistic/synergistic miRNA repression model enables accurate predictions of multi-input miRNA sensor activity. *Nat. Commun.* 9 (1), 2430. doi:10.1038/s41467-018-04575-0

Geng, B., and Craig, T. J. (2021). Small molecule drugs for atopic dermatitis, rheumatoid arthritis, and hereditary angioedema. *Ann. Allergy Asthma Immunol.* 128, 263–268. doi:10.1016/j.anai.2021.10.015

Ghini, F., Rubolino, C., Climent, M., Simeone, I., Marzi, M. J., and Nicassio, F. (2018). Endogenous transcripts control miRNA levels and activity in mammalian cells by target-directed miRNA degradation. *Nat. Commun.* 9 (1), 3119. doi:10.1038/s41467-018-05182-9

Gorbea, C., Mosbruger, T., and Cazalla, D. (2017). A viral Sm-class RNA base-pairs with mRNAs and recruits microRNAs to inhibit apoptosis. *Nature* 550 (7675), 275–279. doi:10.1038/nature24034

Gottlieb, A., Stein, G. Y., Ruppin, E., and Sharan, R. (2011). Predict: A method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* 7, 496. doi:10.1038/msb.2011.26

Habib, G., and Qureshi, S. (2020). Optimization and acceleration of convolutional neural networks: A survey. *J. King Saud Univ. - Comput. Inf. Sci.* 34, 4244. doi:10.1016/j.jksuci.2020.10.004

Hammond, S. M. (2015). An overview of microRNAs. *Adv. Drug Deliv. Rev.* 87, 3–14. doi:10.1016/j.addr.2015.05.001

Haniff, H. S., Liu, X., Tong, Y., Meyer, S. M., Knerr, L., Lemurell, M., et al. (2021). A structure-specific small molecule inhibits a miRNA-200 family member precursor and reverses a type 2 diabetes phenotype. *Cell Chem. Biol.* 29, 300–311.e10. doi:10.1016/j.chembiol.2021.07.006

Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M. (2003). Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.* 125 (39), 11853–11865. doi:10.1021/ja036030u

Healy, N., Schiff, R., Osborne, C. K., and Kerin, M. (2012). Mirnas: Small molecules, big players in tamoxifen resistance in breast cancer. *Int. J. Surg.* 10 (8), S4. doi:10.1016/j.ijsu.2012.06.025

Iwata, T., Mizuno, N., Nagahara, T., Kaneda-Ikeda, E., Kajiya, M., Sasaki, S., et al. (2021). Cytokines regulate stemness of mesenchymal stem cells via miR-628-5p during periodontal regeneration. *J. Periodontol.* 93, 269–286. doi:10.1002/jper.21-0064

Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., et al. (2009). miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* 37, D98–D104. doi:10.1093/nar/gkn714

Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., et al. (2011). DrugBank 3.0: A comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* 39, D1035–D1041. doi:10.1093/nar/gkq1126

Kumar Kingsley, S. M., and Vishnu Bhat, B. (2017). Role of MicroRNAs in the development and function of innate immune cells. *Int. Rev. Immunol.* 36 (3), 154–175. doi:10.1080/08830185.2017.1284212

Kumari, R., Kumar, S., and Kant, R. (2018). Role of circulating miRNAs in the pathophysiology of CVD: As a potential biomarker. *Gene Rep.* 13, 146–150. doi:10.1016/j.genrep.2018.10.003

Lai-Kwon, J., Tiu, C., Pal, A., Khurana, S., and Minchom, A. (2021). Moving beyond epidermal growth factor receptor resistance in metastatic non-small cell lung cancer - a drug development perspective. *Crit. Rev. Oncol. Hematol.* 159, 103225. doi:10.1016/j.critrevonc.2021.103225

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75 (5), 843–854. doi:10.1016/0092-8674(93)90529-y

Lemaire, M., Chabot, G. G., Raynal, N. J., Momparler, L. F., Hurtubise, A., Bernstein, M. L., et al. (2008). Importance of dose-schedule of 5-aza-2'-deoxycytidine for epigenetic therapy of cancer. *BMC Cancer* 8, 128. doi:10.1186/1471-2407-8-128

Li, X., Rao, S., Wang, Y., and Gong, B. (2004). Gene mining: A novel and powerful ensemble decision approach to hunting for disease genes using microarray expression profiling. *Nucleic Acids Res.* 32 (9), 2685–2694. doi:10.1093/nar/gkh563

Li, Y., Qiu, C., Tu, J., Geng, B., Yang, J., Jiang, T., et al. (2014). HMDD v2.0: A database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* 42, D1070–D1074. doi:10.1093/nar/gkt1023

Li, X., Li, X., Liao, D., Wang, X., Wu, Z., Nie, J., et al. (2015). Elevated microRNA-23a expression enhances the chemoresistance of colorectal cancer cells with microsatellite instability to 5-fluorouracil by directly targeting ABCF1. *Curr. Protein Pept. Sci.* 16 (4), 301–309. doi:10.2174/1389203716041504291533309

Li, J., Peng, D., Xie, Y., Dai, Z., Zou, X., and Li, Z. (2021). Novel potential small molecule-MiRNA-cancer associations prediction model based on fingerprint, sequence, and clinical symptoms. *J. Chem. Inf. Model.* 61 (5), 2208–2219. doi:10.1021/acs.jcim.0c01458

Liu, X., Zhu, F., Ma, X., Tao, L., Zhang, J., Yang, S., et al. (2011). The therapeutic target database: An internet resource for the primary targets of approved, clinical trial and experimental drugs. *Expert Opin. Ther. Targets* 15 (8), 903–912. doi:10.1517/14728222.2011.586635

Liu, X., Wang, S., Meng, F., Wang, J., Zhang, Y., Dai, E., et al. (2013). SM2miR: A database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics* 29 (3), 409–411. doi:10.1093/bioinformatics/bts698

Liu, R., Lu, Z., Gu, J., Liu, J., Huang, E., Liu, X., et al. (2018). MicroRNAs 15A and 16-1 activate signaling pathways that mediate chemotaxis of immune regulatory B cells to colorectal tumors. *Gastroenterology* 154 (3), 637–651. e637. doi:10.1053/j.gastro.2017.09.045

Liu, F., Peng, L., Tian, G., Yang, J., Chen, H., Hu, Q., et al. (2020). Identifying small molecule-miRNA associations based on credible negative sample selection and random walk. *Front. Bioeng. Biotechnol.* 8, 131. doi:10.3389/fbioe.2020.00131

Longley, D. B., Harkin, D. P., and Johnston, P. G. (2003). 5-fluorouracil: Mechanisms of action and clinical strategies. *Nat. Rev. Cancer* 3 (5), 330–338. doi:10.1038/nrc1074

Lu, T. X., and Rothenberg, M. E. (2018). MicroRNA. *J. Allergy Clin. Immunol.* 141 (4), 1202–1207. doi:10.1016/j.jaci.2017.08.034

Lv, S., Li, Y., Wang, Q., Ning, S., Huang, T., Wang, P., et al. (2012). A novel method to quantify gene set functional association based on gene ontology. *J. R. Soc. Interface* 9 (70), 1063–1072. doi:10.1098/rsif.2011.0551

Lv, Y., Wang, S., Meng, F., Yang, L., Wang, Z., Wang, J., et al. (2015). Identifying novel associations between small molecules and miRNAs based on integrated molecular networks. *Bioinformatics* 31 (22), 3638–3644. doi:10.1093/bioinformatics/btv417

Meng, F., Yang, X., and Zhou, C. (2014). The augmented Lagrange multipliers method for matrix completion from corrupted samplings with application to mixed Gaussian-impulse noise removal. *Plos One* 9 (9), e108125. doi:10.1371/journal.pone.0108125

Monroig, P. d. C., Chen, L., Zhang, S., and Calin, G. A. (2015). Small molecule compounds targeting miRNAs for cancer therapy. *Adv. Drug Deliv. Rev.* 81, 104–116. doi:10.1016/j.addr.2014.09.002

Nair, A., Chung, H. C., Sun, T., Tyagi, S., Dobrolecki, L. E., Dominguez-Vidana, R., et al. (2018). Combinatorial inhibition of PTPN12-regulated receptors leads to a broadly effective therapeutic strategy in triple-negative breast cancer. *Nat. Med.* 24 (4), 505–511. doi:10.1038/nm.4507

Niu, Z., Zhong, G., and Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing* 452, 48–62. doi:10.1016/j.neucom.2021.03.091

Oh, J. Y., Choi, G. E., Lee, H. J., Jung, Y. H., Chae, C. W., Kim, J. S., et al. (2019). 17β-Estradiol protects mesenchymal stem cells against high glucose-induced mitochondrial oxidants production via Nrf2/Sirt3/MnSOD signaling. *Free Radic. Biol. Med.* 130, 328–342. doi:10.1016/j.freeradbiomed.2018.11.003

Parsons, J., Castaldi, M. P., Dutta, S., Dibrov, S. M., Wyles, D. L., and Hermann, T. (2009). Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. *Nat. Chem. Biol.* 5 (11), 823–825. doi:10.1038/nchembio.217

Peng, Y., Ganesh, A., Wright, J., Xu, W., and Ma, Y. (2012). Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (11), 2233–2246. doi:10.1109/TPAMI.2011.282

Qu, J., Chen, X., Sun, Y. Z., Li, J. Q., and Ming, Z. (2018). Inferring potential small molecule-miRNA association based on triple layer heterogeneous network. *J. Cheminform.* 10 (1), 30. doi:10.1186/s13321-018-0284-9

Qu, J., Chen, X., Sun, Y. Z., Zhao, Y., Cai, S. B., Ming, Z., et al. (2019). *In silico* prediction of small molecule-miRNA associations based on the HeteSim algorithm. *Mol. Ther. Nucleic Acids* 14, 274–286. doi:10.1016/j.omtn.2018.12.002

Rachoń, D., Myśliwska, J., Suchecka-Rachoń, K., Wieckiewicz, J., Myśliwski, A., and Mysliwski, A. (2002). Effects of oestrogen deprivation on interleukin-6 production by peripheral blood mononuclear cells of postmenopausal women. *J. Endocrinol.* 172 (2), 387–395. doi:10.1677/joe.0.1720387

Ruepp, A., Kowarsch, A., Schmidl, D., Buggenthin, F., Brauner, B., Dunger, I., et al. (2010). PhenomiR: A knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol.* 11 (1), R6. doi:10.1186/gb-2010-11-1-r6

Rupaimoole, R., and Slack, F. J. (2017). MicroRNA therapeutics: Towards a new era for the management of cancer and other diseases. *Nat. Rev. Drug Discov.* 16 (3), 203–222. doi:10.1038/nrd.2016.246

Saikia, M., Paul, S., and Chakraborty, S. (2020). Role of microRNA in forming breast carcinoma. *Life Sci.* 259, 118256. doi:10.1016/j.lfs.2020.118256

Seth, P. P., Miyaji, A., Jefferson, E. A., Sannes-Lowery, K. A., Osgood, S. A., Propp, S. S., et al. (2005). SAR by MS: Discovery of a new class of RNA-binding small molecules for the hepatitis C virus: Internal ribosome entry site IIA subdomain. *J. Med. Chem.* 48 (23), 7099–7102. doi:10.1021/jm050815o

Shang, J., Yang, F., Wang, Y., Wang, Y., Xue, G., Mei, Q., et al. (2014). MicroRNA-23a antisense enhances 5-fluorouracil chemosensitivity through APAF-1/caspase-9 apoptotic pathway in colorectal cancer cells. *J. Cell. Biochem.* 115 (4), 772–784. doi:10.1002/jcb.24721

Singh, S., Sharma, A., and Chauhan, V. K. (2021). Online handwritten Gurmukhi word recognition using fine-tuned Deep Convolutional Neural Network on offline features. *Mach. Learn. Appl.* 5, 100037. doi:10.1016/j.mlwa.2021.100037

Sun, L. Y., Wang, N., Ban, T., Sun, Y. H., Han, Y., Sun, L. L., et al. (2014). MicroRNA-23a mediates mitochondrial compromise in estrogen deficiency-induced concentric remodeling via targeting PGC-1α. *J. Mol. Cell. Cardiol.* 75, 1–11. doi:10.1016/j.yjmcc.2014.06.012

Tagliafierro, L., Glenn, O. C., Zamora, M. E., Beach, T. G., Woltjer, R. L., Lutz, M. W., et al. (2017). Genetic analysis of α-synuclein 3' untranslated region and its corresponding microRNAs in relation to Parkinson's disease compared to dementia with Lewy bodies. *Alzheimers Dement.* 13 (11), 1237–1250. doi:10.1016/j.jalz.2017.03.001

Thomou, T., Mori, M. A., Dreyfuss, J. M., Konishi, M., Sakaguchi, M., Wolfrum, C., et al. (2017). Adipose-derived circulating miRNAs regulate gene expression in other tissues. *Nature* 542 (7642), 450–455. doi:10.1038/nature21365

Thorne, N., Inglese, J., and Auld, D. S. (2010). Illuminating insights into firefly luciferase and other bioluminescent reporters used in chemical biology. *Chem. Biol.* 17 (6), 646–657. doi:10.1016/j.chembiol.2010.05.012

Tse, J., Martin-McNaulty, B., Halks-Miller, M., Kauser, K., DelVecchio, V., Vergona, R., et al. (1999). Accelerated atherosclerosis and premature calcified cartilaginous metaplasia in the aorta of diabetic male Apo E knockout mice can be prevented by chronic treatment with 17β-estradiol. *Atherosclerosis* 144 (2), 303–313. doi:10.1016/S0021-9150(98)00325-6

Vidal, R. (2011). Subspace clustering. *IEEE Signal Processing Magazine* 28 (2), 52–68. doi:10.1109/msp.2010.939739

Wang, C. C., and Chen, X. (2019). A unified framework for the prediction of small molecule-MicroRNA association based on cross-layer dependency inference on multilayered networks. *J. Chem. Inf. Model.* 59 (12), 5281–5293. doi:10.1021/acs.jcim.9b00667

Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., and Bryant, S. H. (2009). PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* 37, W623–W633. doi:10.1093/nar/gkp456

Wang, C. C., Chen, X., Qu, J., Sun, Y. Z., and Li, J. Q. (2019). Rfsmma: A new computational model to identify and prioritize potential small molecule-MiRNA associations. *J. Chem. Inf. Model.* 59 (4), 1668–1679. doi:10.1021/acs.jcim.9b00129

Wang, X., Zhao, Y., and Pourpanah, F. (2020). Recent advances in deep learning. *Int. J. Mach. Learn. Cybern.* 11 (4), 747–750. doi:10.1007/s13042-020-01096-5

Wang, S. H., Wang, C. C., Huang, L., Miao, L. Y., and Chen, X. (2021). Dual-Network Collaborative Matrix Factorization for predicting small molecule-miRNA associations. *Brief. Bioinform.* 23, bbab500. doi:10.1093/bib/bbab500

Wen, D., Danquah, M., Chaudhary, A. K., and Mahato, R. I. (2015). Small molecules targeting microRNA for cancer therapy: Promises and obstacles. *J. Control. Release* 219, 237–247. doi:10.1016/j.jconrel.2015.08.011

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* 75 (5), 855–862. doi:10.1016/0092-8674(93)90530-4

Xia, X., Wang, Y., Huang, Y., Zhang, H., Lu, H., and Zheng, J. C. (2019). Exosomal miRNAs in central nervous system diseases: Biomarkers, pathological mediators, protective factors and therapeutic agents. *Prog. Neurobiol.* 183, 101694. doi:10.1016/j.pneurobio.2019.101694

Yekkirala, A. S., Roberson, D. P., Bean, B. P., and Woolf, C. J. (2017). Breaking barriers to novel analgesic drug development. *Nat. Rev. Drug Discov.* 16 (8), 545–564. doi:10.1038/nrd.2017.87

Yin, J., Chen, X., Wang, C. C., Zhao, Y., and Sun, Y. Z. (2019). Prediction of small molecule-MicroRNA associations by sparse learning and heterogeneous graph inference. *Mol. Pharm.* 16 (7), 3157–3166. doi:10.1021/acs.molpharmaceut.9b00384

Yu, A. M., Choi, Y. H., and Tu, M. J. (2020). RNA drugs and RNA targets for small molecules: Principles, progress, and challenges. *Pharmacol. Rev.* 72 (4), 862–898. doi:10.1124/pr.120.019554

Zeng, X., Zhu, S., Hou, Y., Zhang, P., Li, L., Li, J., et al. (2020). Network-based prediction of drug-target interactions using an arbitrary-order proximity embedded deep forest. *Bioinformatics* 36 (9), 2805–2812. doi:10.1093/bioinformatics/btaa010

Zhong, H., Zhou, Y., Xu, Q., Yan, J., Zhang, X., Zhang, H., et al. (2021). Low expression of miR-19a-5p is associated with high mRNA expression of diacylglycerol O-acyltransferase 2 (DGAT2) in hybrid tilapia. *Genomics* 113 (4), 2392–2399. doi:10.1016/j.ygeno.2021.05.016