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 in 1993 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; Tagliaferro 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 may 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; Cristiano 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 (BNRSMMA), which treated the problem of potential SM-miRNA association prediction as a matrix complementation problem. In addition, BNRSMMA 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 PhenomirR (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 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,

$$SSM = \frac{\sum_{i} \alpha_i SS_i}{\sum_i \alpha_i}, \quad (i = 1, 2, 3, 4)$$

(1)

where $\alpha$ represents the weights of SM similarities. All of the measures are important in terms of biology. Thus, we set the values of all $\alpha$ 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.

$$SSM(s_i, s_j) = \frac{SSM(s_i, s_j)}{\sqrt{\sum_{i=1}^{4}SSM(s_i, s_i)} \sqrt{\sum_{i=1}^{4}SSM(s_i, s_i)}}$$

(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
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 (v2.0), 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 \( SM_1 \) and \( SM_2 \), respectively. Moreover, the integrated miRNA similarity \( SMR \) was obtained by the following equation,

\[
SMR = \frac{\sum \beta_j SM_j}{\sum \beta_j} \quad (j = 1, 2)
\]

(3)

where \( \beta_1 \) and \( \beta_2 \) represent the weights of miRNA similarities. Also, we set the values of \( \beta_1 \) and \( \beta_2 \) 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.

\[
SMR(m_i, m_j) = \frac{SMR(m_i, m_j)}{\sqrt{\sum_{i=1}^{n} SMR(m_i, m_j)} \sqrt{\sum_{j=1}^{m} SMR(m_i, m_j)}}
\]

(4)

2.4 GCNLASMMA

GCNLASMMA was separated into two steps. The known SM-miRNA association matrix was initially decomposed and reconstructed to obtain the new SM-miRNA association matrix \( A^* \). The reconstructed SM-miRNA association matrix \( A' \) was then calculated using the 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.

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.
Table 1: The illustration of the IALM algorithm.

Algorithm: Inexact augmented lagrange multipliers

Input: Known SM-miRNA association matrix $X$ and $\alpha = 0.1$
Initial: $X = 0$, $E = 0$, $Y_1 = 0$, $Y_2 = 0$, $\mu = 10^{-4}$, $\max_j = 10^{10}$, $\rho = 1.1$, $\epsilon = 10^{-10}$
While true
1. Fix others and $J = \text{arg}\min_{X,E} \|Y_1 - (X + Y_1/\rho)\|_2^2$
2. Fix others and $X = (I + A^T A)(A^T A - A^T E + J + (A^T Y_1 - Y_2/\rho))$
3. Fix others and $E = \text{arg}\min_{E,J} (\|E\|_2^2 + \|E - (A - AX + Y_1/\rho)\|_F^2)$
4. Update $Y_1 = Y_1 + \mu(A - AX - E)$. $Y_2 = Y_2 + \mu(X - J)$
5. Update $\mu = \min(\mu, \max_j)$
If $\|A - AX - E\|_\infty < \epsilon$ and $\|X - J\|_\infty < \epsilon$
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 GCNLSMMA 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:

$$A = X \times E + E$$ (5)

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

$$\min_{X,E} \|X\|_* + \alpha \|E\|_{1,1} \text{ s.t. } A = X \times E + E$$ (6)

where $\|X\|_* = \sum_i \sigma_i$, ($\sigma_i$ is the singular value of matrix $X$), $\|E\|_{1,1} = \sum_{i,j} \sqrt{E_{ij}^2}$. 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 $\alpha$ 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).

$$\min_{X,E} \|X\|_* + \alpha \|E\|_{1,1} \text{ s.t. } A = X \times E + E$$ (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.

$$L = \|I\|_* + \frac{\alpha}{2} \|E\|_{1,1} + \frac{\delta}{2} (\|A - AX - E\|_F^2 + \|X - J\|_F^2)$$ (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 $Y_j$ 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:

$$A^* = X \times E^*$$ (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:

$$A_H = \begin{bmatrix} -SMR & A^* \\ A^* \times \text{SSM} & -SSM \end{bmatrix}$$ (10)

where $A^T$ represents the transpose of $A^*$. In Eq. 10, we normalized the similarity matrix of SM and miRNA by $-SSM = D_{m}^{-1}SSMD_{m}^{1}$ and $-SMR = D_{n}^{-1}SMRD_{n}^{1}$, respectively. Specifically, $D_i = \text{diag}(\sum_j SSM_{ij})$ and $D_m = \text{diag}(\sum_j SMR_{ij})$.

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 GCNLSMMA, 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...
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:

\[ H^{(l+1)} = f(H^{(l)}, G) = \sigma(D^{-1/2}GD^{1/2}H^{(l)}W^{(l)}) \]  

(11)

In Eq. 11, \( H^{(0)} \) and \( H^{(l+1)} \) denote the embeddings of nodes in the \( l_{th} \) and \( (l+1)_{th} \) layers, respectively. \( D = \text{diag}(\sum_j G_{ij}) \) 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 \( \mu \) in the input graph \( G \) during the propagation process as follows:

\[ G = \begin{bmatrix} \mu - \text{SMR} & A^* \\ A^T & \mu - \text{SSM} \end{bmatrix} \]  

(12)

Then, we initialized the input layer embeddings as:

\[ H^{(0)} = \begin{bmatrix} 0 \\ A^T \end{bmatrix} \]  

(13)

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

\[ H^{(1)} = \sigma(D^{-1/2}GD^{1/2}H^{(0)}W^{(0)}) \]  

(14)

In Eq. 12, \( W^{(0)} \) is a weight matrix that acts only between the input layer and the first hidden layer. \( H^{(l)} \) 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 \( l = 1, 2, \ldots, L \). After \( L \) iterations, we obtained \( L \) \( k \)-dimensional embeddings from different graph convolution.
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 $AH$. 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 $\frac{1}{\sqrt{L}}, l = 1, 2, \cdots, L$. Finally, we obtained the reconstructed SM-miRNA association matrix $A'$ by an activation function as follows,

$$A' = sigmoid(H_m W'H_s^T)$$

(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 GCNLASMA, 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, dataset1 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 $0.9721 \pm 0.0018$ and $0.8393 \pm 0.0047$, 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 GCNLASMA 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 GCNLASMA in identifying
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 studies.

**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.
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 GCNLASMA, 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 GCNLASMA’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 calculate 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, GCNLASMA achieved the AUCs of 0.9721 ± 0.0018 and 0.8393 ± 0.0047 under 5-fold cross-validation based on dataset1 and dataset2, respectively.

# 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  | 28928882 |
| CID 3385 | hsa-let-7e   | 23220571 | CID 3385 | hsa-miR-130a | unconfirmed |
| CID 3385 | hsa-miR-19a  | 23220571 | CID 3385 | hsa-miR-10b  | 23232955 |
| 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 |
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:

| 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 | uncon |
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 GCNLASMMMA. 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

| SM     | miRNA     | Evidence | SM     | miRNA     | Evidence |
|--------|-----------|----------|--------|-----------|----------|
| CID 451668 | hsa-miR-20a | 23220571 | CID 451668 | hsa-miR-30a | unconffirmed |
| 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 | unconffirmed |
| CID 451668 | hsa-miR-204 | unconffirmed | CID 451668 | hsa-miR-92a-1 | unconffirmed |
| CID 451668 | hsa-miR-200b | 23626803 | CID 451668 | hsa-miR-181a-1 | 23220571 |
| CID 451668 | hsa-miR-23a | unconffirmed | CID 451668 | hsa-let-7e | 22053057 |
| CID 451668 | hsa-let-7f-1 | 23220571 | CID 451668 | hsa-miR-26a-1 | unconffirmed |
| CID 451668 | hsa-let-7b | 26708866 | CID 451668 | hsa-miR-1233-1 | unconffirmed |
| CID 451668 | hsa-miR-290c | 23626803 | CID 451668 | hsa-miR-130a | 23220571 |
| CID 451668 | hsa-miR-25 | 23220571 | CID 451668 | hsa-miR-30c-1 | unconffirmed |
| CID 451668 | hsa-miR-128-1 | 27705931 | CID 451668 | hsa-miR-22 | 23220571 |
| CID 451668 | hsa-miR-145 | 26198104 | CID 451668 | hsa-miR-301a | unconffirmed |
| CID 451668 | hsa-miR-221 | unconffirmed | CID 451668 | hsa-let-7g | 23220571 |
| CID 451668 | hsa-miR-19b-1 | unconffirmed | CID 451668 | hsa-miR-195 | 23333942 |
| CID 451668 | hsa-miR-197 | unconffirmed | CID 451668 | hsa-miR-302b | unconffirmed |
| CID 451668 | hsa-let-7i | 23220571 | CID 451668 | hsa-miR-26b | unconffirmed |
| CID 451668 | hsa-miR-181b-1 | unconffirmed | CID 451668 | hsa-miR-205 | unconffirmed |
| CID 451668 | hsa-miR-338 | unconffirmed | CID 451668 | hsa-miR-218-1 | unconffirmed |
| CID 451668 | hsa-let-7d | 26802971 | CID 451668 | hsa-miR-93 | 23220571 |
| CID 451668 | hsa-miR-139 | unconffirmed | CID 451668 | hsa-miR-124-1 | unconffirmed |
| CID 451668 | hsa-miR-328 | unconffirmed | CID 451668 | hsa-miR-15b | unconffirmed |
| CID 451668 | hsa-miR-126 | 23220571 | CID 451668 | hsa-miR-10b | unconffirmed |
| CID 451668 | hsa-miR-17 | 23220571 | CID 451668 | hsa-miR-128-2 | unconffirmed |
| CID 451668 | hsa-miR-19a | 23220571 | CID 451668 | hsa-miR-27a | 23220571 |
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 | unconformed |
| CID 451668 | hsa-miR-181b-1 | unconformed | CID 451668 | hsa-miR-125a | 23220571 |
| CID 451668 | hsa-miR-205 | unconformed | CID 451668 | hsa-let-7b | 26708866 |
| CID 451668 | hsa-miR-19a | 23220571 | CID 451668 | hsa-miR-302b | unconformed |
| CID 451668 | hsa-miR-181a-1 | 23220571 | CID 451668 | hsa-miR-30a | unconformed |
| 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 | unconformed |
| CID 451668 | hsa-miR-126 | 23220571 | CID 451668 | hsa-miR-15b | unconformed |
| CID 451668 | hsa-miR-320a | 26198104 | CID 451668 | hsa-miR-124-1 | unconformed |
| CID 451668 | hsa-miR-30c-1 | unconformed | CID 451668 | hsa-miR-26b | unconformed |
| CID 451668 | hsa-miR-328 | unconformed | CID 451668 | hsa-miR-128-1 | 27705931 |
| CID 451668 | hsa-let-7e | 22033057 | CID 451668 | hsa-let-7a-1 | unconformed |
| CID 451668 | hsa-miR-10b | unconformed | CID 451668 | hsa-miR-218-1 | unconformed |
| CID 451668 | hsa-let-7f-1 | 23220571 | CID 451668 | hsa-miR-200c | 23626803 |
| CID 451668 | hsa-miR-221 | unconformed | CID 451668 | hsa-miR-26a-1 | unconformed |
| CID 451668 | hsa-miR-182 | 23220571 | CID 451668 | hsa-miR-338 | unconformed |
| CID 451668 | hsa-let-7i | 23220571 | CID 451668 | hsa-miR-93 | 23220571 |
| CID 451668 | hsa-miR-195 | 23333942 | CID 451668 | hsa-miR-139 | unconformed |
| CID 451668 | hsa-miR-27a | 23220571 | CID 451668 | hsa-miR-145 | 26198104 |
| CID 451668 | hsa-miR-204 | unconformed | CID 451668 | hsa-miR-107 | 23220571 |
| CID 451668 | hsa-miR-25 | 23220571 | CID 451668 | hsa-let-7d | 26802971 |
| CID 451668 | hsa-miR-23a | unconformed | CID 451668 | hsa-miR-19b-1 | unconformed |
| CID 451668 | hsa-let-7f-1 | 23220571 | CID 451668 | hsa-miR-22 | 23220571 |
| CID 451668 | hsa-miR-17 | 23220571 | CID 451668 | hsa-miR-197 | unconformed |

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
et al., 1999; Rachoni 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, significantly enhancing GCN’s performance. GCNLASMMA was an attempt to identify

| SM | miRNA | Evidence | SM | miRNA | Evidence |
|----|-------|----------|----|-------|----------|
| CID 5757 | hsa-miR-183 | unconconfirmed | CID 5757 | hsa-miR-181b-1 | unconconfirmed |
| CID 5757 | hsa-miR-181a-2 | unconconfirmed | CID 5757 | hsa-miR-19b-1 | unconconfirmed |
| CID 5757 | hsa-miR-125a | 21914226 | CID 5757 | hsa-miR-141 | unconconfirmed |
| 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 | 29391043 |
| CID 5757 | hsa-miR-195 | unconconfirmed | CID 5757 | hsa-miR-7F-1 | 23220571 |
| CID 5757 | hsa-miR-128-2 | 23220571 | CID 5757 | hsa-miR-30b | 23220571 |
| CID 5757 | hsa-miR-181a-1 | unconconfirmed | CID 5757 | hsa-miR-199b | unconconfirmed |
| CID 5757 | hsa-miR-128-1 | 23220571 | CID 5757 | hsa-miR-181c | unconconfirmed |
| CID 5757 | hsa-miR-130a | unconconfirmed | CID 5757 | hsa-miR-106b | 28422740 |
| CID 5757 | hsa-miR-338 | 22996663 | CID 5757 | hsa-miR-23a | 23220571 |
| CID 5757 | hsa-miR-37e | 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 | unconconfirmed |
| CID 5757 | hsa-miR-27a | 23220571 | CID 5757 | hsa-miR-25 | unconconfirmed |
| CID 5757 | hsa-miR-200b | 23220571 | CID 5757 | hsa-miR-218-1 | unconconfirmed |
| CID 5757 | hsa-miR-221 | 21057337 | CID 5757 | hsa-miR-218-1 | unconconfirmed |
| CID 5757 | hsa-miR-151a | unconconfirmed | CID 5757 | hsa-miR-22 | 24715036 |
| CID 5757 | hsa-miR-204 | 29109714 | CID 5757 | hsa-miR-15b | 23220571 |
| CID 5757 | hsa-miR-106a | unconconfirmed | CID 5757 | hsa-miR-130a | unconconfirmed |
| CID 5757 | hsa-miR-205 | unconconfirmed | CID 5757 | hsa-miR-23b | 23220571 |
| CID 5757 | hsa-miR-92a-1 | unconconfirmed | CID 5757 | hsa-miR-26a-1 | unconconfirmed |
| CID 5757 | hsa-miR-130b | unconconfirmed | CID 5757 | hsa-miR-30c-1 | 23220571 |
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.

**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 | 24735645 |
| 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-7c | 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 |
Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Agarwal, V., Bell, G. W., Nam, J. W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Cell 160, 176–186. doi:10.1016/j.cell.2014.11.050

Ahn, J. S., Kim, K. H., and Kim, S. (2012). Target site sequence influences microRNA targeting in mammalian cells. Genes Dev. 26, 1770–1775. doi:10.1101/gad.190729.111

Angelucci, A., Beltrami, S., and Capogna, M. (2010). miR-138 promotes myogenic differentiation by targeting paired box gene 8. Nucleic Acids Res. 38, 369–382. doi:10.1093/nar/gkp1165

Bali, A., Kim, J., and Kim, S. (2014). miR-99a-5p activates EphB4 by targeting its 3'-UTR and promotes tumorigenicity in breast cancer cells. Cancer Lett. 347, 39–46. doi:10.1016/j.canlet.2014.07.020

Barbosa, M., and Burstein, A. J. (2017). The importance of targeting miR-155 in cancer therapy. Cancer Treat. Rev. 57, 1–8. doi:10.1016/j.ctrv.2017.05.001

Bartel, D. P. (2016). MicroRNAs and their targets: A view from the inside. Annu. Rev. Genet. 50, 557–583. doi:10.1146/annurev-genet-011416-024209

Bartelt, A., and Bartel, D. P. (2012). The microRNA database (miRbase). Nucleic Acids Res. 40, D169–D172. doi:10.1093/nar/gkr966

Bartel, D. P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell 116 (2), 281–297. doi:10.1016/j.cell.2004.05.005

Borges Oliveira, D. A., and Calin, G. A. (2021). Classical and noncanonical pathways of microRNA biogenesis. Cancer Treat. Rev. 90, 101682. doi:10.1016/j.ctrv.2021.101682

Bose, S., and Pandey, S. (2015). A new paradigm for cancer treatment: miRNA therapy. J. Cell. Biochem. 116, 279–289. doi:10.1002/jcb.25370

Bourc’his, D., and Bestor, T. H. (2003). Structure and function of the DNA methyltransferases. Annu. Rev. Genet. 37, 689–718. doi:10.1146/annurev.genet.37.122102.141539

Braithwaite, A. R., and Matzke, M. A. (2015). miR-145 as a regulator of the Wnt/beta-catenin pathway: role in cancer. Cancer Biol. Ther. 16, 526–533. doi:10.4161/cbt.34112

Cao, Y., Liu, Y., and Liu, X. (2014). miR-29a promotes breast cancer progression by targeting AOC1. Cancer Lett. 347, 64–71. doi:10.1016/j.canlet.2014.05.012

Carmesin, G., Lichtensteiger, W., and Pachinger, O. K. (2015). miRNA in clinical practice: Caveats and challenges. Nat. Rev. Mol. Cell Biol. 16, 201–213. doi:10.1038/nrm.2014.32

Chen, L. T., Chen, J. L., Lin, T. C., and Lee, Y. C. (2012). miR-146b regulates cyclin D1 expression and cell cycle progression in gastric cancer cells. World J. Gastroenterol. 18, 1908–1915. doi:10.3748/wjg.v18.i9.1908

Chen, X., Ding, M., Sun, Y., Chen, Y., and Yang, J. (2018). Hsa-miR-21 promotes colorectal cancer development via modulating PDK4 expression. Mol. Med. Rep. 17, 305–311. doi:10.3892/mmr.2017.7487

Chen, X., Zhou, Y., and Mao, Y. (2015). miR-19a expression in MM and its clinical significance. Clin. Lab. 61, 1979–1984. doi:10.3233/CL-150947

Chen, Z., Liu, Z., and Wang, Y. (2015). miR-210 is involved in colorectal cancer cell proliferation and invasion through regulating miR-146b. Oncol. Lett. 9, 251–256. doi:10.3892/ol.2015.3502

Chen, Z., Zeng, H., Song, Y., and Li, R. (2017). miR-132 promotes progression of non-small-cell lung cancer via targeting PTENations. Int. J. Mol. Med. 40, 1473–1480. doi:10.3892/ijmm.2017.3464

Chen, J., Wang, X., Liu, Y., Sun, X., and Wu, J. (2017). miR-132 promotes cell proliferation and invasion through targeting PTEN in breast cancer. J. Cell Physiol. 232, 8060–8070. doi:10.1002/jcp.26172

Chen, J.-C., Guo, Y., and Zhu, L.-H. (2019). miR-146b expression is associated with the recurrence and prognosis of non-small-cell lung cancer. J. Exp. Clin. Cancer Res. 38, 201. doi:10.1186/s13046-019-1277-7

Chen, J.-C., Guo, Y., and Zhu, L.-H. (2019). miR-146b expression is associated with the recurrence and prognosis of non-small-cell lung cancer. J. Exp. Clin. Cancer Res. 38, 201. doi:10.1186/s13046-019-1277-7

Chen, J.-C., Guo, Y., and Zhu, L.-H. (2019). miR-146b expression is associated with the recurrence and prognosis of non-small-cell lung cancer. J. Exp. Clin. Cancer Res. 38, 201. doi:10.1186/s13046-019-1277-7

Chen, J.-C., Guo, Y., and Zhu, L.-H. (2019). miR-146b expression is associated with the recurrence and prognosis of non-small-cell lung cancer. J. Exp. Clin. Cancer Res. 38, 201. doi:10.1186/s13046-019-1277-7

Chen, J.-C., Guo, Y., and Zhu, L.-H. (2019). miR-146b expression is associated with the recurrence and prognosis of non-small-cell lung cancer. J. Exp. Clin. Cancer Res. 38, 201. doi:10.1186/s13046-019-1277-7

Chen, J.-C., Guo, Y., and Zhu, L.-H. (2019). miR-146b expression is associated with the recurrence and prognosis of non-small-cell lung cancer. J. Exp. Clin. Cancer Res. 38, 201. doi:10.1186/s13046-019-1277-7
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:10.1517/17422822.2011.586635
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/btq698
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 chemotherapy of immune regulatory B cells to colorectal tumors. Gastroenterology 154 (5), 637–651. 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. Biomater. 8, 131. doi:10.3389/fbioe.2020.00013
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, X., Yang, 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 multiplier method for matrix completion from corrupted samples with application to mixed Gaussian-biological/btv417
Mechanisms of action and clinical strategies. Bioinformatics 1517/14728222.2011.586635
and random walk. Front. Bioeng. Biomater. 8, 131. doi:10.3389/fbioe.2020.00013
Ni et al. 10.3389/fmolb.2022.1009099
production by peripheral blood mononuclear cells of postmenopausal women. J. Med. Chem. 54, 505–511. doi:10.1021/acs.jmedchem.5b01452
and random walk. Front. Bioeng. Biomater. 8, 131. doi:10.3389/fbioe.2020.00013
PhenomiR: A knowledgebase for microRNA expression in diseases and phenotypes. Front. Cell Dev. Biol. 8, 131. doi:10.3389/fcell.2020.00131
Vidal, R. (2011). Subspace clustering. IEEE Signal Processing Magazine 28 (2), 52–68. doi:10.1109/MSP.2010.937937
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., Suzeck, 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). Rfmmeta: 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.9b00029
Wang, X., Zhao, Y., and Pourpanah, F. (2020). Recent advances in deep learning in the development of drug discovery. Mol. Pharm. 17 (7), 2812–2830. doi:10.1021/acs.molpharmaceut.9b00852
Wang, S. H., Wang, C. C., Huang, L., Miao, L. Y., and Chen, X. (2021). Dual-Netword 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
Wrightman, 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:10.1016/0092-8240(93)90530-4
Xia, X., Wang, Y., Huang, Y., Zhang, H., Lu, H., and Zheng, J. C. (2019). Exosomal microRNAs in central nervous system diseases: Biomarkers, pathogenic mediators, protective factors and therapeutic agents. Prog. Neurobiol. 183, 101694. doi:10.1016/j.pneurobio.2019.101694
Yekikala, 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 molecules-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.119.011954
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., Zhong, H., et al. (2021). Low expression of miR-19a-5p is associated with high microRNA expression of diacylglycerol O-acetyltransferase 2 (DGAT2) in hybrid tilapia. Genomics 113 (4), 2392–2399. doi:10.1016/j.ygeno.2021.05.016