



Application of Bidirectional Generative Adversarial Networks to Predict Potential miRNAs Associated With Diseases

Long Xu¹, Xiaokun Li^{1,2*}, Qiang Yang³, Long Tan¹, Qingyuan Liu² and Yong Liu^{1*}

¹School of Computer Science and Technology, Heilongjiang University, Harbin, China, ²Postdoctoral Program of Heilongjiang Hengxun Technology Co., Ltd., Heilongjiang University, Harbin, China, ³School of Electronic Engineering, Heilongjiang University, Harbin, China

OPEN ACCESS

Edited by:

Dongxiao Zhu,
Wayne State University, United States

Reviewed by:

Jin-Xing Liu,
Qufu Normal University, China
Leon Wong,
Guangxi Academy of Sciences, China

*Correspondence:

Xiaokun Li
li.xiaokun@163.com
Yong Liu
liuyong123456@hju.edu.cn

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 05 May 2022

Accepted: 08 June 2022

Published: 12 July 2022

Citation:

Xu L, Li X, Yang Q, Tan L, Liu Q and
Liu Y (2022) Application of Bidirectional
Generative Adversarial Networks to
Predict Potential miRNAs Associated
With Diseases.
Front. Genet. 13:936823.
doi: 10.3389/fgene.2022.936823

Substantial evidence has shown that microRNAs are crucial for biological processes within complex human diseases. Identifying the association of miRNA–disease pairs will contribute to accelerating the discovery of potential biomarkers and pathogenesis. Researchers began to focus on constructing computational models to facilitate the progress of disease pathology and clinical medicine by identifying the potential disease-related miRNAs. However, most existing computational methods are expensive, and their use is limited to unobserved relationships for unknown miRNAs (diseases) without association information. In this manuscript, we proposed a creatively semi-supervised model named bidirectional generative adversarial network for miRNA–disease association prediction (BGANMDA). First, we constructed a microRNA similarity network, a disease similarity network, and Gaussian interaction profile kernel similarity based on the known miRNA–disease association and comprehensive similarity of miRNAs (diseases). Next, an integrated similarity feature network with the full underlying relationships of miRNA–disease pairwise was obtained. Then, the similarity feature network was fed into the BGANMDA model to learn advanced traits in latent space. Finally, we ranked an association score list and predicted the associations between miRNA and disease. In our experiment, a five-fold cross validation was applied to estimate BGANMDA's performance, and an area under the curve (AUC) of 0.9319 and a standard deviation of 0.00021 were obtained. At the same time, in the global and local leave-one-out cross validation (LOOCV), the AUC value and standard deviation of BGANMDA were 0.9116 ± 0.0025 and 0.8928 ± 0.0022 , respectively. Furthermore, BGANMDA was employed in three different case studies to validate its prediction capability and accuracy. The experimental results of the case studies showed that 46, 46, and 48 of the top 50 prediction lists had been identified in previous studies.

Keywords: microRNAs, complex disease, miRNA–disease association prediction, similarity network, bidirectional generative adversarial network

1 INTRODUCTION

MicroRNAs (miRNAs) are endogenous gene-encoded non-coding single-stranded RNA molecules with a length of about 20–24 nucleotides (Bartel, 2004; Cheng et al., 2005). They participated in the post-transcriptional management of gene expression and are crucial for multiple biochemical processes (such as cellular apoptosis, cell proliferation, metabolism, and so on) by targeting specific message RNAs (mRNAs) and regulating gene expression and miRNA degradation (Ambros, 2003; Karp and Ambros, 2005; Shukla et al., 2011; Ebert and Sharp, 2012; Shi et al., 2013). Therefore, scientists are increasingly interested in human biochemical processes at the miRNA level (Xu et al., 2004). Currently, accumulating evidence indicates that most miRNAs are directly related to complex human diseases (Meltzer, 2005; Krützfeldt and Stoffel, 2006; Jiang et al., 2009; Chen et al., 2019). For instance, it was confirmed that mir-367 can facilitate the proliferation of hepatocellular carcinoma and invasion of cells, displaying aberrant expressions in the tumor tissues of hepatocellular cancer patients compared to normal tissues (Guo et al., 2020). In addition, recent studies revealed that mir-21 is correlated with the prognosis of patients with brain tumors, and particularly, its overexpression may cause a worse prognosis (He et al., 2016). Another example of disease-related miRNA is mir-125a-5p, which plays a critical role in lung cancer development under the regulation of epidermal growth factor signal transduction (Wang et al., 2009). Therefore, utilizing experimental or computational models to identify the underlying miRNA–disease associations will contribute to facilitating the discovery of pathogenic mechanisms and potential biomarkers (Calin and Croce, 2006; Jones et al., 2014). Previous studies have indicated that predicting possible miRNA–disease associations through traditionally biological methods is laborious and expensive. With the rapid evolution of technology and science, numerous advanced computational methods have been developed to establish the representation of pairwise miRNA–disease associations based on accumulated genomic data.

Based on the assumption that functionally similar mirnas are likely to be associated with phenotypically similar diseases (Perez-Iratxeta et al., 2005), researchers had made great progress over the past few decades in building computational models to infer potential miRNA–disease associations. Most of them were score function-based models, which analyze biological information to establish the score function and predict the associations (Zeng et al., 2016b). For example, an innovative model named MiRNA–protein–disease Association was proposed (Mørk et al., 2014), which utilized proteins as mediums between miRNAs and diseases to predict the associations. A group of proteins correlated with a specific miRNA (disease) was listed, the association scores of miRNA–protein pairs and protein–disease pairs were calculated, and the maximum value was selected as the final association score is between miRNA and disease. In addition Chen et al. (2016) mixed the known miRNA–disease associations, the diseases semantic similarity, and the miRNAs functional similarity with GIP kernel similarity to figure out unknown miRNA–disease associations. Then, a model named within and

between score for MiRNA–disease association prediction (WBSMDA) was developed, which can simultaneously prioritize miRNAs for all diseases. A new similarity matrix of miRNA–disease associations was constructed (Ma et al., 2019) by integrating gene similarity information, miRNA target gene information, disease gene information, and other data sources. They applied the nuclear neighborhood similarity algorithm to calculate the similarity feature of miRNA–disease pairs. Ultimately, a bidirectional propagation algorithm was adopted to obtain the predicted score.

In the past 2 decades, machine learning-based models have been widely proposed to predict the underlying associations of miRNA–disease pairs. For example, researchers constructed a vector spacer model named MiRAI (Pasquier and Gardès, 2016). First, an adjacent matrix was acquired by splicing four types of miRNA–related associations among disease, target, neighbor, and cluster. Then, they applied a singular value decomposition algorithm to cut down the new matrix dimension and computed an eigenvector representation of each miRNA–disease pair. Finally, the correlation grade was obtained using the cosine similarity of miRNA–disease vector representation. Based on combining two ideal classifiers in disease (miRNA) space to optimize the association probability, a model named Regularized Least Squares for predicting miRNA–disease associations (RLSMDA) was proposed (Chen and Yan, 2014). It is worth mentioning that this method does not require any representation of unknown miRNA–disease pairs. An inductive matrix completion model was developed (Chen et al., 2018) to predict the disease-related miRNAs, which is applicable to new diseases with unknown miRNAs. First, they set up two metrics to indicate the low-dimensional matrix of miRNA–disease representation. Then, an optimal algorithm was used in iteration to update them. When the stopping threshold was reached, the two updated matrices were directly fused into the miRNA–disease similarity matrix. As an enhancement, researchers innovatively constructed an updated model named Neighborhood Constraint Matrix Completion for miRNA–disease association prediction (NCMCMDA), which utilized the similarity information of miRNAs and diseases (Chen et al. (2021)). They applied a fast iterative shrinkage-thresholding algorithm based on the known miRNA–disease associations and comprehensive miRNA (disease) similarity to recover the missing association information. A label propagation-based method was proposed (Li et al., 2018) for scoring miRNA–disease pairs by calculating pairwise neighborhood similarity (LPLNS). Due to unvalidated miRNA–disease pairs presenting few known associations, an additional processing step was included in the LPLNS to drive new interaction likelihood profiles. To list the candidate miRNAs for diseases and explore the potential associations, a novel framework called GBDT-LR was constructed by combining logistic regression and gradient boosting decision tree (Zhou et al., 2020). Besides, researchers developed a computational model based on Similarity Constrained Matrix Factorization for miRNA–disease association prediction (Li et al., 2021a), which creatively expanded L2 regular term and similarity constraint term to infer disease-related miRNAs.

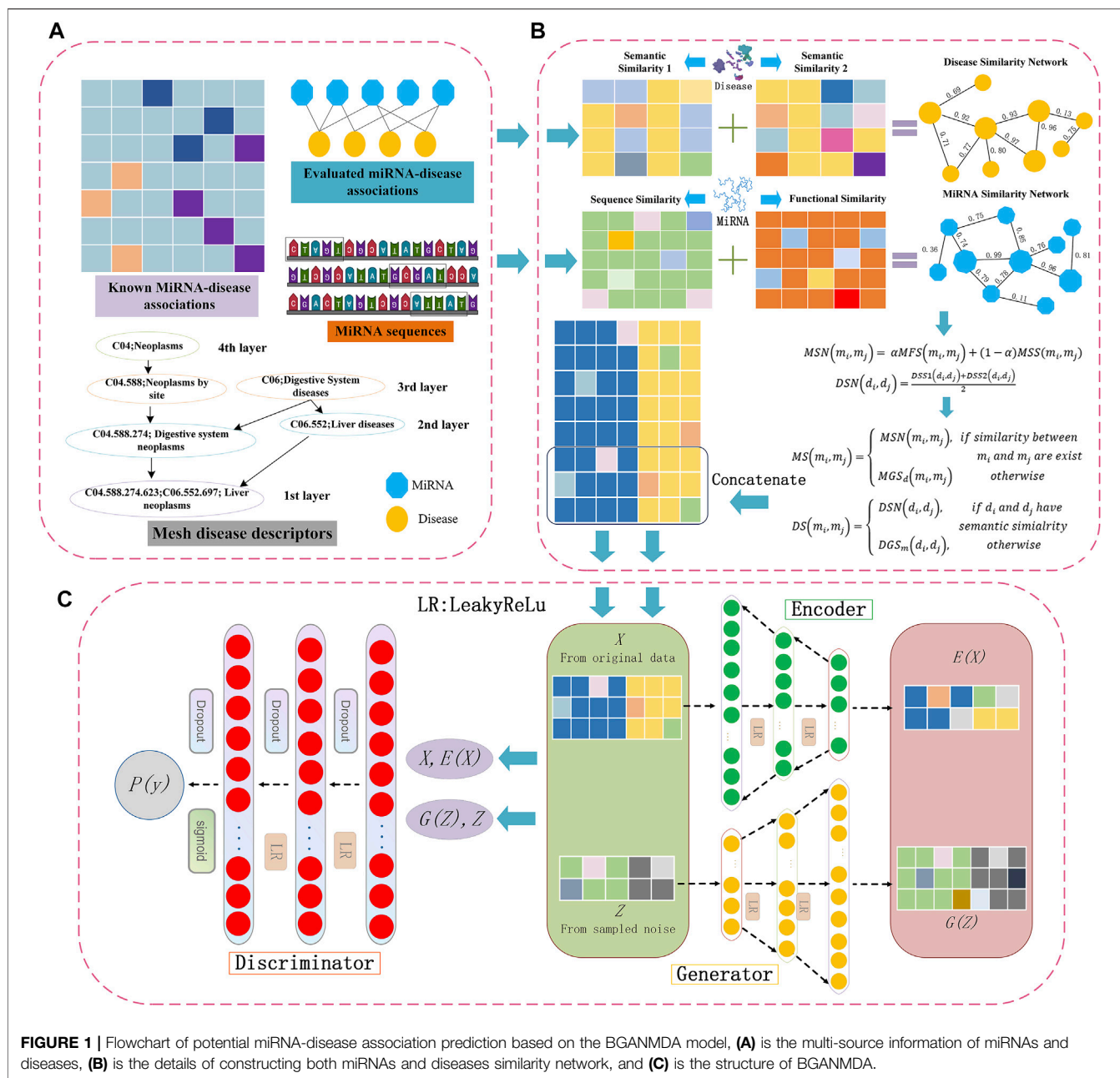


FIGURE 1 | Flowchart of potential miRNA-disease association prediction based on the BGANMDA model. **(A)** is the multi-source information of miRNAs and diseases, **(B)** is the details of constructing both miRNAs and diseases similarity network, and **(C)** is the structure of BGANMDA.

Many novel neural network-based methods that extract similarity features and learn the latent representations had been proposed to predict potential miRNA-disease associations (Yang and Li, 2021). For example, a model called Deep Belief Network for miRNA-disease association prediction (DBNMDA) was mentioned (Chen et al., 2021), which innovatively constructed the feature vectors with all microRNA and disease information to pre-train restricted Boltzmann machines. A graph neural network-based auto-encoder model called GAEMDA (Li et al., 2021b) adopted an end-to-end way to identify the underlying associations between miRNAs and diseases. By combining an auto-encoder and a convolutional neural network (Peng et al., 2019), a learning-based neural network

model was constructed to figure out the potential miRNA-disease associations. Based on four integrated biological networks and verified protein-protein interaction in humans, researchers developed a new computational framework named Heterogeneous Graph Convolutional Network for miRNA-disease associations prediction (Li et al., 2019). For predicting the correlations between miRNAs and diseases, some studies presented a supervised end-to-end method, termed the neural inductive matrix completion with graph convolutional network (Li et al., 2020), which can effectively learn the representation of underlying traits from the known miRNA (disease) information. To learn the original and global miRNA-disease representations in a low-dimensional feature space (Xuan et al., 2019), a novel

network-based model was proposed, termed Convolutional Neural network for miRNA-disease associations prediction (CNNMDA).

Despite the great progress achieved in the techniques used to explore the potential miRNA and disease relationships, the above-mentioned models, and also others, still present some restrictions and disadvantages. Overall, the studies published in the past decade show that great success has been achieved in the field of bioinformatics based on deep learning-based models. In particular, computational models based on neural networks have made outstanding contributions to the task of prediction (Zeng et al., 2016a). As a neural network, the auto-encoder can learn input data through unsupervised learning, strongly represent potential features, effectively reduce sample noise, and randomly generate data (Suk et al., 2015). For example, aiming to discover complex feature representation of disease-related miRNAs, a novel deep learning method for predicting miRNA-disease associations through deep autoencoder with multiple kernel learning (DAEMKL) was presented (Zhou et al., 2021). In this study, a creative computational model named bidirectional generative adversarial network (BGANMDA) is proposed to predict potential pairwise miRNA-disease associations. More specifically, similarity information networks were initially constructed from comprehensive similarity characteristics. Then, a similarity feature network with the full underlying relationships between miRNA and disease pairs was obtained by integrating all the similarity information. The whole integrated similarity network was loaded into the BGANMDA which employed an encoder to learn high-level features in latent space, a generator to produce a brand-new correlation between miRNAs and diseases, and a discriminator to decide whether the predicted associations were real. In this study, five-fold cross-validation and leave-one-out cross-validation (LOOCV) were adopted to evaluate the model's prediction performance. During the five-fold cross-validation, the BGANMDA acquired an area under the curve (AUC) of 0.9319 and a standard deviation of 0.0021 was obtained. At the same time, in the global and local leave-one-out cross-validation (LOOCV), the AUC value and standard deviation of BGANMDA were 0.9116 ± 0.0025 and 0.8928 ± 0.0022 , respectively. Furthermore, BGANMDA was employed in three different case studies to validate its prediction capability and accuracy. The experimental results of the case studies showed that 46, 46, and 48 of the top 50 prediction lists had been identified in previous studies.

2 MATERIALS AND METHODS

2.1 Human miRNA-Disease Associations

The Human MiRNA Disease Database (HMDD V3.0) was adopted as a benchmark dataset (Huang et al., 2019), which can be directly downloaded for experimental verification of disease-related miRNA information from <http://www.cuilab.cn/hmdd> (version v 3.2, published on 27 Mar 2019). After erasing the pairwise miRNA-disease associations that did not have IDs or lacked traits, duplicate samples describing the miRNA-disease

relationships were removed based on experimental support. In the process, 18,733 miRNA-disease associations were obtained, including 1,208 miRNAs and 985 diseases in the HMDD v 3.2 database. Based on the sorted dataset, we constructed an association binary matrix BM , consisting of 984 rows and 1,207 columns, was constructed to maintain the interaction information between miRNAs and diseases, which has 984 rows and 1,207 columns. If an experimentally verified miRNA-disease association was detected, the element value at the corresponding position of the matrix was set to 1; otherwise, it was set to 0.

2.2 Multi-Source Similarity Information for miRNAs and Diseases

2.2.1 MiRNA Similarity Network

To calculate the network of miRNA sequence similarity, the miRBase database containing almost all the miRNA sequence information, (Kozomara and Griffiths-Jones, 2014), was downloaded from <https://www.mirbase.org>, as shown in **Figure 1A**. The similarities of any two miRNAs were quantified using the Levenshtein distance, which represented the minimum cost of converting a single string to another string after replacing, inserting, and deleting one letter. The editing penalty was set to 2, while the deleted and inserted penalties were set to 1. Let $MSS(m_i, m_j)$ be the miRNA similarity score, where m_i denotes the i th miRNA and m_j denotes the j th miRNA, and the definition formula is shown as follows:

$$MSS(m_i, m_j) = 1 - \frac{x}{Len(m_i) - Len(m_j)}, \quad (1)$$

where x denotes the minimum penalty and $Len(m)$ is the sequence length of miRNA.

Then, we applied the max-min normalization to rescale and normalize MSS as follows:

$$MSS' = \frac{MSS(m_i, m_j) - MSS_{min}}{MSS_{max} - MSS_{min}}, \quad (2)$$

where MSS_{max} and MSS_{min} represent the maximum and minimum MSS , respectively.

Based on the universally acknowledged conjecture that pathologically similar diseases are more likely to be relevant to functionally similar miRNAs (and vice versa), a popular strategy (Zou et al., 2016) was employed to calculate the functional similarity (MFS) for a miRNA-miRNA pair m_i and m_j as follows:

$$MFS = \frac{\sum_{d \in D(m_i) \cap D(m_j)} (DSS(d, D(m_i)) + DSS(d, D(m_j)))}{|D(m_i)| + |D(m_j)|}, \quad (3)$$

where $D(m_i)$ and $D(m_j)$ denote the disease sets linked with m_i and m_j , respectively, and $|D(\cdot)|$ is the cardinal number of disease set. $DSS(d, D(m_j))$ is calculated through the max value of disease semantic similarity.

In this study, the sequence and functional similarity of miRNAs were employed to describe the miRNA similarity

characteristics. However, an independent similarity descriptor is not capable of accurately expressing the global similarity between any two miRNAs. Therefore, a similarity factor was set to determine which similarity descriptor can better represent similar features of miRNAs. Finally, a miRNA similarity network $MSN(m_i, m_j)$ between miRNA m_i and m_j was obtained by synthesizing the two similarity descriptors as follows:

$$MSN(m_i, m_j) = \alpha MFS(m_i, m_j) + (1 - \alpha) MSS(m_i, m_j), \quad (4)$$

where α is the similarity factor. In order to better describe the similarity characteristics between mirnas, we carefully compared the differences in miRNA functional similarity and sequence similarity in the description of similarity characteristics. Referring to previous literature (Lei et al., 2021), it was found that mirnas with similar functions were more likely to cause similar diseases. Experimental comparison shows that the optimal miRNA similarity network can be constructed when the value of α is set to 0.6.

2.2.2 miRNA GIP Kernel Similarity

Based on the previous study, the Gaussian interaction profile can be applied to capture the topological features of interaction networks, which is a successful method to evaluate the nuclear similarity of biomolecules for each pair in computational biology (Chen et al., 2016; Bao et al., 2017). Therefore, it is of great importance to search for GIP kernel similarity in the binary matrix with miRNA-disease interaction information. The GIP kernel similarity of microRNA $MGS_d(m_i, m_j)$ between m_i and m_j can be defined as follows:

$$MGS_d(m_i, m_j) = \exp(-\lambda \|BM(m_i), BM(m_j)\|^2), \quad (5)$$

where $BM(m_i)$ and $BM(m_j)$ are the i th and j th columns miRNA kernel information of each disease in binary matrix BM , respectively, λ represents a parameter that controls the kernel boundary's width, which can be calculated based on λ' as follows:

$$\lambda = \frac{\lambda'}{\frac{1}{nm} \sum_{i=1}^{nm} \|BM(m_i)\|^2}, \quad (6)$$

where nm denotes the number of miRNAs, λ' is the original bandwidth and set to 1 as suggested in other studies (Xiao et al., 2019).

2.2.3 Disease Similarity Network

As described in previous studies (Xuan et al., 2013), by using the medical subject headings disease descriptors (MeSH), downloaded from the NCBI website (<https://www.ncbi.nlm.nih.gov/>), it was possible to estimate the semantic similarity of diseases based on directed acyclic graph (DAG) structures. Specifically, a disease d can be described as $DAG_d = (d, G(d), E(d))$, in which $G(d)$ denotes the disease d itself and all the node of its ancestors, $E(d)$ is the corresponding edge set that contains the direct relationships from parents to child nodes in DAGs. Therefore, we computed the semantic contribution of disease d_k and d_t as follows:

$$C1_{d_t}(d_k) = \begin{cases} \max\{\mu \cdot C1_{d_t}(d'_k)\}, & \text{if } d_k \neq d_t \\ 1, & \text{if } d_k = d_t \end{cases}, \quad (7)$$

where μ is the semantic contribution factor, as suggested by recent study we set it as 0.5 (Wang et al., 2010). Then, the disease semantic value d_t is defined as

$$SV1(d_t) = \sum_{d_k \in G(d_t)} C1_{d_t}(d_k). \quad (8)$$

Let $DSS1 \in R^{nd \times nd}$ be the pairwise disease semantic similarity, which can be computed as follows:

$$DSS1(d_i, d_j) = \frac{\sum_{x \in G(d_i) \cap G(d_j)} (C1_{d_i}(x) + C1_{d_j}(x))}{SV1(d_i) + SV1(d_j)}, \quad (9)$$

where $G(d_i)$ and $G(d_j)$ represent the disease d_i and d_j in DAGs, respectively, and $DSS1$ is the first matrix to store the disease semantic similarity.

Furthermore, because diseases are more common when they appear in more DAGs, they are more specific when they appear in fewer GAGs, and in the same DAG layer, the diseases' semantic contribution value is almost different. Therefore, based on previous research (Pasquier and Gardès, 2016), another measurement was applied to obtain the semantic similarity of diseases as follows:

$$C2_{d_t}(d_k) = -\log\left(\frac{NG(d_k)}{nd}\right), \quad (10)$$

where $NG(d_k)$ is the number of DAGs including d_k .

Similarly, both the disease semantic value d_t and the pairwise disease semantic similarity $DSS2(d_i, d_j)$ can be described as follows:

$$SV2(d_t) = \sum_{d_k \in G(d_t)} C2_{d_t}(d_k) \quad (11)$$

$$DSS2(d_i, d_j) = \frac{\sum_{x \in G(d_i) \cap G(d_j)} (C2_{d_i}(x) + C2_{d_j}(x))}{SV2(d_i) + SV2(d_j)}, \quad (12)$$

where $DSS2$ is the second matrix to store the semantic similarity of diseases.

To obtain a more persuasive semantic similarity of diseases, a disease similarity network $DSN(m_i, m_j)$ was constructed between disease d_i and d_j by coalescing the two semantic disease similarities as follows:

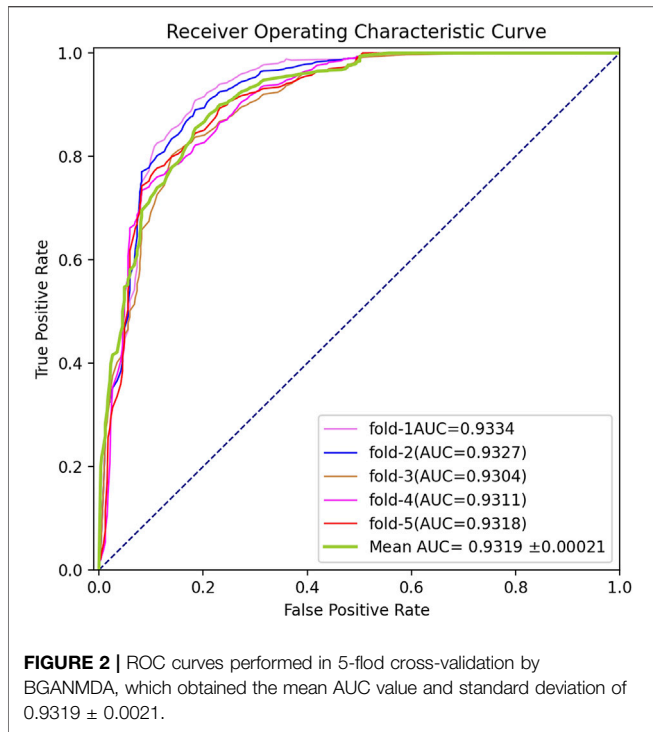
$$DSN(d_i, d_j) = \frac{DSS1(d_i, d_j) + DSS2(d_i, d_j)}{2}. \quad (13)$$

2.2.4 Disease GIP Kernel Similarity

Correspondingly, the GIP kernel similarity of diseases can be established as follows:

$$DGS_m(d_i, d_j) = \exp(-\lambda \|BM(d_i), BM(d_j)\|^2) \quad (14)$$

$$\lambda = \frac{\lambda'}{\frac{1}{nd} \sum_{i=1}^{nd} \|BM(d_i)\|^2}, \quad (15)$$



where $BM(d_i)$ and $BM(d_j)$ represent the i th and j th rows disease kernel information of each miRNA in binary matrix BM , respectively, λ represents a parameter that controls the width of the kernel boundary, and nd is the number of diseases.

2.4 Integrated Similarity Characteristic

Based on the above section, the miRNAs (diseases) similarity network and the GIP kernel similarity of miRNAs (diseases) were collected to acquire comprehensive similarity information. Considering that many sparse values may exist in the above-mentioned similarity network, the GIP kernel similarity was integrated with the similarity networks of miRNAs and diseases using the following formulas:

$$MS(m_i, m_j) = \begin{cases} MSN(m_i, m_j), & \text{if similarity between} \\ & m_i \text{ and } m_j \text{ are exist} \\ MGS_d(m_i, m_j), & \text{otherwise} \end{cases}, \quad (16)$$

$$DS(d_i, d_j) = \begin{cases} DSN(d_i, d_j), & \text{if } d_i \text{ and } d_j \text{ have} \\ & \text{semantic similarity} \\ DGS_m(d_i, d_j), & \text{otherwise} \end{cases}. \quad (17)$$

In the integrated similarity network, the miRNA similarity vector of miRNA m_i stores the similarity values of all the miRNAs to m_i . At the same time, the similarity values of all other diseases to disease d_i are included in the similarity vector of d_i . Thus, all the similarity eigenvectors for the corresponding pairwise miRNA-disease were concatenated to create a long feature vector of size $nm + nd$, where the nm and nd represent the number of microRNAs and diseases, respectively. Overall, the $nm \times nd$

generated eigenvectors were taken as samples, each sample corresponding to a miRNA-disease pair. As shown in **Figure 1B**, the details of processing similarity for potential miRNA-disease prediction is displayed.

2.5 Bidirectional Generative Adversarial Network

The present study introduced a computational model named bidirectional generative adversarial network (BGANMDA), which combined a nonlinear auto-encoder (consisting of an encoder and a generator), and an optimal discriminator to complete the task of identifying the potential associations between miRNAs and diseases. The Framework of BGANMDA is shown in **Figure 1C** (See more details in **Supplementary Figure S2**). Generally, the encoder of BGANMDA maps the original data point x to the feature representation $E(x)$ in latent space. At the same time, z is captured to generate a new relationship between a miRNA and a disease $G(z)$ through hidden layers in the generator. Then, the BGANMDA discriminator discriminates both in the traditional data space, and in the joint data and latent space ($(x, E(x))$ versus $(G(z), z)$), where the output of encoder $E(x)$ and the input of generator z are the latent components.

In the model, the encoder $E: \Omega_X \rightarrow \Omega_Z$ included a distribution $P_E(Z|X) = \delta(Z - E(x))$, which maps the original data points x into a latent space of the generator. At the same time, the generator $G: \Omega_Z \rightarrow \Omega_{XZ}$ randomly extracts sampling noise from the latent space of the encoder to generate new miRNA-disease associations under the distribution $Q_G(X|Z) = \delta(X - G(z))$. To “fool” a discriminator perfectly, both of encoder and generator must learn to invert each other through the joint probability distribution, satisfying the following two properties:

$$(a) X \in \hat{\Omega}_X \wedge E(X) = Z \quad (b) Z \in \hat{\Omega}_Z \wedge G(Z) = X. \quad (18)$$

The discriminator will take (X, Z) as input from the latent space to forecast the deterministic relationship of miRNA-disease pairs under the distribution of $P_D(Y|X, Z)$. If only property (a) is satisfied, the discriminator can infer the source of input (X, Z) must be come from the encoder pair $(X, E(X))$, and the value of discriminator $D_{EG}^*(X, Z)$ is 1; if the source of (X, Z) only satisfies (b), it must be come from the generator pair $(G(Z), Z)$ and $D_{EG}^*(X, Z)$ is 0. Therefore, a minimax objective can be defined to displace the BGANMDA training objective as follows:

$$\min_{G,E} \max_D V(D, E, G), \quad (19)$$

where $V(D, E, G)$ can be represented by the following formulas:

$$\mathbb{E}_{X \sim p_X} [\log D(X, E(X))] + \mathbb{E}_{Z \sim p_Z} [\log (1 - D(G(Z), Z))] \quad (20)$$

and

$$\log D(X, E(X)) = \mathbb{E}_{Z \sim p_Z(\cdot|X)} [\log D(X, Z)], \quad (21)$$

$$\log (1 - D(G(Z), Z)) = \mathbb{E}_{X \sim p_X(\cdot|Z)} [\log (1 - D(X, Z))]. \quad (22)$$

When a discriminator input (X, Z) satisfies noth (a) and (b), E and G invert each other almost everywhere, that is $X = G(E(X))$

TABLE 1 | The AUC AUPR, F1-scores, recall, precision, and MCC of ten methods on miRNA-disease associations prediction task in five-fold cross validation.

	AUC	AUPR	F1-score	Recall	Precision	MCC
BGANMDA	0.9316	0.9237	0.9024	0.8968	0.9215	0.8913
NCMCMDBA	0.9187	0.9093	0.9031	0.8621	0.9033	0.8541
DBNMDA	0.8994	0.8972	0.8793	0.8517	0.8802	0.8632
GAEMDA	0.8674	0.8625	0.8609	0.8485	0.8756	0.8427
TDRCMDA	0.8713	0.8706	0.8569	0.8990	0.8783	0.8409
SSCMDA	0.8531	0.8654	0.8542	0.8501	0.8697	0.8371
CNNMDA	0.8494	0.8583	0.8512	0.8453	0.8724	0.8347
HDMP	0.8226	0.8407	0.8373	0.8486	0.8531	0.8287
KADZMDA	0.8317	0.8351	0.8219	0.8312	0.8418	0.8216
WBSMDA	0.8163	0.8216	0.8097	0.8277	0.8391	0.8163

Bolded numbers are the best performance in each category.

and $Z = E(G(Z))$. Compared with other advanced miRNA-disease predicting models, the BGANMDA here employed focuses more on processing complex data and effectively learning the gradient information to ensure the correct allocation of parameter weights.

In preprocessed similarity eigenvectors, each disease (miRNA) contains the similarity information of all miRNAs (diseases), which integrate the miRNAs similarity network, diseases similarity network, and GIP kernel similarity of miRNAs (diseases). As mentioned above, BGANMDA is an innovative computational model inspired by a nonlinear auto-encoder. A BGANMDA encoder is one of the two parts of an auto-encoder, and it shows a strong performance in terms of compressing complex data, eliminating extra noise, and learning the additional features of latent space. The eigenvector samples are used as encoder input and the parameters of the similarity vectors can be calculated through three fully connected layers of the neural network. Besides, LeakyReLU was applied as an activation function for each network layer with dropout to accelerate the convergence rate and prevent the occurrence of over-fitting. It assigns a non-zero slope to all negative values, which can accelerate gradient descent and better carry out backpropagation. The function is defined as follows:

$$LeakyReLU = \begin{cases} x, & x \geq 0 \\ \eta x, & x < 0 \end{cases}, \quad (23)$$

where η , which is set to 0.01, denotes a fixed learning parameter.

In the BGANMDA encoder, the dimensions of the similarity eigenvector samples between miRNA and disease are compressed into low-dimensional vectors when passing through network layers. The dense information of compressed low-dimensional vectors allows the model to learn how to map the miRNA-disease relationship into latent space. In this way, a trained encoder can precisely identify the feature representation by capturing semantic attributes, in order to obtain a data pair $(x, E(x))$. To better understand the further representation of latent space, the number of neurons in an encoder output layer was set to 100. At the same time, the binary cross-entropy was used as the loss function, as shown in the following equation:

$$Loss_p(q) = -\frac{1}{\Gamma} \sum_{i=1}^{\Gamma} y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)), \quad (24)$$

where Γ represents the output size of BGANMDA, y is the label (1 for known miRNA-disease pairs), and $p(y)$ is the predicted probability of the association between miRNA and disease.

In most generative adversarial network models, generators always play a role in studying the features of original data to generate new data based on the learned characteristics. However, in the BGANMDA, the generator was preferentially used to select a random sample as input instead of the original one. The structure of the generator network is similar to that of the encoder which has three fully connected layers with dropout. The generator output was calculated as follows:

$$G(Z) = W^G z + b^G, \quad (25)$$

where z is the sampling noise from the encoder latent space, and W^G and b^G represent the weights and bias of the generator, respectively.

It is noteworthy that each layer in the BGANMDA generator increases the dimension of the potential representation layer by layer and the final output dimension is the same as the encoder input. The sampling noise representation of the miRNA-disease association is decoded by the generator, then the new associations are generated. As a result, a series of data pairs $(G(z), z)$ is obtained.

The data pairs $(x, E(x))$ and $(G(z), z)$ are taken as inputs to try to fool the discriminator. Initially, if the data pair derives from the encoder, the discriminator can easily recognize the input source and discriminate it as real, namely $D_{EG}^*(X, Z)$ is set to 1; whereas, if the data pair derives from the generator, the $D_{EG}^*(X, Z)$ is set to 0. As the model comprehensively analyzes the underlying features of miRNA-disease relationships, the encoder and the generator learn to convert each other. It becomes difficult for the discriminator to distinguish the source of the input, so that we can obtain predictions that are more representative of reality. The sigmoid function was employed to calculate the final probability of miRNA and disease pairs, which is defined as follows:

$$sigmoid(\theta) = \frac{1}{1 - exp(-\theta)}, \quad (26)$$

where θ denotes the sigmoid function input.

This BGANMDA encoder, with its excellent representation ability, can learn the potential association of miRNA-disease pairs. The generator can extract the features from the sampled noise latent space and generate new miRNA-disease associations. Finally, $x = G(E(x))$ and $z = E(G(z))$ are almost everywhere through a union probability distribution to obtain a bidirectional structure. The experimental results of the present study reveal that BGANMDA is robust and has a strong representational learning ability to predict potential miRNA-disease associations and, compared with other state-of-the-art methods, it performs remarkably well.

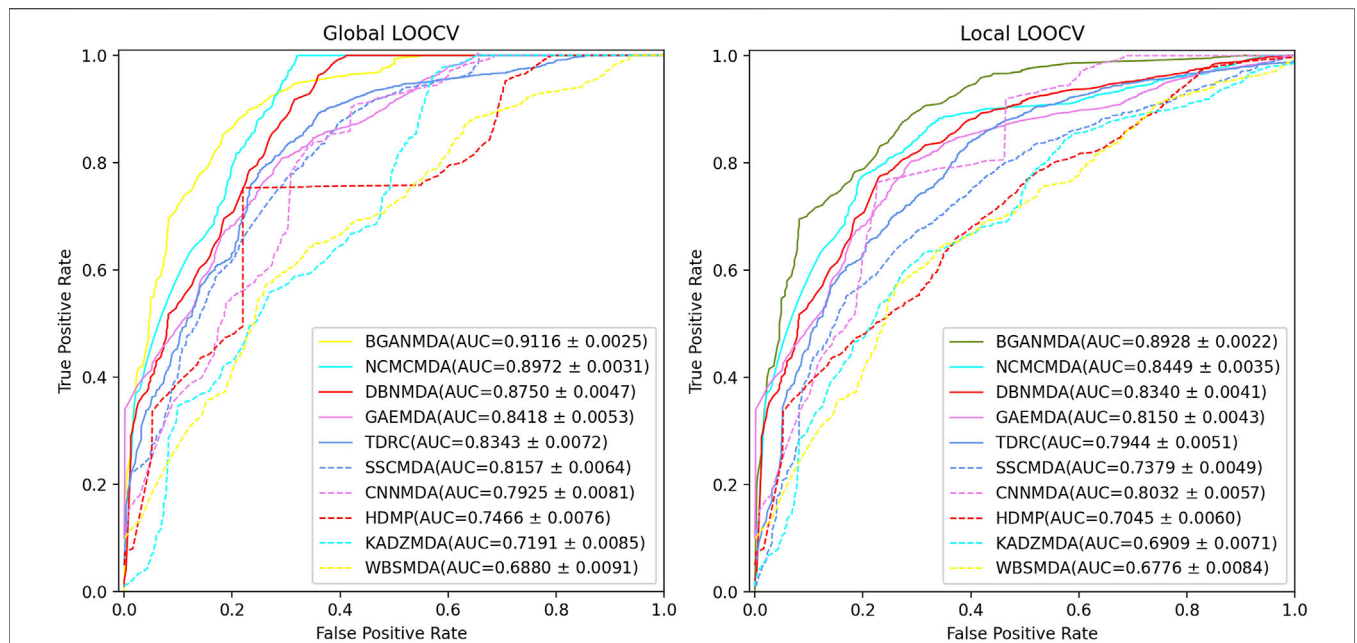


FIGURE 3 | The performance of BGANMDA and the other nine disease-related miRNA prediction methods, namely NCMCMDA, DBNMDA, GAEMDA, TDR, SSCMDA, CNNMDA, HDMP, KADZMDA and WBSMDA, were compared based on the ROC curve and the AUC value in global and local LOOCV. BGANMDA obtained the AUC and standard deviation of 0.9116 ± 0.0025 in global LOOCV and 0.8928 ± 0.0022 in local LOOCV, respectively.

3 RESULTS

3.1 Performance Evaluation

In this study, we implemented BGANMDA based on the structure of combined generative adversarial networks with auto-encoder. The input size of BGANMDA encoder is 2,191, and the output size is 100. The generator in our model has the same network structure with encoder, its input size and output size are 100 and 2,191, respectively. The data pair $(x, E(x))$ or $(G(z), z)$ was concatenated and fed to the BGANMDA discriminator, which has the dimension 2,291 of input size. We adopted Adam as a gradient descent algorithm to optimize parameters and the learning rate was fixed at $2e-4$. To avoid over-fitting, the cross-entropy function and LeakyReLU (Wang et al., 2018) were used as the loss function and activation function, respectively. The suitable dropout rate strategy ranged from 0.1 to 0.9, and we set it at 0.5 after the effective validation. To train the model, 20% of the evaluated samples in training sets were randomly removed five times, and the epochs and batch size were set at 20,000 and 128 each time.

To systematically measure the prediction capability of BGANMDA, three different validation methods were employed, namely five-fold cross-validation, global LOOCV, and local LOOCV. In the first, the acknowledged miRNA-disease samples were stochastically split into five subsets, where each was considered as the dataset for testing, and the others were treated as training sets. The BGANMDA model was used to prioritize the unverified miRNA-disease candidates and test samples according to the score given obtained. In order to ensure the reliability of the evaluation results, the five-fold cross-validation was repeated 150 times to reach the

distribution of original samples. As shown in Figure 2, our model calculated a mean AUC and standard deviation of 0.9116 ± 0.00021 . Meanwhile, we compared with other state-of-the-art models, namely the NCMCMDA, DBNMDA, GAEMDA, TDR, SSCMDA, CNNMDA, HDMP, KATZMDA, and WBSMDA via different evaluation index. As shown in Table 1, BGANMDA obtains the values of AUPR (0.9237), F1-score (0.9024), Recall (0.8968), Precision (0.9215), and MCC (0.8913), which outperforms than other advanced models in 5-fold cross validation. Noting that the miRNA (diseases) similarity network and GIP kernel similarity in 5-fold cross-validation process, because the correlation binary matrix was altered when part of the known association of miRNA-disease pairs was removed.

In LOOCV, the distinction between global and local LOOCV depends on whether the whole miRNA-disease information is included simultaneously specifically, in global LOOCV, the focus was on all underlying pairwise miRNA-disease correlations and each known association between a specific disease and miRNA as test sample was excluded in turn. In contrast, in local LOOCV, only the unknown miRNA-disease associations concerned in test samples were ranked by comparing the association scores. In the rank list, a threshold was given: samples with miRNA-disease association prediction scores above the threshold were considered true positives (TP). For each given threshold, it was possible to find the corresponding TP to determine the true positive ratio (sensitivity). Similarly, the false negatives among the candidate samples could be obtained by setting a threshold, and the corresponding false-positive ratio (1-specificity). The term sensitivity denotes the ratio of the test sample ranking over a specific threshold, whereas specificity indicates the proportion of

TABLE 2 | Based on the known associations provided by dbDEMC and HMDD v3.2, the top 50 miRNAs related to colon neoplasm were predicted by employing BGANMDA model, and 46 predictions were confirmed based on dbDEMC and miR2Disease. The first column records the top 1–25 predicted potentially related miRNAs and the third column records the 26–50 predicted potentially relevant miRNAs.

miRNA	Evidence	miRNA	Evidence
hsa-mir-145	dbDEMC; HMDD v3.2	hsa-mir-139-5p	dbDEMC; HMDD v3.2
hsa-mir-21	dbDEMC	hsa-let-7c	dbDEMC; HMDD v3.2
hsa-mir-143	dbDEMC; HMDD v3.2	hsa-mir-96	dbDEMC; HMDD v3.2
hsa-mir-195	dbDEMC; HMDD v3.2	hsa-mir-106a	dbDEMC; HMDDv3.2
hsa-mir-502-5p	dbDEMC	hsa-mir-628-3p	HMDD v3.2
hsa-mir-215	dbDEMC; HMDD v3.2	hsa-mir-210	dbDEMC
hsa-mir-503	dbDEMC	hsa-mir-140-5p	dbDEMC
hsa-mir-100	dbDEMC	hsa-mir-20a	dbDEMC
hsa-mir-155	dbDEMC	hsa-mir-28-5p	HMDD v3.2
hsa-mir-497	dbDEMC; HMDD v3.2	hsa-mir-342-3p	dbDEMC; HMDD v3.2
hsa-mir-548d-3p	dbDEMC; HMDD v3.2	hsa-mir-556-5p	dbDEMC
hsa-mir-150	unconfirmed	hsa-mir-23a	dbDEMC
hsa-mir-552	dbDEMC; HMDD v3.2	hsa-mir-93	dbDEMC; HMDD v3.2
hsa-mir-650	HMDD v3.2	hsa-mir-133b	HMDD v3.2
hsa-mir-491-5p	dbDEMC	hsa-mir-518b	dbDEMC; HMDD v3.2
hsa-mir-183	dbDEMC	hsa-mir-581	unconfirmed
hsa-mir-30a	HMDD v3.2	hsa-mir-421	dbDEMC; HMDD v3.2
hsa-mir-182	dbDEMC; HMDD v3.2	hsa-mir-192a	dbDEMC
hsa-mir-378	HMDD v3.2	hsa-mir-32-3p	dbDEMC
hsa-mir-34a	unconfirmed	hsa-mir-18a	dbDEMC; HMDD v3.2
hsa-mir-17	dbDEMC	hsa-mir-330-3p	dbDEMC; HMDD v3.2
hsa-mir-665	dbDEMC	hsa-mir-203	dbDEMC; HMDD v3.2
hsa-mir-155-3p	dbDEMC; HMDD v3.2	hsa-mir-583	HMDD v3.2
hsa-mir-623	HMDD v3.2	hsa-mir-889	unconfirmed
hsa-mir-486-5p	HMDD v3.2	hsa-mir-10b	dbDEMC; HMDD v3.2

unknown miRNA-disease association samples ranked under this threshold. Based on the results obtained, the receiver operating characteristic curve was drawn and the AUC at different thresholds was calculated. The higher the AUC value, the better the model performance; if it was close to 1, the BGANMDA was considered an excellent predictor of potential miRNA-disease correlations, whereas the association was regarded as a random prediction if the AUC value was close to 0.5. To comprehensively verify BGANMDA’s predictive ability for complex disease-related miRNAs, we compared it with other nine models for global and local LOOCV. As shown in Figure 3, BGANMDA acquired the AUC and standard deviation of 0.9116 ± 0.0025 in global LOOCV and 0.8928 ± 0.0022 in local LOOCV, respectively. Obviously, our model superior to other advanced methods, which recorded values of 0.8972 ± 0.0031 , 0.8750 ± 0.0047 , 0.8418 ± 0.0053 , 0.8343 ± 0.0072 , 0.8157 ± 0.0064 , 0.7925 ± 0.0081 , 0.7466 ± 0.0076 , 0.7191 ± 0.0085 , and 0.6880 ± 0.0091 in global LOOCV. However, in local LOOCV, the CNNMDA (0.8032 ± 0.0057) obtained a higher AUC value, while the other methods recorded lower ones, as observed for the NCMCMA (0.8449 ± 0.0035), DBNMDA (0.8340 ± 0.0041), GAEMDA (0.8150 ± 0.0043), TDRC (0.7944 ± 0.0051), SSCMDA (0.7379 ± 0.0049), HDMP (0.7045 ± 0.0060), KADZMDA (0.6909 ± 0.0071), and WBSMDA (0.6776 ± 0.0084). We assumed that CNNMDA could outperform other methods in local LOOCV because of the effective pattern of convolutional feature extracting. NCMCMA integrated neighborhood constraint with matrix completion, aiming at transforming the task of recovering the

missing miRNA-disease associations into an optimization problem. As can be seen, though NCMCMA has a good performance in global and local LOOCV, the insufficiency of present known miRNA-disease associations which NCMCMA strongly depends on will limit the prediction performance. Compared with the other advanced models, BGANMDA has the advantage of excellent learning capability for underlying similar traits, making it inclusive and robust to the lack of unknown miRNA-disease associations.

3.2 Case Studies

In order to further evaluate the practical capability of the BGANMDA to predict the potential disease-related miRNAs, three different case studies of fatal cancers were considered, namely the neoplasms of the colon, esophagus, and kidney. Here, the known information on the associations between microRNAs and diseases obtained from the HMDD v 3.2 databases was used as the positive samples for the BGANMDA, and the miRNAs were prioritized based on the predicted score given by the model. Then, the top 50 forecast candidates were ranked based on the dbDEMC database (Yang et al., 2017) and HMDD v3.2 database.

Colon neoplasm is a dangerous malignant tumor causing a high mortality rate in humans, and its morbidity rates are only second to those of esophageal and gastric cancers (Brody, 2015; Ji et al., 2018). Studies have predicted the occurrence of 101,420 new colon neoplasm cases, representing 8.3% of all new cancer cases reported in the United States in 2019, which will result in the death of 51,200 people (Siegel et al., 2019). Thus, it is urgent to

TABLE 3 | Based on the validated associations provided by dbDEMC and HMDD v3.2, the top 50 miRNAs related to kidney neoplasm were predicted by employing BGANMDA model, and 48 predictions were confirmed based on dbDEMC and miR2Disease. The first column records the top 1–25 predicted potentially relevant miRNAs and the third column records the 26–50 predicted potentially related miRNAs.

miRNA	Evidence	miRNA	Evidence
hsa-mir-21	dbDEMC; HMDD v3.2	hsa-mir-21–5p	dbDEMC; HMDD v3.2
hsa-mir-429	dbDEMC	hsa-mir-548d-3p	Unconfirmed
hsa-mir-299–5p	dbDEMC	hsa-mir-30c-2-3p	dbDEMC; HMDD v3.2
hsa-mir-200c	dbDEMC	hsa-mir-30a-5p	dbDEMC; HMDDv3.2
hsa-mir-204	dbDEMC; HMDD v3.2	hsa-mir-513c-5p	dbDEMC
hsa-mir-1293	HMDD v3.2	hsa-mir-584–3p	dbDEMC; HMDD v3.2
hsa-mir-184	dbDEMC; HMDD v3.2	hsa-mir-20b	HMDD v3.2
hsa-mir-193a-3p	dbDEMC; HMDD v3.2	hsa-mir-18a	HMDD v3.2
hsa-mir-210	unconfirmed	hsa-mir-144–5p	HMDD v3.2
hsa-mir-211–5p	dbDEMC; HMDD v3.2	hsa-mir-244–5p	HMDD v3.2
hsa-mir-199a-5p	dbDEMC	hsa-mir-106b	dbDEMC; HMDD v3.2
hsa-mir-532–5p	dbDEMC; HMDD v3.2	hsa-mir-133b	dbDEMC; HMDD v3.2
hsa-mir-433–3p	dbDEMC; HMDD v3.2	hsa-mir-483–5p	dbDEMC
hsa-mir-206	HMDD v3.2	hsa-mir-580–5p	dbDEMC; HMDD v3.2
hsa-mir-489–5p	HMDD v3.2	hsa-mir-484	dbDEMC
hsa-mir-660–5p	dbDEMC	hsa-mir-363	dbDEMC
hsa-mir-3654	dbDEMC; HMDD v3.2	hsa-mir-93	dbDEMC; HMDD v3.2
hsa-mir-196a-3p	unconfirmed	hsa-mir-342–3p	dbDEMC; HMDD v3.2
hsa-mir-320b	dbDEMC; HMDD v3.2	hsa-mir-215	dbDEMC; HMDD v3.2
hsa-mir-199b-5p	dbDEMC; HMDD v3.2	hsa-mir-194	dbDEMC
hsa-mir-301b-3p	dbDEMC	hsa-mir-6843–3p	HMDD v3.2
hsa-mir-199b-3p	dbDEMC	hsa-mir-496	HMDD v3.2
hsa-mir-17–3p	dbDEMC; HMDD v3.2	hsa-mir-30e-3p	HMDD v3.2
hsa-mir-199a-3p	HMDD v3.2	hsa-mir-415a	dbDEMC; HMDD v3.2
hsa-mir-676–5p	dbDEMC; HMDD v3.2	hsa-mir-185	dbDEMC; HMDD v3.2

develop sensitive and novel biomarkers that can effectively and timely detect colon neoplasms. Studies have confirmed that miRNAs are becoming a crucial target for colon tumor prevention, diagnosis, and therapy. Some research revealed that the abundance of hsa-mir-145 is negatively correlated with its expression in colon neoplasm cells (Wang et al., 2012), a relationship confirmed by luciferase reporter assay. In addition (Yu et al., 2018), identified the association between has-mir-21–5p and the clinicopathological features of colon adenocarcinoma (CODA) patients, as well as its overexpression in CODA cells. The experiments conducted revealed that has-mir-21–5p promoted the migration, and proliferation of colon neoplasm cells, and invasion of tissues, by inhibiting CHL1 expression. It was also shown that the overexpression of has-mir-143 raised oxaliplatin-induced apoptosis relevant to oxygen generation (Gomes et al., 2018). This suggests that has-mir-143 may bypass oxaliplatin resistance in the cells of HCT116 human colon neoplasm by increasing oxidative stress. In this study, the BGANMDA was used to determine the potential miRNAs associated with colon neoplasms and 100, 90, and 92%, were confirmed in the top 10, 20, and 50, respectively, based on the dbDEMC and HMDD v3.2 (see **Table 2**).

In order to illustrate the capability of the BGANMDA to predict diseases with unconfirmed miRNAs, the esophageal neoplasm was selected as a concrete example, based on the HMDD v 3.2 database. More specifically, the validated disease-related microRNAs of this tumor were omitted and it was considered a new disease. Hence, the model only extracted the associated miRNAs of other diseases and all the miRNA-disease similarity information to train its prediction ability. As a dangerous and high-incidence tumor worldwide, the

etiology of esophageal neoplasms is associated with inflammation, chronic nitrosamine stimulation, and content of microelements in edibles (Kollarova et al., 2007). A number of studies have revealed that hsa-mir-133b may be a potential therapeutic target for esophageal squamous cell carcinoma. Its overexpression can inhibit the MAPK/ERK and PI3K/AKT signaling pathways by regulating epidermal growth factor receptors to suppress the proliferation and migration, of esophageal squamous carcinoma, and invasion of tissues cells (Zeng et al., 2019). Thus, detecting the existing miRNA biomarkers is of great importance to discover esophageal neoplasm cases. Some studies indicated that hsa-mir-17–5p is a crucial biomarker to predict the response to neoadjuvant chemoradiation therapy in esophageal adenocarcinoma (EAC), which would help improve patient stratification and serve as a new therapeutic target to boost the efficacy of this therapy in EAC (Lynam-Lennon et al., 2017). After the implementation of BGANMDA resulted in values of 100, 95 and 92% in the top 10, 20, and 50 potential miRNAs related to esophageal neoplasm based on the dbDEMC and HMDD v3.2 databases (**Supplementary Table S3**).

To evaluate the stability of the model performance using various data sources, all the known kidney neoplasm-related miRNA associations and similarity information (obtained from the HMDD v 2.0 database) database and unverified pairwise miRNA-disease associations were used to train its prediction ability. Then, the prediction scores obtained from the model were verified using the HMDD v 3.2 and dbMDEMC databases, and the literature. Kidney neoplasm, also known as renal cancer, is recognized as one of the top 10 frequent diseases, with over

250,000 unheard cases verified each year. Therefore, finding the association between kidney neoplasm progression and the dysregulation of certain miRNAs can accelerate the prevention, diagnosis, and treatment of renal cancer, and reduce their costs (Shephard et al., 2013). Has-mir-429, the second-highest probability associated with kidney neoplasm predicted by the model, has been reported to be downregulated in contrast-induced acute kidney injury (CI-AKI). In this model, the overexpression of mir-429 reduced apoptosis and increased cell viability by targeting PDCD4 to inhibit the NF- κ B signaling pathway (Niu et al., 2021). According to previous research, hsa-mir-210 can directly target HIF-1 α and inhibit the HIF-1 α pathway by participating in the molecular response of hypoxic kidney injury *in vitro*, thus protecting renal tumor cells from hypoxia-induced apoptosis (Liu et al., 2017). In addition, hsa-mir-206 also suppresses kidney neoplasm carcinoma proliferation and epithelial-mesenchymal transformation by inhibiting CDK6 expression (Guo et al., 2020b). After training the stable capability of the BGANMDA, the results showed that the accuracy in the determination of potential miRNAs correlated with kidney neoplasm was 90, 95, and 96% in the top 10, 20, and 50, respectively, based on the dbDEMC and HMDD v3.2 databases, as shown in **Table 3**.

4 DISCUSSION AND CONCLUSION

Predicting the underlying associations of miRNA-disease pairs contributes to the understanding of disease mechanisms at the miRNA level, ultimately resulting in better prevention, diagnosis, and treatment. In the present study the model known as BGANMDA, based on auto-encoder and traditional generative adversarial networks, was proposed to determine the probability score of unknown miRNA-disease pairs by constructing the miRNAs (diseases) similarity network and GIP kernel similarity of miRNAs (diseases). The BGANMDA showed a superior performance compared to other advanced methods in three types of cross validation, which also reflected its stable capability. Furthermore, case studies of various diseases also confirmed that the model's predictions are reliable and accurate.

The model's successful performance of can be illustrated by the following factors. First, the similarity network was constructed from miRNAs (diseases) as traits to train the model. Second, the BGANMDA retained the advantages of both the auto-encoder and GAN, which can automatically recognize the comprehensive similarity characteristics of miRNAs and diseases, eliminate noise, and reduce dimensions. In addition, the model had an excellent performance in terms of learning the annotated biological patterns. Third, and most importantly, it achieved a bidirectional GAN structure, which means that the model's encoder mapped the data points x into latent space and the generator's sampling noise from the latent space to generate new miRNA-disease associations. Ultimately, the encoder and generator of BGANMDA can invert each other based on the joint probability distribution.

However, the model has some limitations. First, the parameter values proposed were set as default, so it was not possible to further consider whether the performance would be impacted.

Parameter settings play an important role in assisting the model to learn privileged information from the eigenvectors, particularly for complex associated features. Second, the model's strongly relied on similarity features, which were computed based on handcrafted measurements. Third, the information in each network layer could not be shared and propagated well, because the component of the auto-encoder was used for compressing the features into low dimensions and learning the latent representation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

LX, QY, and XKL conceived the study. LX, XKL, and YL developed the method. LX, QY, and LT implemented the algorithms and collected the data. QY and XKL performed the data analyses. LX and QYL wrote the manuscript. Both authors have read and approved the article.

FUNDING

This project was supported by the National Natural Science Foundation of China (Nos. 81,273,649, 61,501,132, and 61,672,181), the Natural Science Foundation of Heilongjiang Province (Nos. LH2019F049 and LH2019A029), the China Postdoctoral Science Foundation (No. 2019M650069), the Research Funds for the Central Universities (No. 3072019CFT0603), the Fund for Young Innovation Team of Basic Scientific Research in Heilongjiang Province (No. RCYJTD201805), the Heilongjiang Basic Scientific Research and Technological Innovation Fund (No. KJCX201805), the Young Eagles Plan (2020CYJBGX0057 and 2020CYJBGX0353), and the Foundation Items: Innovation fund for SMEs (No. 2017FF1GJ023). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

ACKNOWLEDGMENTS

The authors are grateful for the suggestions that helped to improve the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.936823/full#supplementary-material>

REFERENCES

- Ambros, V. (2003). MicroRNA Pathways in Flies and Worms. *Cell* 113, 673–676. doi:10.1016/s0092-8674(03)00428-8
- Bao, W., Jiang, Z., and Huang, D. S. (2017). Novel Human Microbe-Disease Association Prediction Using Network Consistency Projection. *BMC Bioinforma.* 18 (Suppl. 16), 543. doi:10.1186/s12859-017-1968-2
- Bartel, D. P. (2004). MicroRNAs. *Cell* 116, 281–297. doi:10.1016/s0092-8674(04)00045-5
- Brody, H. (2015). Colorectal Cancer. *Nature* 521, S1. doi:10.1038/521s1a
- Calin, G. A., and Croce, C. M. (2006). MicroRNA Signatures in Human Cancers. *Nat. Rev. Cancer* 6, 857–866. doi:10.1038/nrc1997
- Chen, X., Li, T. H., Zhao, Y., Wang, C. C., and Zhu, C. C. (2021a). Deep-belief Network for Predicting Potential miRNA-Disease Associations. *Brief. Bioinform* 22, bbaa186. doi:10.1093/bib/bbaa186
- Chen, X., Wang, L., Qu, J., Guan, N. N., and Li, J. Q. (2018). Predicting miRNA-Disease Association Based on Inductive Matrix Completion. *Bioinformatics* 34, 4256–4265. doi:10.1093/bioinformatics/bty503
- Chen, X., Sun, L.-G., and Zhao, Y. (2021b). NCMCMDA: miRNA-Disease Association Prediction through Neighborhood Constraint Matrix Completion. *Brief. Bioinform* 22, 485–496. doi:10.1093/bib/bbz159
- Chen, X., Xie, D., Zhao, Q., and You, Z.-H. (2019). MicroRNAs and Complex Diseases: from Experimental Results to Computational Models. *Brief. Bioinform* 20, 515–539. doi:10.1093/bib/bbx130
- Chen, X., Yan, C. C., Zhang, X., You, Z.-H., Deng, L., Liu, Y., et al. (2016). WBSMDA: Within and between Score for MiRNA-Disease Association Prediction. *Sci. Rep.* 6, 21106. doi:10.1038/srep21106
- Chen, X., and Yan, G.-Y. (2014). Semi-supervised Learning for Potential Human microRNA-Disease Associations Inference. *Sci. Rep.* 4, 5501. doi:10.1038/srep05501
- Cheng, A. M. (2005). Antisense Inhibition of Human miRNAs and Indications for an Involvement of miRNA in Cell Growth and Apoptosis. *Nucleic Acids Res.* 33, 1290–1297. doi:10.1093/nar/gki200
- Ebert, M. S., and Sharp, P. A. (2012). Roles for microRNAs in Conferring Robustness to Biological Processes. *Cell* 149, 515–524. doi:10.1016/j.cell.2012.04.005
- Gomes, S. E., Pereira, D. M., Roma-Rodrigues, C., Fernandes, A. R., Borralho, P. M., and Rodrigues, C. M. P. (2018). Convergence of miR-143 Overexpression, Oxidative Stress and Cell Death in HCT116 Human Colon Cancer Cells. *PLoS one* 13, e0191607. doi:10.1371/journal.pone.0191607
- Guo, M., Gan, L., Si, J., Zhang, J., Liu, Z., Zhao, J., et al. (2020a). Role of miR-302/367 Cluster in Human Physiology and Pathophysiology. *Acta Biochim. Biophys. Sin.* 52, 791–800. doi:10.1093/abbs/gmaa065
- Guo, Z., Jia, H., and Ge, J. (2020b). MiR-206 Suppresses Proliferation and Epithelial-Mesenchymal Transition of Renal Cell Carcinoma by Inhibiting CDK6 Expression. *Hum. Cell* 33, 750–758. doi:10.1007/s13577-020-00355-5
- He, X.-Y., Liao, Y.-D., Guo, X.-Q., Wang, R., Xiao, Z.-Y., and Wang, Y.-G. (2016). Prognostic Role of microRNA-21 Expression in Brain Tumors: a Meta-Analysis. *Mol. Neurobiol.* 53, 1856–1861. doi:10.1007/s12035-015-9140-3
- Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., et al. (2019). HMDD v3.0: a Database for Experimentally Supported Human microRNA-Disease Associations. *Nucleic Acids Res.* 47, D1013–D1017. doi:10.1093/nar/gky1010
- Ji, X., Peng, Q., and Wang, M. (2018). Anti-colon-cancer Effects of Polysaccharides: A Mini-Review of the Mechanisms. *Int. J. Biol. Macromol.* 114, 1127–1133. doi:10.1016/j.ijbiomac.2018.03.186
- 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
- Jones, K., Nourse, J. P., Keane, C., Bhatnagar, A., and Gandhi, M. K. (2014). Plasma microRNA Are Disease Response Biomarkers in Classical Hodgkin Lymphoma. *Clin. Cancer Res.* 20, 253–264. doi:10.1158/1078-0432.ccr-13-1024
- Karp, X., and Ambros, V. (2005). Encountering MicroRNAs in Cell Fate Signaling. *Science* 310, 1288–1289. New York, N.Y. doi:10.1126/science.1121566
- Kollarova, H., Machova, L., Horakova, D., Janoutova, G., and Janout, V. (2007). Epidemiology of Esophageal Cancer - an Overview Article. *Biomed. Pap.* 151, 17–28. doi:10.5507/bp.2007.003
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: Annotating High Confidence microRNAs Using Deep Sequencing Data. *Nucl. Acids Res.* 42, D68–D73. doi:10.1093/nar/gkt1181
- Krützfeldt, J., and Stoffel, M. (2006). MicroRNAs: a New Class of Regulatory Genes Affecting Metabolism. *Cell. Metab.* 4, 9–12. doi:10.1016/j.cmet.2006.05.009
- Lei, X., Mudiyansele, T. B., Zhang, Y., Bian, C., Lan, W., Yu, N., et al. (2021). A Comprehensive Survey on Computational Methods of Non-coding RNA and Disease Association Prediction. *Brief. Bioinform* 22, bbaa350. doi:10.1093/bib/bbaa350
- Li, C., Liu, H., Hu, Q., Que, J., and Yao, J. (2019). A Novel Computational Model for Predicting microRNA-Disease Associations Based on Heterogeneous Graph Convolutional Networks. *Cells* 8, 977. doi:10.3390/cells8090977
- Li, G., Luo, J., Xiao, Q., Liang, C., and Ding, P. (2018). Predicting microRNA-Disease Associations Using Label Propagation Based on Linear Neighborhood Similarity. *J. Biomed. Inf.* 82, 169–177. doi:10.1016/j.jbi.2018.05.005
- Li, J., Zhang, S., Liu, T., Ning, C., Zhang, Z., and Zhou, W. (2020). Neural Inductive Matrix Completion with Graph Convolutional Networks for miRNA-Disease Association Prediction. *Bioinformatics* 36, 2538–2546. doi:10.1093/bioinformatics/btz965
- Li, L., Gao, Z., Wang, Y.-T., Zhang, M.-W., Ni, J.-C., Zheng, C.-H., et al. (2021a). SCMFMDA: Predicting microRNA-Disease Associations Based on Similarity Constrained Matrix Factorization. *PLoS Comput. Biol.* 17, e1009165. doi:10.1371/journal.pcbi.1009165
- Li, Z., Li, J., Nie, R., You, Z. H., and Bao, W. (2021b). A Graph Auto-Encoder Model for miRNA-Disease Associations Prediction. *Brief. Bioinform* 22, bbaa240. doi:10.1093/bib/bbaa240
- Liu, L.-L., Li, D., He, Y.-L., Zhou, Y.-Z., Gong, S.-H., Wu, L.-Y., et al. (2017). miR-210 Protects Renal Cell against Hypoxia-Induced Apoptosis by Targeting HIF-1 Alpha. *Mol. Med.* 23, 258–271. doi:10.2119/molmed.2017.00013
- Lynam-Lennon, N., Heavey, S., Sommerville, G., Bibby, B. A. S., Ffrench, B., Quinn, J., et al. (2017). MicroRNA-17 Is Downregulated in Esophageal Adenocarcinoma Cancer Stem-like Cells and Promotes a Radioresistant Phenotype. *Oncotarget* 8, 11400–11413. doi:10.18632/oncotarget.13940
- Ma, Y., He, T., Ge, L., Zhang, C., and Jiang, X. (2019). MiRNA-disease Interaction Prediction Based on Kernel Neighborhood Similarity and Multi-Network Bidirectional Propagation. *BMC Med. Genomics* 12 (Suppl. 10), 185. doi:10.1186/s12920-019-0622-4
- Meltzer, P. S. (2005). Small RNAs with Big Impacts. *Nature* 435, 745–746. doi:10.1038/435745a
- Mørk, S., Pletscher-Frankild, S., Palleja Caro, A., Gorodkin, J., and Jensen, L. J. (2014). Protein-driven Inference of miRNA-Disease Associations. *Bioinformatics* 30, 392–397. doi:10.1093/bioinformatics/btt677
- Niu, H.-M., Guo, L.-Q., Qiao, Y.-H., and Jiao, H.-Y. (2021). MiR-429 Prohibited NF-Kb Signalling to Alleviate Contrast-Induced Acute Kidney Injury via Targeting PDCD4. *Autoimmunity* 54, 243–253. doi:10.1080/08916934.2021.1919878
- Pasquier, C., and Gardès, J. (2016). Prediction of miRNA-Disease Associations with a Vector Space Model. *Sci. Rep.* 6, 27036. doi:10.1038/srep27036
- Peng, J., Hui, W., Li, Q., Chen, B., Hao, J., Jiang, Q., et al. (2019). A Learning-Based Framework for miRNA-Disease Association Identification Using Neural Networks. *Bioinformatics* 35, 4364–4371. doi:10.1093/bioinformatics/btz254
- Perez-Iratxeta, C., Wjst, M., Bork, P., and Andrade, M. A. (2005). G2D: a Tool for Mining Genes Associated with Disease. *BMC Genet.* 6, 45. doi:10.1186/1471-2156-6-45
- Shepherd, E., Neal, R., Rose, P., Walter, F., and Hamilton, W. T. (2013). Clinical Features of Kidney Cancer in Primary Care: a Case-Control Study Using Primary Care Records. *Br. J. Gen. Pract.* 63, e250–e255. doi:10.3399/bjgp13x665215
- Shi, H., Xu, J., Zhang, G., Xu, L., Li, C., Wang, L., et al. (2013). Walking the Interactome to Identify Human miRNA-disease Associations through the Functional Link between miRNA Targets and Disease Genes. *BMC Syst. Biol.* 7, 101. doi:10.1186/1752-0509-7-101
- Shukla, G. C., Singh, J., and Barik, S. (2011). MicroRNAs: Processing, Maturation, Target Recognition and Regulatory Functions. *Mol. Cell. Pharmacol.* 3, 83–92. doi:10.4255/mcpharmacol.11.13
- Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer Statistics, 2019. *CA A Cancer J. Clin.* 69, 7–34. doi:10.3322/caac.21551

- Suk, H.-I., Lee, S.-W., and Shen, D. (2015). Latent Feature Representation with Stacked Auto-Encoder for AD/MCI Diagnosis. *Brain Struct. Funct.* 220, 841–859. doi:10.1007/s00429-013-0687-3
- Wang, D., Wang, J., Lu, M., Song, F., and Cui, Q. (2010). Inferring the Human microRNA Functional Similarity and Functional Network Based on microRNA-Associated Diseases. *Bioinformatics* 26, 1644–1650. doi:10.1093/bioinformatics/btq241
- Wang, G., Mao, W., Zheng, S., and Ye, J. (2009). Epidermal Growth Factor Receptor-Regulated miR-125a-5p - a Metastatic Inhibitor of Lung Cancer. *FEBS J.* 276, 5571–5578. doi:10.1111/j.1742-4658.2009.07238.x
- Wang, S.-H., Phillips, P., Sui, Y., Liu, B., Yang, M., and Cheng, H. (2018). Classification of Alzheimer's Disease Based on Eight-Layer Convolutional Neural Network with Leaky Rectified Linear Unit and Max Pooling. *J. Med. Syst.* 42, 85. doi:10.1007/s10916-018-0932-7
- Wang, Z., Zhang, X., Yang, Z., Du, H., Wu, Z., Gong, J., et al. (2012). MiR-145 Regulates PAK4 via the MAPK Pathway and Exhibits an Antitumor Effect in Human Colon Cells. *Biochem. Biophysical Res. Commun.* 427, 444–449. doi:10.1016/j.bbrc.2012.06.123
- Xiao, Q., Dai, J., Luo, J., and Fujita, H. (2019). Multi-view Manifold Regularized Learning-Based Method for Prioritizing Candidate Disease miRNAs. *Knowledge-Based Syst.* 175, 118–129. doi:10.1016/j.knsys.2019.03.023
- Xu, P., Guo, M., and Hay, B. A. (2004). MicroRNAs and the Regulation of Cell Death. *Trends Genet.* 20, 617–624. doi:10.1016/j.tig.2004.09.010
- Xuan, P., Han, K., Guo, M., Guo, Y., Li, J., Ding, J., et al. (2013). Prediction of microRNAs Associated with Human Diseases Based on Weighted K Most Similar Neighbors. *PLoS one* 8, e70204. doi:10.1371/journal.pone.0070204
- Xuan, P., Sun, H., Wang, X., Zhang, S., and Pan, S. (2019). Inferring the Disease-Associated miRNAs Based on Network Representation Learning and Convolutional Neural Networks. *Ijms* 20, 3648. doi:10.3390/ijms20153648
- Yang, Q., and Li, X. (2021). BiGAN: LncRNA-Disease Association Prediction Based on Bidirectional Generative Adversarial Network. *BMC Bioinforma.* 22, 357. doi:10.1186/s12859-021-04273-7
- Yang, Z., Wu, L., Wang, A., Tang, W., Zhao, Y., Zhao, H., et al. (2017). dbDEMOC 2.0: Updated Database of Differentially Expressed miRNAs in Human Cancers. *Nucleic Acids Res.* 45, D812–D818. doi:10.1093/nar/gkw1079
- Yu, W., Zhu, K., Wang, Y., Yu, H., and Guo, J. (2018). Overexpression of miR-21-5p Promotes Proliferation and Invasion of Colon Adenocarcinoma Cells through Targeting CHL1. *Mol. Med.* 24, 36. doi:10.1186/s10020-018-0034-5
- Zeng, H., Edwards, M. D., Liu, G., and Gifford, D. K. (2016a). Convolutional Neural Network Architectures for Predicting DNA-Protein Binding. *Bioinformatics* 32, i121–i127. doi:10.1093/bioinformatics/btw255
- Zeng, W., Zhu, J.-F., Liu, J.-Y., Li, Y.-L., Dong, X., Huang, H., et al. (2019). miR-133b Inhibits Cell Proliferation, Migration and Invasion of Esophageal Squamous Cell Carcinoma by Targeting EGFR. *Biomed. Pharmacother.* 111, 476–484. doi:10.1016/j.biopha.2018.12.057
- Zeng, X., Zhang, X., and Zou, Q. (2016b). Integrative Approaches for Predicting microRNA Function and Prioritizing Disease-Related microRNA Using Biological Interaction Networks. *Brief. Bioinform* 17, 193–203. doi:10.1093/bib/bbv033
- Zhou, F., Yin, M.-M., Jiao, C.-N., Zhao, J.-X., Zheng, C.-H., and Liu, J.-X. (2021). Predicting miRNA-Disease Associations through Deep Autoencoder with Multiple Kernel Learning. *IEEE Trans. Neural Netw. Learn. Syst.* doi:10.1109/tnnls.2021.3129772
- Zhou, S., Wang, S., Wu, Q., Azim, R., and Li, W. (2020). Predicting Potential miRNA-Disease Associations by Combining Gradient Boosting Decision Tree with Logistic Regression. *Comput. Biol. Chem.* 85, 107200. doi:10.1016/j.compbiolchem.2020.107200
- Zou, Q., Li, J., Song, L., Zeng, X., and Wang, G. (2016). Similarity Computation Strategies in the microRNA-Disease Network: a Survey. *Brief. Funct. Genomics* 15, 55–64. doi:10.1093/bfpg/elv024

Conflict of Interest: XL and QL were employed by Postdoctoral Program of Heilongjiang Hengxun Technology Co., Ltd.

The remaining 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: 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.

Copyright © 2022 Xu, Li, Yang, Tan, Liu and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.