

Article

Identification of MiRNA–Disease Associations Based on Information of Multi-Module and Meta-Path

Zihao Li ¹, Xing Huang ¹, Yakun Shi ¹, Xiaoyong Zou ^{2,*}, Zhanchao Li ^{3,*}  and Zong Dai ^{1,*}

¹ School of Biomedical Engineering, Sun Yat-sen University, Shenzhen 518107, China; lizh377@mail2.sysu.edu.cn (Z.L.); huangx48@mail2.sysu.edu.cn (X.H.); shiyk6@mail2.sysu.edu.cn (Y.S.)

² School of Chemistry, Sun Yat-sen University, Guangzhou 510275, China

³ School of Chemistry and Chemical Engineering, Guangdong Pharmaceutical University, Guangzhou 510006, China

* Correspondence: ceszxy@mail.sysu.edu.cn (X.Z.); zhanchao8052@gdpu.edu.cn (Z.L.); daizong@mail.sysu.edu.cn (Z.D.)

Abstract: Cumulative research reveals that microRNAs (miRNAs) are involved in many critical biological processes including cell proliferation, differentiation and apoptosis. It is of great significance to figure out the associations between miRNAs and human diseases that are the basis for finding biomarkers for diagnosis and targets for treatment. To overcome the time-consuming and labor-intensive problems faced by traditional experiments, a computational method was developed to identify potential associations between miRNAs and diseases based on the graph attention network (GAT) with different meta-path mode and support vector (SVM). Firstly, we constructed a multi-module heterogeneous network based on the meta-path and learned the latent features of different modules by GAT. Secondly, we found the average of the latent features with weight to obtain a final node representation. Finally, we characterized miRNA–disease-association pairs with the node representation and trained an SVM to recognize potential associations. Based on the five-fold cross-validation and benchmark datasets, the proposed method achieved an area under the precision–recall curve (AUPR) of 0.9379 and an area under the receiver–operating characteristic curve (AUC) of 0.9472. The results demonstrate that our method has an outstanding practical application performance and can provide a reference for the discovery of new biomarkers and therapeutic targets.

Keywords: MiRNA–disease association; graph neural network; meta-path



Citation: Li, Z.; Huang, X.; Shi, Y.; Zou, X.; Li, Z.; Dai, Z. Identification of MiRNA–Disease Associations Based on Information of Multi-Module and Meta-Path. *Molecules* **2022**, *27*, 4443. <https://doi.org/10.3390/molecules27144443>

Academic Editors: Ruoxu Gu, Zunnan Huang and Fengxu Wu

Received: 17 May 2022

Accepted: 8 July 2022

Published: 11 July 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

MicroRNA (miRNA), with a length between 18 and 24 nucleotides, is one of the types of non-coding RNAs in cells. Previously, miRNA was considered as a useless clip of human gene and even once called ‘junk gene’ because it could not encode protein [1]. However, more and more research studies show that miRNA is able to regulate the gene expression affecting some essential biological processes, such as proliferation, division, growth and apoptosis of the cell [2–4]. It commonly binds with messenger RNA (mRNA) at the three prime untranslated region (3′UTR) to achieve the transcription repression or degradation of the mRNA target [5]. Therefore, a high or low level of miRNA can lead to chaotic protein synthesis, which may destroy normal metabolism and cause dysfunction, further inviting diseases [6]. In addition, some studies have also shown that miRNA serving as an epigenetic regulator of gene expression goes hand in hand with human diseases [7–10]. Therefore, to identify the association between miRNAs and diseases is helpful for understanding the pathogenesis of disease. Moreover, miRNA can serve as a promising biomarker for diagnosis or a target for treatment [11]. Nevertheless, traditional biological experiments, limited by high cost, and being laborious and time consuming, are prone to failure to find all the relations between miRNA and disease. With the development of database and biotechnology, the massive accumulation of biological data enables researchers to

extract potential information and further adopt it to identify miRNA–disease association (MDA). Protein is an essential component of all cells and tissues in the body. Protein-related information has been utilized in plenty of bioinformatics studies, such as protein interaction and drug–protein interaction [12–14].

Up to now, a large number of computational methods have been developed to recognize MDA. These methods, roughly classified into three categories of similarity-based methods, network-based methods and machine learning based methods, are all based on the hypothesis that miRNAs with similar function tend to be associated with diseases with similar phenotypes. For similarity-based methods, throughout the development of the computational method for MDA, Jiang et al. [15] used a computation method instead of a traditional experiment to find unknown MDA. The work made use of the genes related to miRNA and hypergeometric distributions to calculate the miRNA similarity score and find the perspective neighbor miRNA by ranking the score; however, it only focused on the direct neighbor miRNA and neglected the undirect ones. Chen et al. [16] tried to integrate various heterogeneous biological datasets and calculate the within-score and between-score to rank the indefinite MDA. Pasquier et al. [17] utilized diverse information to construct miRNA and disease vector and find MDA by vector similarity. The network-based method predicts MDA by implementing random walk and other propagation algorithms in miRNA and disease network. Chen et al. [18] constructed the miRNA functional similarity network (MFSN) and implemented the random walk algorithm on it to obtain the score of candidate miRNAs. Xuan et al. [19] divided the miRNA into two categories, labeled and unlabeled, and also carried out a random walk on the MFSN, which enabled the prior information to improve the current information. However, these methods can only be used for miRNAs that have similar function to other miRNAs. To extend the prediction, some researchers have integrated a diverse biological dataset. You et al. [20] constructed a heterogeneous graph and developed a path-based method adopting a depth-first search algorithm to surmise MDA. Chen et al. [21] designed a method which implements random walk on the miRNA–miRNA and disease–disease network constructed by Laplacian score of graphs, respectively. The development of machine learning and deep learning breathes new life into the fields of healthcare and bioinformatics, such as disease prediction, sleep monitoring and medical image processing [22–24]. In addition, many methods based on machine learning and deep learning have been proposed to distinguish associations between miRNA and diseases. Jiang et al. [25] employed the miRNA and disease similarity score as a feature vector and randomly selected some unobserved MDA as negative samples to classify by support vector machine. Zhao et al. [26] attempted to integrate several decision trees to obtain the score with a respective weight, forming a strong classifier, which achieved an adaptive boosting improvement for prediction. Li et al. [27] utilized a graph convolution network to learn the latent feature of miRNA and disease. Subsequently, they acquired MDA by neural inductive matrix completion. Xuan et al. [28] constructed a dual convolutional neural network framework to learn the global and local representation for the subsequent prediction. Ji et al. [29] gained the miRNA and disease representation, respectively, by minimizing the squared losses between the value of cosine distance and the score of the function similarity, and then adopted the auto encoder to predict the probability of MDA.

Recently, graph neural network, depending on its ability to fuse the feature of node and graph topological structure, has been introduced into bioinformatics [13,30–33]. What is more, the introduction of meta-path is able to enrich the semantic information of the network and provide the extra structure information for uncovering the complexity of the network. As mentioned above, a protein whose chaos in synthesis may cause diseases plays an essential role in life activity as well as being regulated by miRNA. Thus, the integration of protein, miRNA and disease information may be able to significantly improve the prediction performance.

Inspired by graph neural networks such as graph convolutional network (GCN) [34], graph attention network (GAT) [35] and heterogenous graph attention network [36], a novel method is proposed for predicting miRNA–disease association. In the current approach,

multi-module meta-path along with graph attention network is employed to extract the network topology features of miRNAs and diseases, and support vector machine (SVM) is used as classifier to identify the potential MDA (MMGAN-SVM). Finally, five-fold cross-validation is conducted to evaluate the prediction performance and the case studies with lymphoma, liver neoplasms and lung neoplasms are performed to demonstrate the practical application performance.

Overall, the main contributions of this work are as follows:

1. Protein information and meta-path strategy were utilized to construct the multi-module, which can enrich the information of miRNAs and diseases.
2. The topological and semantic information can be better learned by Graph attention network and attention mechanism.
3. A reliable negative sample selection strategy was utilized to overcome the imbalance between positive and negative samples.

2. Results

2.1. Dimension Optimization of Node Representation

The node representation implies the complex information in latent feature space, and its dimensionality affects the predictive performance of the model. A low number of dimensions may lead to the loss of information, while a high number of dimensions will lead to the introduction of noise and time consuming for calculation. Thus, discovery of the optimized dimension of the node representation is attempted based on the 5-CV through changing dimension in the range of (32, 64, 128, 256, 512). The experiment is repeated 10 times for each dimension. Here, Acc, Roc and Aupr are utilized to evaluate the effect of dimension on model performance and statistical average results are shown in Figure 1. The Acc of each dimension is 0.8591, 0.8628, 0.8719, 0.8751 and 0.8700 and its standard deviation (std) is 0.0030, 0.0038, 0.0033, 0.0023 and 0.0040. The Roc of each dimension is 0.9178, 0.9251, 0.9392, 0.9472 and 0.9448 and its std is 0.0031, 0.0054, 0.0030, 0.0016 and 0.0034. The Aupr of each dimension is 0.8965, 0.9058, 0.9180, 0.9379 and 0.9401 and its std is 0.0052, 0.0054, 0.0063, 0.0042 and 0.0043. We can conclude that higher dimensionality tends to be better performance. However, a high feature vector can lead to a huge computational burden and long model training time. Therefore, the optimal feature dimension for node representation is set to 256. The learning curve of our model is shown in Figure 2 and the result illustrates that the model has been trained in an optimal state.

2.2. Classifier Optimization

Here, deep neural networks (DNNs), such as multi-layer perceptron (MLP), convolutional neural networks (CNN) and the traditional machine learning method, including SVM and random forest (RF), are utilized to construct a model. The 5-CV is conducted with a different model 10 times in the same condition as well as with the optimal parameter, and the result is shown in the Figures 3–5 and in Table 1. In the 5-CV experiment, SVM shows the best performance in Auc and Aupr. Although the evaluation measures of SVM in 10 repetitive experiments are a little better than those of other models, SVM performs a lower std than other models. In conclusion, SVM shows the better performance in the majority of evaluation measures.

Table 1. The performance comparison of different classifier.

	Acc	Auc	Aupr	Sens	Spec	Prec	F1	Mcc
MLP	0.8661	0.9460	0.9420	0.8924	0.8398	0.8504	0.8693	0.7364
CNN	0.8689	0.9458	0.9411	0.8984	0.8393	0.8490	0.8725	0.7399
RF	0.8639	0.9398	0.9327	0.8776	0.8502	0.8542	0.8657	0.7281
SVM	0.8752	0.9470	0.9374	0.9156	0.83491	0.8473	0.8801	0.7531

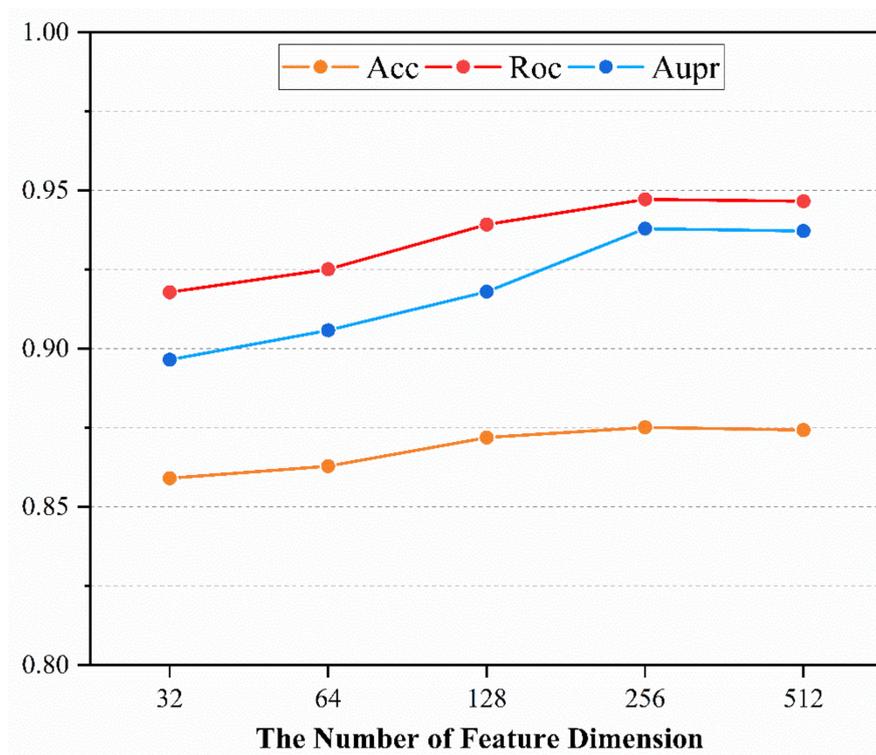


Figure 1. The main performance under different dimensions.

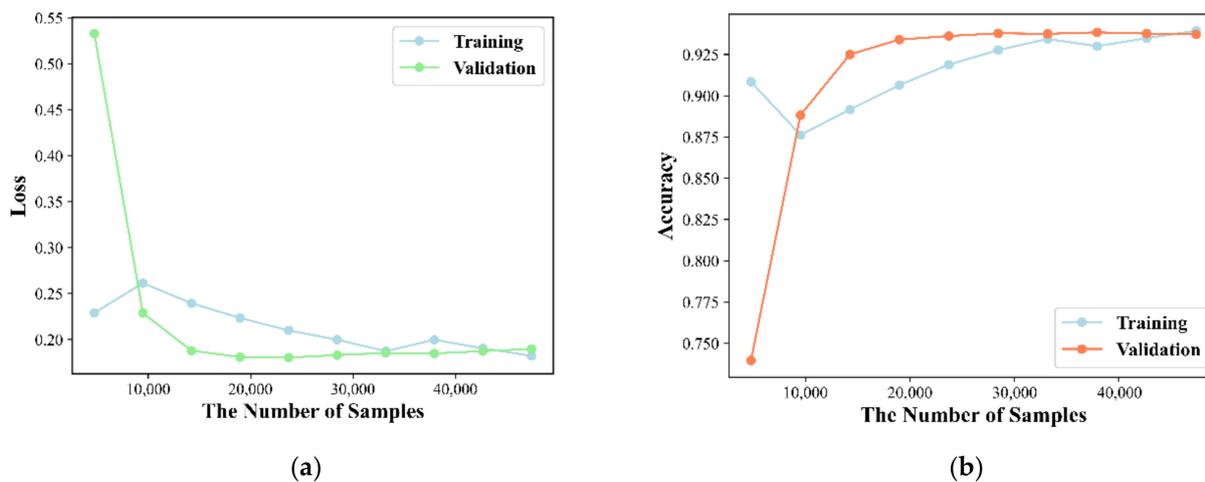


Figure 2. The learning curve of our method. (a) Training and validation accuracy graph; (b) Training and validation error graph.

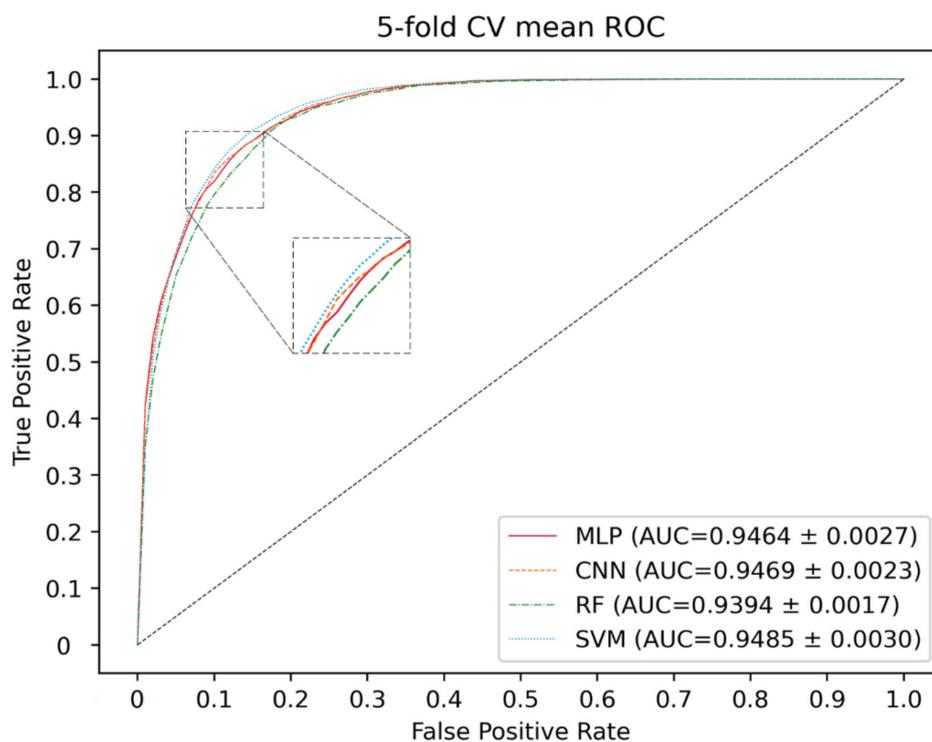


Figure 3. ROC curves for different classifier.

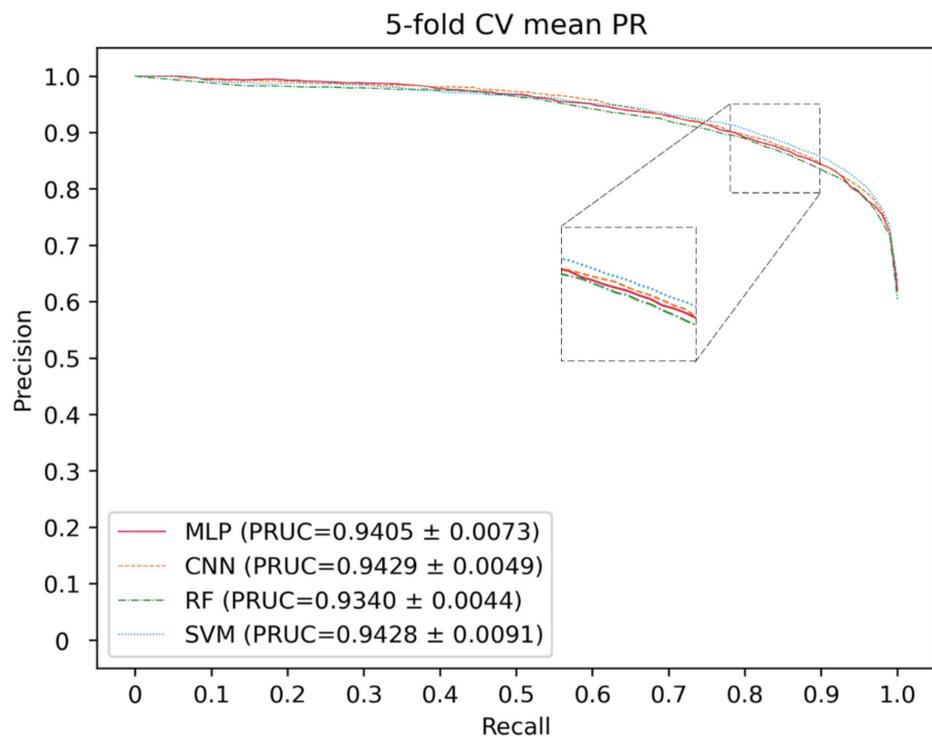


Figure 4. PR curves for different classifier.

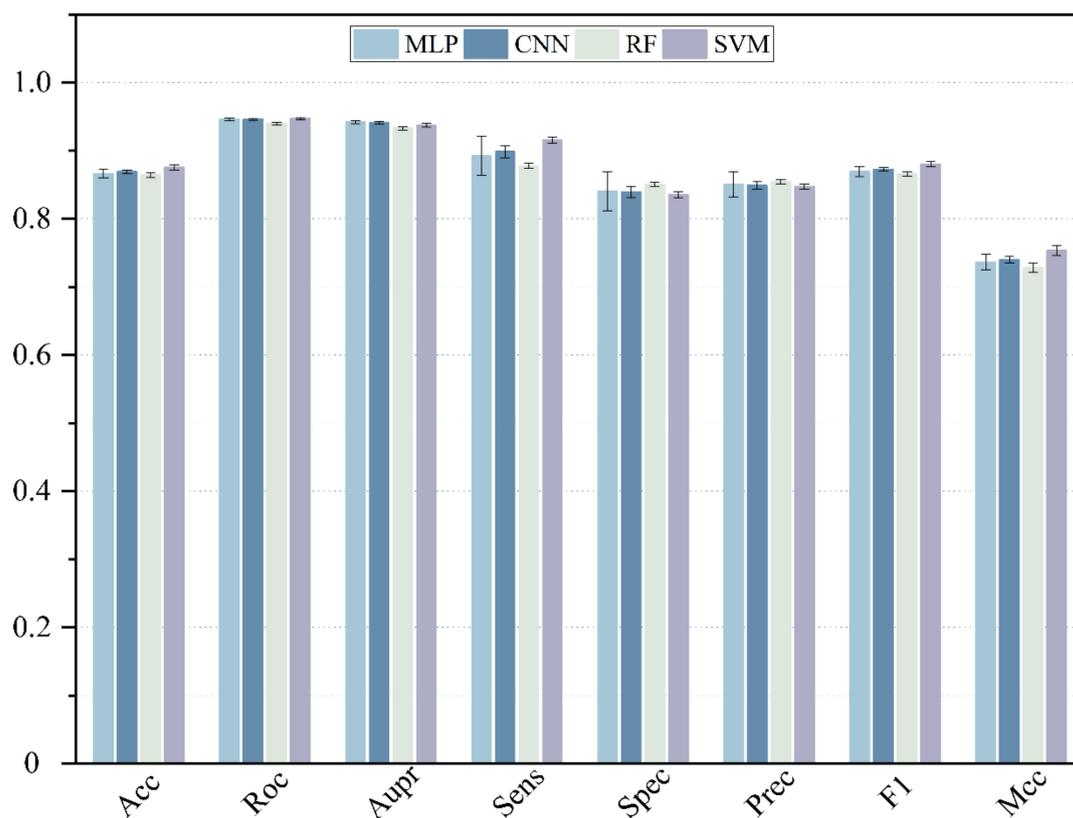


Figure 5. The performance comparison of different classifier.

2.3. Comparison with Other Methods

To further demonstrate the performance of the current method, a comparison is performed with some state of art methods including PBMDA [20], WBNPMD [37], NIMCGCN [27], DNRLMF-MDA [38] and VGAE-MDA [39]. PBMDA is a path-based method which aims at eliminating weak interactions. WBNPMD predicted the MDA by the bipartite network projection with weight. NIMCGCN is a matrix completion-based method which learns the feature by GCN. DNRLMF-MDA is a matrix factorization-based method and it utilized dynamic neighborhood regularization to improve performance. VGAE-MDA adopted variational graph auto-encoders to integrate the score from well-trained two subgraphs. Based on the benchmark dataset, the best results of 5-CV from our model are shown in Figures 6 and 7. The average of Acc, Roc, AuPr and F1 measured ten times in the experiment are 0.8753, 0.9472, 0.9374 and 0.8801 with the std 0.0036, 0.0015, 0.0030 and 0.0034, respectively. The five-fold cross-validation results of the existing methods are shown in Figure 8. The AUCs of PBMDA, WBNPMD, NIMCGCN, DNRLMF-MDA and VGAE-MDA are 0.9172, 0.9173, 0.9291, 0.9357 and 0.9394, respectively. In our method, the features of miRNAs and diseases are not only enriched by extra information of the protein but also integrated with the structure semantic information of MDA. In addition, SVM can show great performances in nonlinear classification tasks. Due to these strategies, the result also illustrated correspondingly that our method presented an outstanding performance.

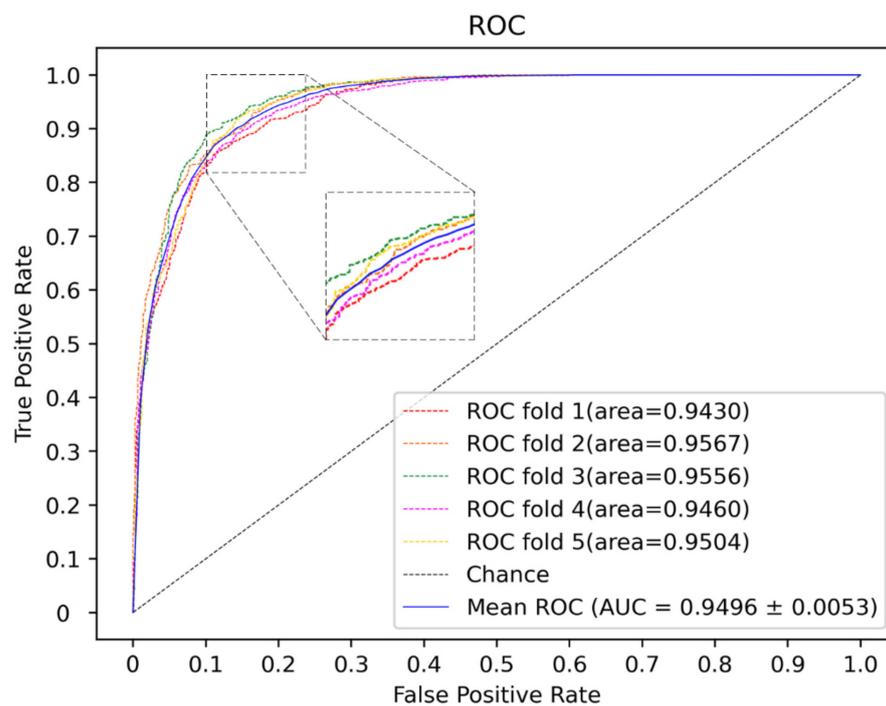


Figure 6. The ROC curves of our method.

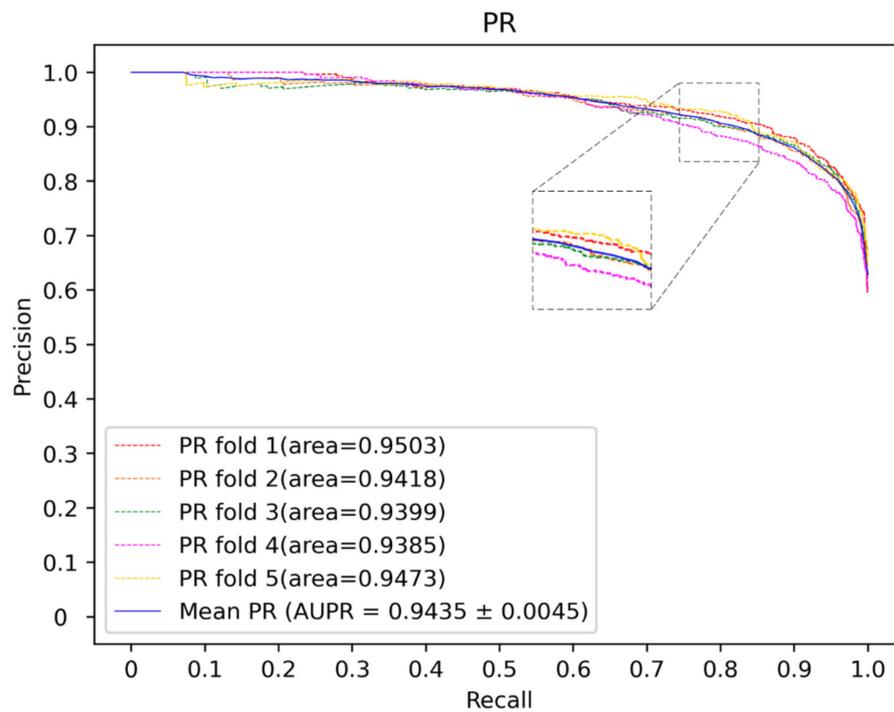


Figure 7. The PR curves of our method.

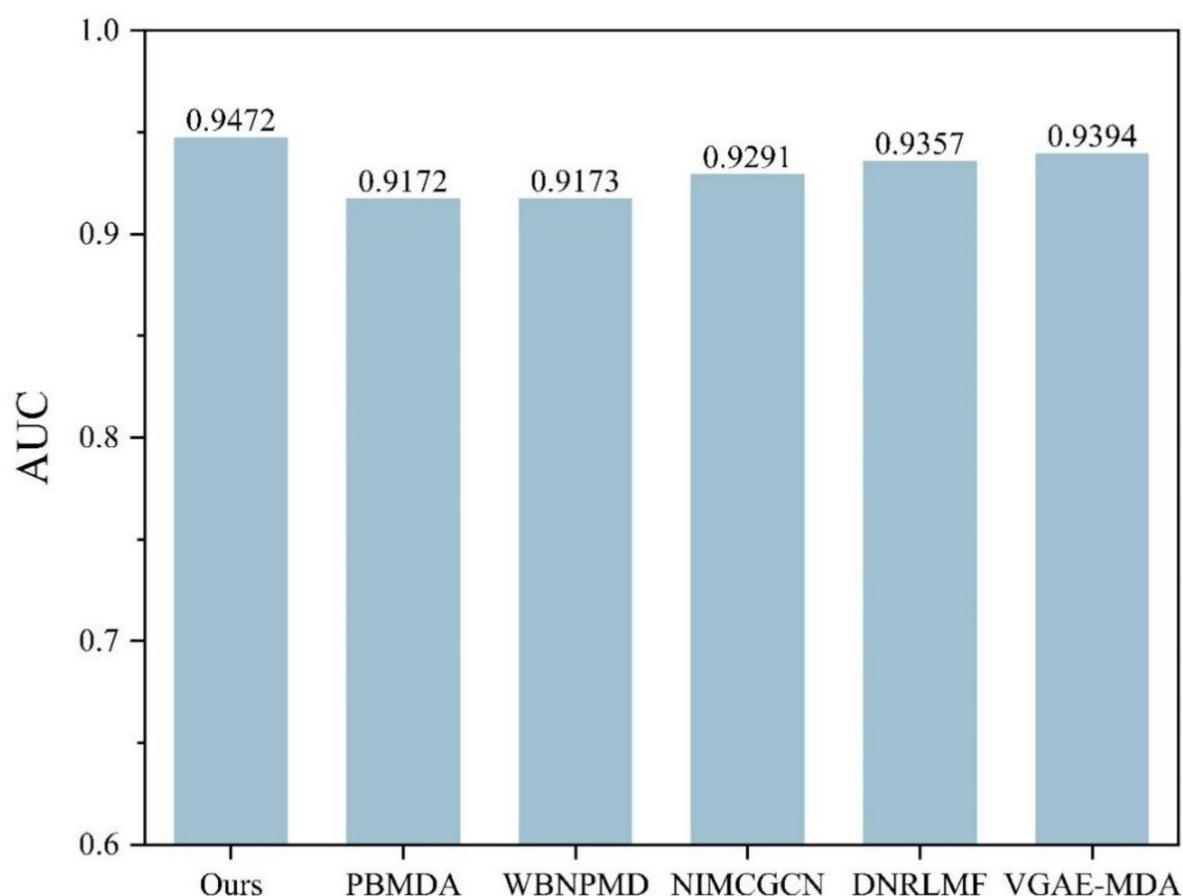


Figure 8. Performance comparison of different methods in 5-CV.

2.4. Proportion of Negative Sample

In fact, the number of negative samples is much larger than the number of positive samples. Therefore, the impact of the ratio of positive and negative samples on the performance of the model is further investigated. A negative sample with different ratios of 1:1, 1:2, 1:3, 1:4 and 1:5 is randomly selected to conduct the 5-CV, and the result is shown in Figure 9 and listed in Table 2. As we can see, some evaluation measures are affected significantly by the unbalance between positive and negative samples, because these evaluation measures are sensitive to the ratio between positive and negative samples. With the increase in negative samples, the value of Aupr, Sens, F1 and Mcc slowly descends. A greater number of negative samples involved in the training procedure makes it easier for the model to identify the negative samples. Thus, the value of Acc increases along with the growth of ratios. The values of Auc and Aupr fluctuate within a controllable range. However, aiming at digging potential MDA, it is necessary for the model to obtain high sensitivity. Thus, to display the best performance of our model, the proportion of negative and positive samples is set as 1:1.

Table 2. Performance comparison of different ratio of positive and negative samples.

Ratio	Acc	Auc	Aupr	Sens	Spec	Prec	F1	Mcc
1:1	0.8753	0.9470	0.9375	0.9157	0.8349	0.8473	0.8801	0.7531
1:2	0.8790	0.9481	0.8989	0.8168	0.9101	0.8199	0.8182	0.7277
1:3	0.8901	0.9460	0.8634	0.7210	0.9464	0.8177	0.7662	0.6971
1:4	0.9002	0.9422	0.8321	0.6461	0.9637	0.8167	0.7213	0.6684
1:5	0.9098	0.9325	0.7984	0.5898	0.9738	0.8186	0.6854	0.6461

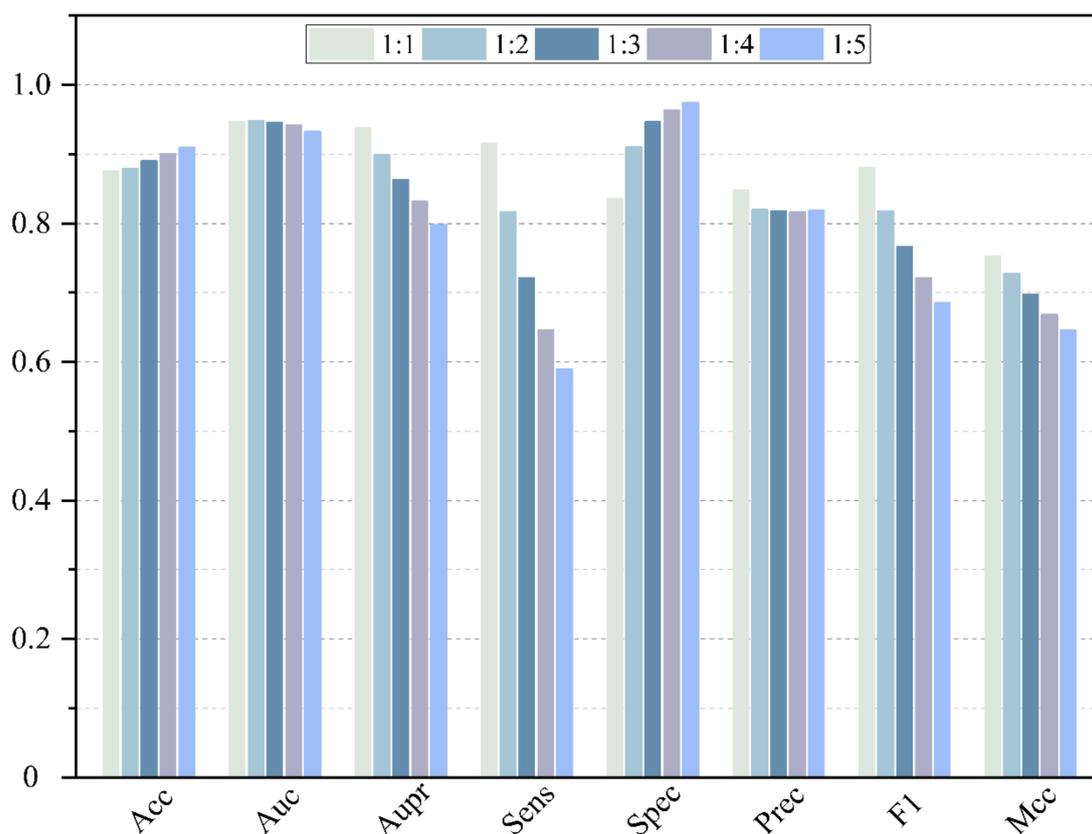


Figure 9. Performance comparison of different ratio of positive and negative samples.

2.5. Reliability of Negative Sample

At present, there is no database dedicated to collecting miRNA–disease non-association pairs because these pairs cannot provide more information to promote the mechanisms' research and drug discovery. To overcome this problem, a random matching method is utilized to construct negative samples; however, it may contain false negative samples. Therefore, the influence of negative sample reliability on model performance is further studied. At first, we calculated the mean values of all dimensions for all the positive samples to form a cluster vector. Then, we obtained the average Euclidean distance (AED) by calculating Euclidean distance between each negative sample and the cluster vector. Depending on different threshold of AED, the original negative sample set was able to be refined and shrunk. The AED threshold was set in the range of (0.4AED, 0.5AED, 0.6AED, 0.7AED, 0.8AED, 0.9AED 1.0AED) and the negative samples whose Euclidean distance was lower than the threshold were removed to obtain different negative sample datasets. Then, negative sample with the same ratios as positive samples were randomly selected from the dataset for the training set in each of the threshold experiments. The results of different threshold are shown in Figure 10. The values of Acc, Auc, AuPr, Sens, Spec, Precision, F1 and Mcc are located in the range of (0.8798–0.9755), (0.9506–0.9932), (0.9423–0.9975), (0.9148–0.9713), (0.8448–0.9798), (0.8551–0.9795), (0.7616–0.9510) and (0.8839–0.9753), respectively. In addition, with the increase in threshold, the selected negative sample is further away from the cluster vector, and there is a degree of improvement for all the evaluation measure. Thus, the strategy of reliable negative sample selection makes a positive difference on the model.

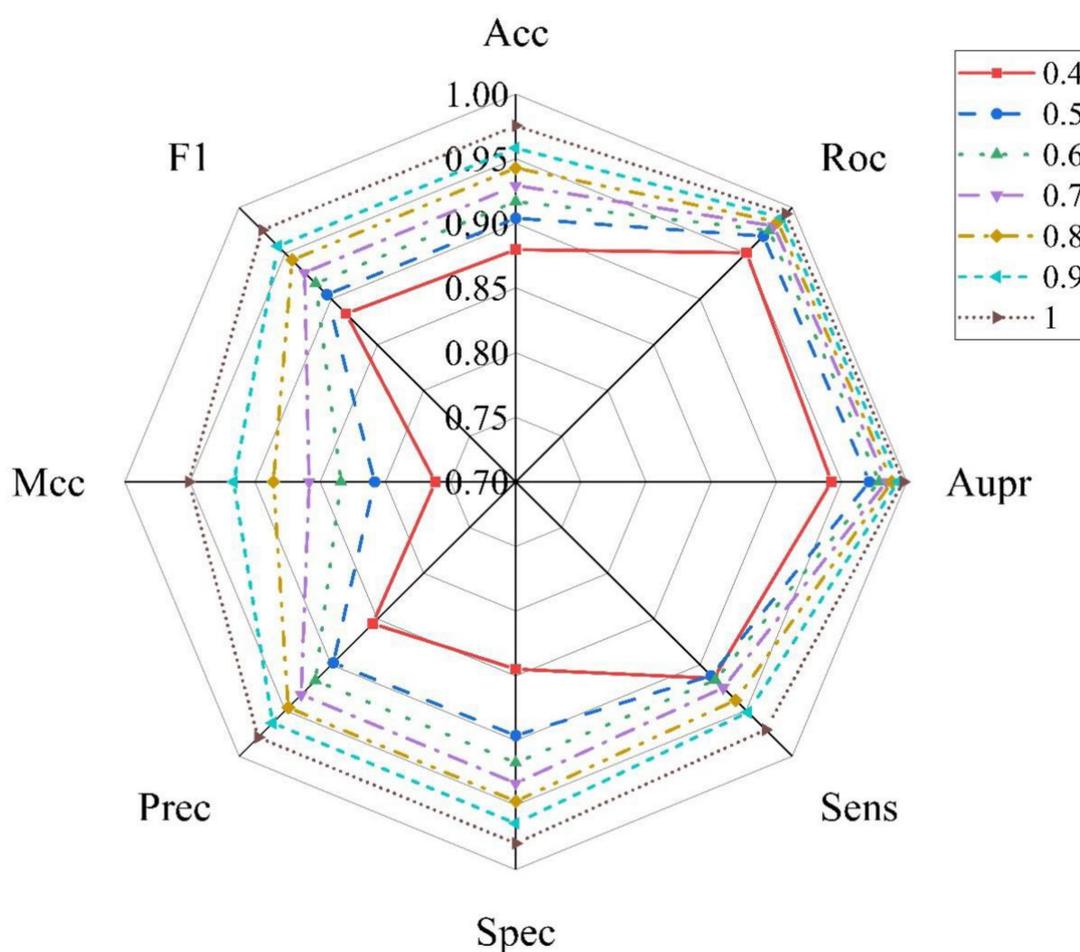


Figure 10. Performance comparison of different thresholds of negative samples.

2.6. Case Studies

To illustrate the practical application performance of our model, the case studies are implemented over the three common human diseases: liver neoplasm, lung neoplasm and leukemia. Specifically, the MDA information of each case study is erased during the model training and the prediction score is acquired for all the miRNA candidates. Here, the ratio of positive and negative samples is set as 1:1 and the strategy of reliable negative samples selection is utilized in training procedure. According to the prediction scores, these identified potential disease-related miRNAs are ranked in descending order. For the three diseases, the recognized top 30 miRNAs and the corresponding scores are listed in Tables 3–5, respectively. Meanwhile, these results are validated by the databases of HMDD V3.0 and dbDEMC. The latter is a database recording the expression profiles of cancer-related miRNA and the published literature [40].

Table 3. Top-30 Predicted Associations of Liver Neoplasms.

Rank	Score	miRNA	Evidence
1	0.9557	hsa-miR-21	HMDD3.0, dbDEMC, PMID: 31037150
2	0.9540	hsa-miR-155	dbDEMC, PMID: 29565484
3	0.9477	hsa-miR-146a	HMDD3.0, dbDEMC, PMID: 29133238
4	0.9345	hsa-miR-29a	HMDD3.0, dbDEMC, PMID: 33891266
5	0.9326	hsa-miR-16	HMDD3.0, dbDEMC, PMID: 30657555
6	0.9323	hsa-miR-29b	dbDEMC, PMID: 34184070
7	0.9309	hsa-miR-125b	HMDD3.0 dbDEMC, PMID: 32609900
8	0.9301	hsa-miR-15a	dbDEMC, PMID: 31099097

Table 3. Cont.

Rank	Score	miRNA	Evidence
9	0.9266	hsa-miR-1	dbDEMC, PMID: 31846694
10	0.9242	hsa-miR-221	HMDD3.0, dbDEMC, PMID: 31069760
11	0.9220	hsa-miR-34a	HMDD3.0, dbDEMC, PMID: 32778238
12	0.9203	hsa-miR-17	dbDEMC, PMID: 32206115
13	0.9195	hsa-miR-20a	dbDEMC, PMID: 32206115
14	0.9184	hsa-miR-199a	HMDD3.0, dbDEMC, PMID: 31144384
15	0.9183	hsa-miR-133a	dbDEMC, PMID: 30086463
16	0.9150	hsa-miR-19b	dbDEMC, PMID: 29889802
17	0.9147	hsa-miR-29c	HMDD3.0 dbDEMC, PMID: 30718452
18	0.9141	hsa-miR-223	HMDD3.0, dbDEMC, PMID: 32233593
19	0.9139	hsa-miR-222	HMDD3.0, dbDEMC, PMID: 34273068
20	0.9101	hsa-miR-150	dbDEMC, PMID: 25549355
21	0.9043	hsa-miR-92a	dbDEMC, PMID: 32587378
22	0.9040	hsa-miR-18a	dbDEMC, PMID: 34221105
23	0.9015	hsa-miR-145	dbDEMC, PMID: 29658584
24	0.9011	hsa-miR-106b	dbDEMC, PMID: 29975452
25	0.9009	hsa-miR-181a	dbDEMC, PMID: 25058462
26	0.9006	hsa-miR-19a	dbDEMC, PMID: 27012708
27	0.8999	hsa-miR-210	HMDD3.0, dbDEMC, PMID: 27666683
28	0.8978	hsa-miR-31	HMDD3.0, dbDEMC, PMID: 25797269
29	0.8957	hsa-miR-122	HMDD3.0, dbDEMC, PMID: 25537773
30	0.8941	hsa-miR-142	HMDD3.0, dbDEMC, PMID: 30092578

Table 4. Top-30 Predicted Associations of Lung Neoplasms.

Rank	Score	miRNA	Evidence
1	0.9690	hsa-miR-21	HMDD3.0, dbDEMC, PMID: 30736829
2	0.9675	hsa-miR-155	HMDD3.0, dbDEMC, PMID:32447486
3	0.9673	hsa-miR-122	HMDD3.0, dbDEMC, PMID: 26604787
4	0.9672	hsa-miR-15a	HMDD3.0, dbDEMC, PMID: 33059020
5	0.9671	hsa-miR-29a	HMDD3.0, dbDEMC, PMID: 33250420
6	0.9670	hsa-miR-16	HMDD3.0, dbDEMC, PMID: 31379227
7	0.9660	hsa-miR-29b	HMDD3.0, dbDEMC, PMID: 31813135
8	0.9647	hsa-miR-133a	HMDD3.0, dbDEMC, PMID: 33074595
9	0.9630	hsa-miR-1	HMDD3.0, dbDEMC, PMID: 34139980
10	0.9626	hsa-miR-15b	dbDEMC, PMID: 32220063
11	0.9617	hsa-miR-199a	HMDD3.0, dbDEMC, PMID: 28363780
12	0.9608	hsa-miR-146a	HMDD3.0, dbDEMC, PMID: 29127520
13	0.9602	hsa-miR-29c	HMDD3.0, dbDEMC, PMID: 29512752
14	0.9598	hsa-miR-26a	HMDD3.0, dbDEMC, PMID: 33407724
15	0.9588	hsa-miR-126	HMDD3.0, dbDEMC, PMID: 34107168
16	0.9586	hsa-miR-192	HMDD3.0, dbDEMC, PMID: 29571988
17	0.9581	hsa-miR-30b	HMDD3.0, dbDEMC, PMID: 33779882
18	0.9578	hsa-miR-106b	dbDEMC, PMID: 34351868
19	0.9575	hsa-miR-19b	HMDD3.0, dbDEMC, PMID: 29455644
20	0.9569	hsa-miR-150	HMDD3.0, dbDEMC, PMID: 24456795
21	0.9575	hsa-miR-23a	HMDD3.0, dbDEMC, PMID: 28436951
22	0.9567	hsa-miR-196a	HMDD3.0, dbDEMC, PMID: 33775710
23	0.9561	hsa-miR-19a	HMDD3.0, dbDEMC, PMID: 28364280
24	0.9558	hsa-miR-23b	dbDEMC, PMID: 32495614
25	0.9556	hsa-miR-206	HMDD3.0, dbDEMC, PMID: 26919096
26	0.9555	hsa-miR-26b	HMDD3.0, dbDEMC, PMID: 26744864
27	0.9552	hsa-miR-223	HMDD3.0, dbDEMC, PMID: 29615147
28	0.9547	hsa-miR-195	HMDD3.0, dbDEMC, PMID: 32406336
29	0.9544	hsa-miR-222	HMDD3.0, dbDEMC, PMID: 32588752
30	0.9539	hsa-miR-34a	HMDD3.0, dbDEMC, PMID: 30700696

Table 5. Top 30 Predicted Associations of Leukemia.

Rank	Score	miRNA	Evidence
1	0.9819	hsa-miR-21	HMDD3.0, dbDEMC, PMID: 32911844
2	0.9804	hsa-miR-155	HMDD3.0, dbDEMC, PMID: 33357126
3	0.9723	hsa-miR-146a	HMDD3.0, dbDEMC, PMID: 32798394
4	0.9643	hsa-miR-17	HMDD3.0, dbDEMC, PMID: 35536524
5	0.9632	hsa-miR-29a	HMDD3.0, dbDEMC, PMID: 31870103
6	0.9631	hsa-miR-125b	HMDD3.0, dbDEMC, PMID: 27637078
7	0.9630	hsa-miR-34a	HMDD3.0, dbDEMC, PMID: 27424989
8	0.9629	hsa-miR-20a	HMDD3.0, dbDEMC, PMID: 34587164
9	0.9622	hsa-miR-16	HMDD3.0, dbDEMC, PMID: 28599250
10	0.9606	hsa-miR-221	HMDD3.0, dbDEMC, PMID: 29172404
11	0.9605	hsa-miR-29b	dbDEMC, PMID: 29435107
12	0.9568	hsa-miR-92a	HMDD3.0, dbDEMC, PMID: 31870103
13	0.9556	hsa-miR-145	HMDD3.0, dbDEMC, PMID: 32538049
14	0.9552	hsa-miR-126	HMDD3.0, dbDEMC, PMID: 34686664
15	0.9546	hsa-miR-1	dbDEMC, PMID: 28042875
16	0.9543	hsa-miR-15a	HMDD3.0, dbDEMC, PMID: 24026141
17	0.9532	hsa-miR-19b	HMDD3.0, dbDEMC, PMID: 29032147
18	0.9520	hsa-miR-18a	HMDD3.0, dbDEMC, PMID: 32146479
19	0.9505	hsa-let-7a	dbDEMC, PMID: 29398802
20	0.9489	hsa-miR-19a	HMDD3.0, dbDEMC, PMID: 34895042
21	0.9473	hsa-miR-222	HMDD3.0, dbDEMC, PMID: 20203269
22	0.9463	hsa-miR-143	dbDEMC, PMID: 28890884
23	0.9454	hsa-miR-31	HMDD3.0, dbDEMC, PMID: 22511990
24	0.9453	hsa-miR-29c	dbDEMC, PMID: 31333331
25	0.9445	hsa-miR-223	HMDD3.0, dbDEMC, PMID: 27900032
26	0.9443	hsa-miR-133a	dbDEMC, PMID: 32647415
27	0.9439	hsa-miR-199a	HMDD3.0, dbDEMC, PMID: 31636666
28	0.9409	hsa-let-7b	HMDD3.0, dbDEMC, PMID: 33283713
29	0.9398	hsa-miR-150	HMDD3.0, dbDEMC, PMID: 27917123
30	0.9386	hsa-miR-200b	PMID: 30574752

Development of liver neoplasm, which has the highest mortality rate in the East Asia region, is contributed to by genetic and epigenetic factors [41]. There are two principal subtype of liver cancer, hepatocellular carcinoma (HCC) and cholangiocarcinoma, and the former is the main type happening to the case in [42]. All the top thirty miRNAs predicted can be confirmed by HMDD V3.0 or dbDEMC. In addition, some researchers reported that the over-expression of miR-221/222 is responsible for the multifocality of HCC, and the over-expression of miR-155 occurs after the cancer recurrence [43,44]. All of those miRNAs appear in the Top thirty predicted results.

Lung neoplasm is a common tumor with the highest morbidity, after breast neoplasm, worldwide, and it can be divided into two categories: small cell lung carcinoma and non-small cell lung carcinoma [45]. As listed in Table 4, all miRNAs can be validated by HMDD V3.0 or dbDEMC. Fan et al. reported the expression of miR-20a and miR-15b to be evidence to distinguish the case from healthy individuals [46]. In addition, miR-223 and miR-145 in plasma can be considered as potential biomarkers for early diagnosis [47].

Leukemia is recognized as a progressive malignant disease and is divided into four main types: acute leukemia, chronic leukemia, myelogenous leukemia and lymphocytic leukemia [48]. Twenty-nine of the top thirty predicted miRNAs in Table 5 can be validated by HMDD V3.0 or dbDEMC. Only one predicted result of miR-200b without recorded in database; however, it is revealed to promote the cell proliferation and invasion in leukemia [49]. MiR-200b acts as an oncogenic regulator in human lung cancer. The proliferation, invasion and apoptosis of leukemia cells can be controlled by miR-200b through its regulatory of NOTCH1 signaling pathway. In addition, the inactivation of miR-155 and miR-29 contribute in leukemia and the expression decrement of miR-223 can be used to distinguish the case from a healthy individual [50,51].

3. Discussion

As the epigenetic controller, miRNAs are involved in gene expression and cellular signaling pathways, which makes a difference in cell propagation, division, growth and apoptosis leading. With these functions, miRNAs are considered to play a critical role in the initiation and progression of human diseases as well as being the promising biomarker or therapeutic target to help with the early diagnosis and treatments. Hence, it is meaningful to discover the potential related miRNAs for a disease. In this study, a model is proposed for feature extraction and to build a model classification. The results have been compared with the state-of-the-art methods in 5-CV and the case studies showed that our method has a great performance. The comparisons of our method and other methods were performed, and advantages and drawbacks are listed in Table 6. The methods of PBMDA and WBNPMD obtained AUC of 0.9172 and 0.9173, respectively, because neither complex network was created, nor weighted edges adopted. Construction of the complex network contribute to enriching potential information of networks and adoption of a weighted edges strategy brings known microRNA disease associations into sharper focus. On the contrary, the methods of NIMCGCN, DNRLMF and VGAE-MDA with AUC of 0.9291, 0.9357 and 0.9394 not only constructed a complex network but also adopted diverse strategy to improve the prediction performance, such as neural inductive matrix completion (NIMCGCN), dynamic regularized weight (DNRLMF) and variational Bayesian inference (VGAE-MDA). Our method obtained the highest AUC of 0.9472, because the complex network was constructed and the weighted edge of microRNA disease association was considered among different modules. In addition, the weighted parameters can be adaptively learned by loss function. However, the unbalance sample problem should be investigated for all methods. The outstanding performance of our method stems from three factors. First, the information of protein is introduced to enrich the feature of miRNAs and diseases, and a composite module based on meta-path is constructed. Second, the latent feature incorporating information of the node and topological structure are extracted by node aggregation in different meta-paths and modules with an attention mechanism. Third, SVM is able to complete the non-linear classification task well on the feature extracted. In the future, much more information, such as miRNAs expression profiles, miRNA sequences and drugs, will be taken into account to improve the MDAs prediction performance. In addition, more efficient feature extraction algorithms will be a novel direction.

Table 6. The advantages and drawbacks of our method and other methods.

Method	AUC	Advantages	Drawbacks
PBMDA	0.9172	Topological information, complex network	No weighted, imbalance problem
WBNPMD	0.9173	Weighted edges	No topological information, imbalance problem
NIMCGCN	0.9291	Topological information, complex network, neural inductive	No weighted, imbalance problem
DNRLMF	0.9357	Complex network, dynamic regularized weight	No topological information, imbalance problem
VGAE-MDA	0.9394	Topological information, complex network, variational Bayesian inference	No weighted, imbalance problem
Ours	0.9472	Topological information, complex network, adaptive weight	Imbalance problem

4. Materials and Methods

The experiment-verified miRNA–disease associations were retrieved from the Human microRNA Disease Database (HMDD) [52]. In this work, HMDD V2.0 was adopted as the benchmark dataset with 5430 human miRNA–disease associations incorporating 495 miRNAs and 383 diseases after deduplication and normalization. For convenience, these associations are described as an adjacent matrix $A \in \{0, 1\}^{m \times n}$, in which m and n are the number of miRNAs and diseases, respectively. If an miRNA i is associated with a

disease j , the value of $A_{i,j}$ is 1, and 0 vice versa. In addition, the miRNA–protein associations were collected from the miRTarBase database and the disease–protein associations from Comparative Toxicogenomics Database (CTD) [53,54].

4.1. Integration Similarity Calculation and Multi-Module Construction

4.1.1. MiRNA Integration Similarity

MiRNA integration similarity was composed of miRNA functional similarity (MFS) and Gaussian interaction profile kernel similarity [55]. The calculation of MFS was defined according to the previous work, which was based on the assumption that the function of two miRNAs are more similar if the number of the common disease associated with them is greater [56,57]. The score of MFS was defined as $FS(i, j)$, i.e., the similarity score between miRNA i and miRNA j . In addition, to supplement the missing entries of MFS, Gaussian interaction profile kernel similarity $mGS(i, j)$ was adopted and defined as Equation (1):

$$mGS(i, j) = \exp\left(-\delta_m \|M(m_i) - M(m_j)\|^2\right) \quad (1)$$

where $M(m_i)$ and $M(m_j)$ indicate the i th and j th row of the adjacent matrix A , respectively. δ_m represents the kernel bandwidth parameter, and is illustrated as Equation (2):

$$\delta_m = \frac{1}{m} \sum_{i=1}^m \|M_i\|^2 \quad (2)$$

where m is the number of miRNAs. Finally, miRNA integration similarity is described as Equation (3):

$$MS(m_i, m_j) = \begin{cases} FS(m_i, m_j) & \text{If there is a function similarity between } m_i \text{ and } m_j \\ mGS(m_i, m_j) & \text{otherwise} \end{cases} \quad (3)$$

4.1.2. Disease Integration Similarity

Disease integration similarity constitutes disease semantic similarity and Gaussian interaction profile kernel similarity. The entry of diseases in the National Library of Medicine (<http://www.ncbi.nlm.nih.gov/> (accessed on 5 November 2021)) describes the relationship among different disease, which can be used to construct a hierarchical directed acyclic graph (DAG). According to the definition by Wang et al. [56], semantic contribution of a disease d is calculated as Equation (4):

$$DC_d(i) = \begin{cases} 1 & \text{if } i = d \\ \max\{\sigma * DC_d(d') \mid d' \in \text{children of } d\} & \text{if } i \neq d \end{cases} \quad (4)$$

where σ is a semantic contribution decay factor, and is maintained it as the same as the previous work that was set as 0.5 [56]. The semantic value of the disease d_i , $DV(d_i)$ is defined as Equation (5):

$$DV(d_i) = \sum_{k \in D(d_i)} DC_{d_i}(k) \quad (5)$$

where $D(d_i)$ is the node set of disease d_i and its ancestor. The semantic similarity score between disease d_i and d_j can be calculated as Equation (6):

$$SS(d_i, d_j) = \frac{\sum_{k \in D(d_i) \cap D(d_j)} (DC_{d_i}(k) + DC_{d_j}(k))}{DV(d_i) + DV(d_j)} \quad (6)$$

Finally, the disease semantic similarity combined with Gaussian kernel similarity is defined as Equation (7):

$$DS(d_i, d_j) = \begin{cases} SS(d_i, d_j) & \text{if there is a semantic similarity between } d_i \text{ and } d_j \\ dGS(d_i, d_j) & \text{otherwise} \end{cases} \quad (7)$$

where dGS is the Gaussian kernel similarity of disease and defined as Equation (8), its formulation is similar with Equation (8).

$$dGS(i, j) = \exp\left(-\delta_m \|Ds(d_i) - Ds(d_j)\|^2\right) \quad (8)$$

where $Ds(d_i)$ and $Ds(d_j)$ indicates the i th and j th column of the adjacent matrix A , respectively.

4.1.3. Multi Module Construction

Meta-path is explained as a path in form of $P = N_1 \xrightarrow{r_1} N_2 \xrightarrow{r_2} \dots \xrightarrow{r_{m-2}} N_{m-1} \xrightarrow{r_{m-1}} N_m$ (simplified as $N_1 N_2 \dots N_{m-1} N_m$), which illustrates that the starting node N_1 is able to reach one of the destination nodes connected by a composite relation $R = r_1 r_2 \dots r_{m-1} r_{m-2} \dots r_{m-1}$ [58]. Based on the meta-path, various significance can be received from the relation between two of the identical type nodes. Thus, miRNA–protein association and disease–protein association were introduced to enrich the information of the miRNA–miRNA and disease–disease association (MMA and DDA) network. Another four association matrices MDMA (MMA based on disease), MPMA (MMA based on protein), DMDA (DDA based on miRNA) and DPDA (DDA based on protein) are shown in Figure 11.

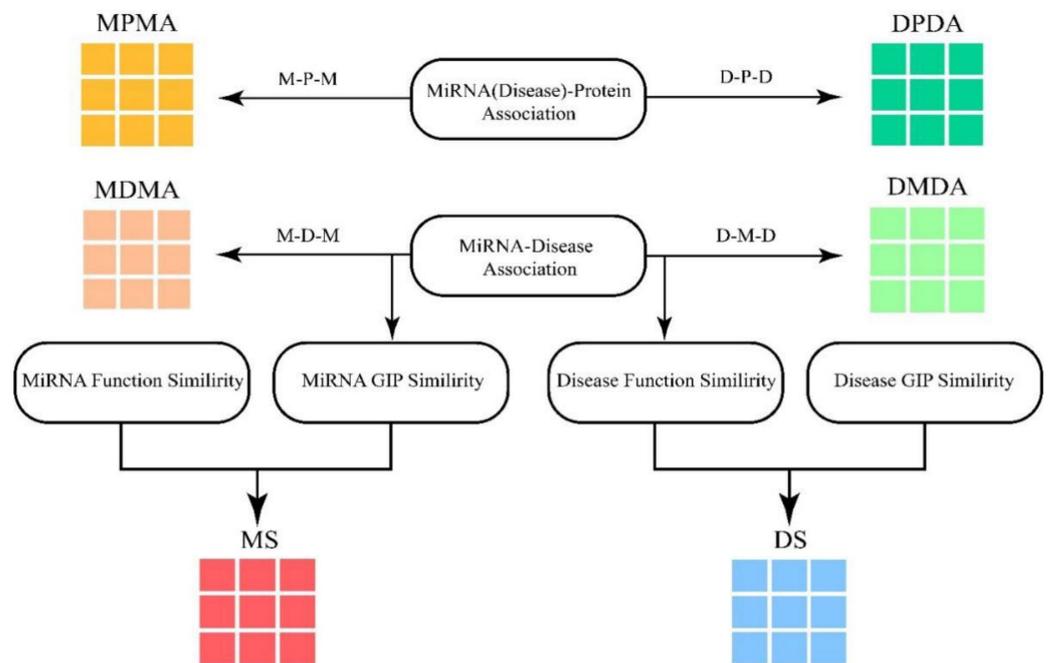


Figure 11. The construction of module. MPMA and MDMA are miRNA adjacent matrices based on proteins and diseases respectively. DPDA and DMDA are disease adjacent matrices based on proteins and diseases respectively. MS and DS are the similarity matrices of miRNAs and diseases respectively.

Take MDMA for example, the value of $MDMA(i, j)$ is 1 when miRNA i can reach miRNA j through a disease d , and it is 0 vice versa. To increase the density of MDA to about 3%, the heterogeneous graph is constructed in term of multi-module as Equations (9)–(11):

$$G_1 = \begin{bmatrix} MS & A \\ A^T & DS \end{bmatrix} \quad (9)$$

$$G_2 = \begin{bmatrix} \text{MDMA} & A \\ A^T & \text{DMDA} \end{bmatrix} \quad (10)$$

$$G_3 = \begin{bmatrix} \text{MPMA} & A \\ A^T & \text{DPDA} \end{bmatrix} \quad (11)$$

where the A^T is the transposition matrix of A .

4.2. Information Aggregation

4.2.1. Node Feature Linear Transformation and Aggregation

The original feature of miRNA and disease had to be projected into the same latent feature space, because their original feature represented two different feature spaces. The latent feature of nodes could be obtained by leveraging the transformation matrix to carry on linear transformation. Specifically, each type of node adopts a respective transformation matrix. For the node $i \in N_c$ of type c , the latent feature $h_i \in R^{d'}$ of it could be obtained by using Equation (12):

$$h_i = W_c \cdot x_c^i \quad (12)$$

where $W_c \in R^{d' \times n}$ is the transformation matrix of type c and $x_c^i \in R^n$ is the original feature of node i .

GAT was able to aggregate the information of neighboring nodes for the central node by of assigning learnable weight, which finally obtained the node representation by fusing the information of the network topological structure and node feature. Specifically, GAT adopted the SoftMax function to calculate the attention score of each node, and then continually updated the information of the central node by aggregating that of neighboring nodes based on their respective attention score. For each graph G constructed by meta-path, the importance e_{ij}^G contributed by neighbor node j to central node i was defined by Equation (13):

$$e_{ij}^G = \text{LeakyReLU}(\omega^G \cdot [h_i || h_j]) \quad (13)$$

where $\omega^G \in R^{2d'}$ is the attention parameter vector for graph G and $||$ represents the concatenation operation. SoftMax function was utilized to normalize the importance of all nodes in order to obtain the final attention score α_{ij} which was defined by Equation (14):

$$\alpha_{ij}^G = \text{SoftMax}(e_{ij}^G) = \frac{\exp(e_{ij}^G)}{\sum_{k \in N_i^G} \exp(e_{ik}^G)} \quad (14)$$

where k indicates the neighbor node of i in the graph G .

4.2.2. Module Aggregation

Based on the attention score, the information of node i was able to aggregate that of its neighbor nodes and eventually obtain the node representation z_i^G of graph G and defined as Equation (15):

$$z_i^G = \sigma \left(\sum_{j \in N_i^G} \alpha_{ij}^G \cdot h_j \right) \quad (15)$$

where $\sigma(\cdot)$ represents the nonlinear activation function, and sigmoid function was used in the current study.

For the sake of the stability and low variance, multi-head attention mechanism was introduced to improve the learning process of attention score. Specifically, the aggregation was repeated for K times and the formulation (15) can be revised by Equation (16):

$$z_i^G = \prod_{k=1}^K \sigma \left(\sum_{j \in N_i^G} \alpha_{ij}^G \cdot h_j \right) \quad (16)$$

In addition, due to different meta-paths, every node obtained more than one representation in various semantic significances. To figure out which meta-path was more essential, an attention mechanism could be also adopted among the representations obtained by different meta-paths. We denoted that different meta-path mode as G^1, G^2, \dots, G^n and the corresponding node representation as $z_i^{G^1}, z_i^{G^2}, \dots, z_i^{G^n}$. Then, an attention mechanism was used to fuse and average all of the representations with their respective weight, defined by Equations (17)–(19).

$$w_{G_k} = \frac{1}{|v_c|} \sum_{i \in v_c} \lambda^T \cdot \tanh(W_c \cdot z_i^{G^k} + \varepsilon) \quad (17)$$

$$\beta_{G_k} = \text{SoftMax}(w_{G_k}) = \frac{\exp(w_{G_k})}{\sum_{j \in \{G^1, G^2, \dots, G^n\}} \exp(w_{G_j})} \quad (18)$$

$$z_i = \sum_{k=1}^n \beta_{G_k} \cdot z_i^{G^k} \quad (19)$$

where $W_c \in \mathbb{R}^{D \times d'}$ and $\varepsilon \in \mathbb{R}^D$ are the weight matrix and bias vector. v_c is a set of neighbor nodes in the same meta-path mode and $\lambda^T \in \mathbb{R}^D$ is the attention vector of all meta-paths for node type c . β_{G_k} indicates the final attention score after normalizing the importance contribution of a meta-path and z_i' is the final node representation.

4.2.3. Training and Prediction

Inspired by some matrix factorization or completion method, the latent feature of miRNA and disease was obtained by pre-training [59–62]. Specifically, the final node representation was used to reconstruct MDA by an inner product operation and the reconstruction error was reduced through minimizing the cross-entropy loss function defined by Equation (20):

$$L(a, a') = -[a \log a' + (1 - a) \log(1 - a')] \quad (20)$$

where a and a' are the original MDA and reconstruction MDA, respectively.

According to minimizing the loss function, the parameter mentioned above is constantly trained. When the well-trained latent feature of miRNAs and disease was acquired, the form of the miRNA–disease pair was concatenated as the input for SVM, which is a binary classifier with a significant accuracy and robustness in sparse and noise data. With the kernel function, it was also able to implement the non-linear classification and cater to the data complexity of miRNA and disease. After the prediction of SVM, the miRNA–disease pair association score can be obtained to indicate the association probability of the miRNA–disease pair.

4.3. Model Experiment and Evaluation

In this work, five-fold cross validation (5-CV) was utilized to evaluate the prediction performance of the model. The known MDA was considered as the positive sample and the stochastically selected identical amount of unobserved MDA as the negative sample. Positive and negative samples were combined into a dataset, which was randomly divided into five equal-sized subsets. Each subset was used as the test set in turn, and the remaining subsets were utilized as a training set. To reduce the variance cause by randomness, the

procedure was repeated 10 times. To evaluate the performance of the model, several evaluation measures were taken into account, including accuracy (Acc), precision (Pre), specificity (Spe), sensitivity (Sens, also called as recall), F1-measure(F1), Matthews correlation coefficient (Mcc), the areas under receiver operating characteristic (ROC) curve (AUC) and areas under precision-recall (PR) curve (AUPR). The Acc, Pre, recall and F1 were calculated by Equations (21)–(26):

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (21)$$

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (22)$$

$$\text{Spe} = \frac{\text{TN}}{\text{TP} + \text{FN}} \quad (23)$$

$$\text{Recall(Sens)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (24)$$

$$\text{F1} = \frac{2 * \text{Pre} * \text{Recall}}{\text{Pre} + \text{Recall}} \quad (25)$$

$$\text{Mcc} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TP} + \text{FP}) \times (\text{TN} + \text{FN})}} \quad (26)$$

where the TP, TN, FP and FN represent the number of true positives, true negatives, false positives and false negatives, respectively.

For the part of feature extraction, we pretrained it for 4000 epochs and employed an Adam optimizer with a learning rate 0.001. In addition, the other hyper parameters only affected the dimension of node representation, and the dimension was set in the range of (32, 64, 128, 256, 512). For the part of prediction with SVM, the penalty factor C was set as 150 and the radial basis function was used as a kernel function. In addition, the architecture and parameters of the model are listed in Table 7.

Table 7. The framework and parameters of model.

	Parameters
GAT	Input (1, 857, 857)
	Node attention layer (1, 857, 32) × 8, activation function
	Concatenate layer (1, 857, 256)
	Module attention layer (857, 256), activation function
	Dense layer (857, 256), activation function
	Learning rate (0.001)
	Epoch (2000)
SVM	Kernel function (radial basis function)
	C factor (50)

The framework of MMGAN-SVM is illustrated in Figure 12. First of all, as shown in Figure 12a, known miRNA–disease association was coordinated with Gaussian interaction profile kernel to calculate the miRNA (disease) integrated similarity (MS/DS). The miRNA–miRNA (disease–disease) relation network was built by adopting meta-path with the help of MDA, miRNA–protein association and disease–protein association. Second, as shown in Figure 12b, with the preparation above, combination matrix constructed the multi-module to be the input of the model. Then, as shown in Figure 12c,d, the latent feature of miRNA (disease) was acquired by model concatenate in the form of an miRNA–disease pair, which later served as the input of SVM for MDA prediction.

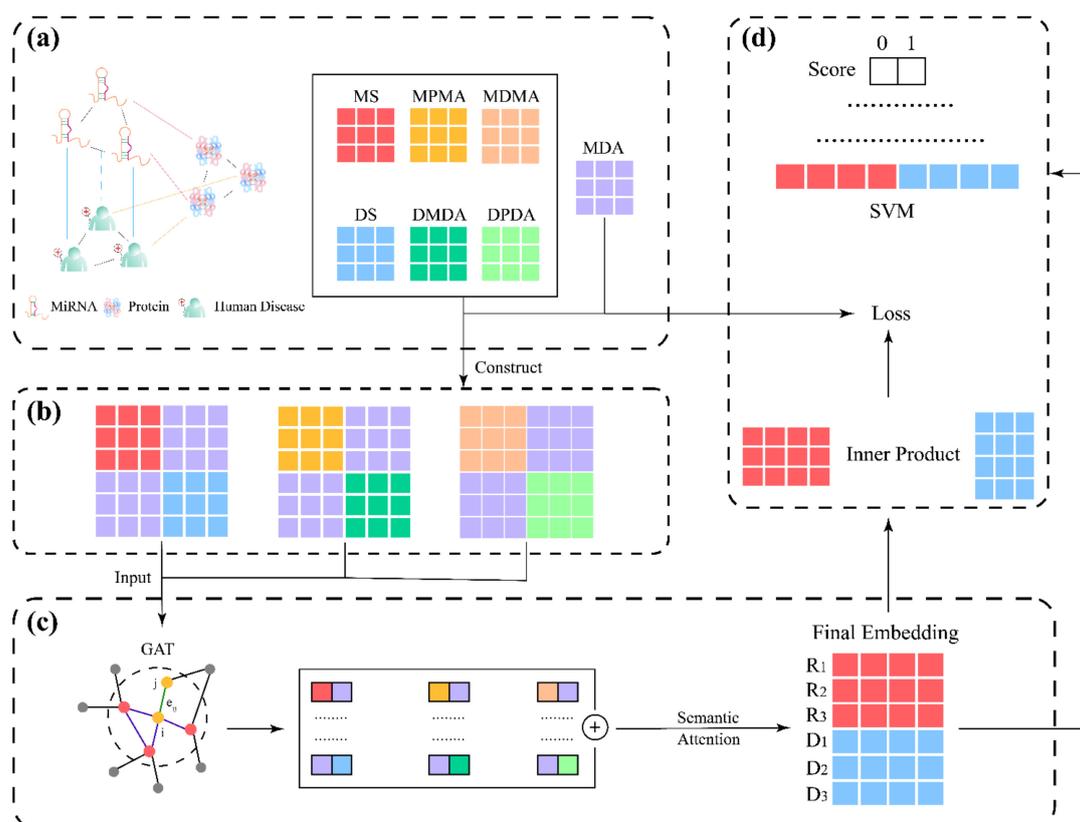


Figure 12. The flowchart of our method. (a) Construction of networks; (b) Construction of multi-module; (c) Feature extraction; (d) Model training and prediction.

5. Conclusions

In this study, extra protein information and meta-path were introduced to construct a multi-module, and GAT was utilized to learn the latent feature of the node in every module. Then, with the attention mechanism, the topological and semantic information of nodes could be aggregated adaptively by their different neighboring nodes and modules. With abundant information of latent features, the latter classification task conducted by SVM obtained a great performance in MDA prediction. In addition, the impact of different ratios of positive and negative samples on the model was explored. To some extent, the unbalance between negative and positive samples actually made some influences on the model. Thus, the strategy of reliable negative sample selection was adopted to reduce the impact of sample unbalance. In addition, the results showed that the performance of prediction can be improved by selecting negative samples within a certain threshold. In conclusion, we propose a new avenue for research to discovery potential biomarker and treatment for diseases.

Author Contributions: Conceptualization, Z.L. (Zihao Li) and X.H.; methodology, Z.L. (Zihao Li); software, Z.L. (Zihao Li) and Z.L. (Zhanchao Li); validation, Z.L. (Zihao Li), X.H. and Y.S.; formal analysis, Z.L. (Zhanchao Li), X.Z. and Z.D.; investigation, Y.S.; resources, Z.L. (Zhanchao Li), X.Z. and Z.D.; data curation, Z.L. (Zihao Li); writing—original draft preparation, Z.L. (Zihao Li); writing—review and editing, Z.L. (Zhanchao Li), X.Z. and Z.D.; visualization, Z.L. (Zhanchao Li); supervision, X.Z.; project administration, Z.D.; funding acquisition, Z.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Scientific Technology Project of Guangzhou City (no.202103000003) and the National Natural Science Foundation of China (no.21974153).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets and codes are available at <https://github.com/Excenmin/MMGAN-SVM> (accessed on 1 March 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Not available.

References

1. Berindan-Neagoe, I.; Monroig, P.d.C.; Pasculli, B.; Calin, G.A. MicroRNAome genome: A treasure for cancer diagnosis and therapy. *CA Cancer J. Clin.* **2014**, *64*, 311–336. [[CrossRef](#)] [[PubMed](#)]
2. Ambros, V. MicroRNAs: Tiny regulators with great potential. *Cell* **2001**, *107*, 823–826. [[CrossRef](#)]
3. Bartel, D.P. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **2004**, *116*, 281–297. [[CrossRef](#)]
4. Ambros, V. The functions of animal microRNAs. *Nature* **2004**, *431*, 350–355. [[CrossRef](#)]
5. Friedman, R.C.; Farh, K.K.-H.; Burge, C.B.; Bartel, D.P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **2009**, *19*, 92–105. [[CrossRef](#)] [[PubMed](#)]
6. Ha, M.; Kim, V.N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell. Biol.* **2014**, *15*, 509–524. [[CrossRef](#)] [[PubMed](#)]
7. Hua, S.; Yun, W.; Zhiqiang, Z.; Zou, Q. A discussion of micrornas in cancers. *Curr. Bioinform.* **2014**, *9*, 453–462. [[CrossRef](#)]
8. Das, J.; Podder, S.; Ghosh, T.C. Insights into the miRNA regulations in human disease genes. *BMC Genom.* **2014**, *15*, 1010. [[CrossRef](#)]
9. Santamaria, X.; Taylor, H. MicroRNA and gynecological reproductive diseases. *Fertil. Steril.* **2014**, *101*, 1545–1551. [[CrossRef](#)]
10. Condorelli, G.; Latronico, M.V.; Cavarretta, E. MicroRNAs in cardiovascular diseases: Current knowledge and the road ahead. *J. Am. Coll. Cardiol.* **2014**, *63*, 2177–2187. [[CrossRef](#)]
11. Gong, H.; Liu, C.M.; Liu, D.P.; Liang, C.C. The role of small RNAs in human diseases: Potential troublemaker and therapeutic tools. *Med. Res. Rev.* **2005**, *25*, 361–381. [[CrossRef](#)]
12. Dimić, D.S.; Kaluđerović, G.N.; Avdović, E.H.; Milenković, D.A.; Živanović, M.N.; Potočňák, I.; Samol'ová, E.; Dimitrijević, M.S.; Saso, L.; Marković, Z.S.; et al. Synthesis, Crystallographic, quantum chemical, antitumor, and molecular docking/dynamic studies of 4-hydroxycoumarin-neurotransmitter derivatives. *Int. J. Mol. Sci.* **2022**, *23*, 1001. [[CrossRef](#)]
13. Wang, S.-H.; Govindaraj, V.V.; Górriz, J.M.; Zhang, X.; Zhang, Y.-D. COVID-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. *Inf. Fusion* **2021**, *67*, 208–229. [[CrossRef](#)] [[PubMed](#)]
14. Li, Z.; Wang, Y.; Xie, Y.; Zhang, L.; Dai, Z.; Zou, X. Predicting the binding affinities of compound-protein interactions by random forest using network topology features. *Anal. Methods* **2018**, *10*, 4152–4161. [[CrossRef](#)]
15. Jiang, Q.; Hao, Y.; Wang, G.; Juan, L.; Zhang, T.; Teng, M.; Liu, Y.; Wang, Y. Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst. Biol.* **2010**, *4*, S2. [[CrossRef](#)] [[PubMed](#)]
16. Chen, X.; Yan, C.C.; Zhang, X.; You, Z.H.; Deng, L.; Liu, Y.; Zhang, Y.; Dai, Q. WBSMDA: Within and between Score for MiRNA-Disease Association prediction. *Sci. Rep.* **2016**, *6*, 21106. [[CrossRef](#)] [[PubMed](#)]
17. Pasquier, C.; Gardès, J. Prediction of miRNA-disease associations with a vector space model. *Sci. Rep.* **2016**, *6*, 27036. [[CrossRef](#)]
18. Chen, X.; Liu, M.X.; Yan, G.Y. RWRMDA: Predicting novel human microRNA-disease associations. *Mol. Biosyst.* **2012**, *8*, 2792–2798. [[CrossRef](#)]
19. Xuan, P.; Han, K.; Guo, Y.; Li, J.; Li, X.; Zhong, Y.; Zhang, Z.; Ding, J. Prediction of potential disease-associated microRNAs based on random walk. *Bioinformatics* **2015**, *31*, 1805–1815. [[CrossRef](#)]
20. You, Z.-H.; Huang, Z.-A.; Zhu, Z.; Yan, G.-Y.; Li, Z.-W.; Wen, Z.; Chen, X. PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. *PLoS Comput. Biol.* **2017**, *13*, e1005455. [[CrossRef](#)]
21. Chen, M.; Liao, B.; Li, Z. Global similarity method based on a two-tier random walk for the prediction of microRNA-disease association. *Sci. Rep.* **2018**, *8*, 6481. [[CrossRef](#)] [[PubMed](#)]
22. Hussain, I.; Hossain, M.A.; Jany, R.; Bari, M.A.; Uddin, M.; Kamal, A.R.M.; Ku, Y.; Kim, J.-S. Quantitative Evaluation of EEG-Biomarkers for Prediction of Sleep Stages. *Sensors* **2022**, *22*, 3079. [[CrossRef](#)] [[PubMed](#)]
23. Hussain, I.; Park, S.J. HealthSOS: Real-time health monitoring system for stroke prognostics. *IEEE Access* **2020**, *8*, 213574–213586. [[CrossRef](#)]
24. Gao, Z.; Liu, X.; Qi, S.; Wu, W.; Hau, W.K.; Zhang, H. Automatic segmentation of coronary tree in CT angiography images. *Int. J. Adapt. Control. Signal Process.* **2019**, *33*, 1239–1247. [[CrossRef](#)]
25. Jiang, Q.; Wang, G.; Jin, S.; Li, Y.; Wang, Y. Predicting human microRNA-disease associations based on support vector machine. *Int. J. Data Min. Bioinform.* **2013**, *8*, 282–293. [[CrossRef](#)]
26. Zhao, Y.; Chen, X.; Yin, J. Adaptive boosting-based computational model for predicting potential miRNA-disease associations. *Bioinformatics* **2019**, *35*, 4730–4738. [[CrossRef](#)]
27. Li, J.; Zhang, S.; Liu, T.; Ning, C.; Zhang, Z.; Zhou, W. Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction. *Bioinformatics* **2020**, *36*, 2538–2546. [[CrossRef](#)]
28. Xuan, P.; Sun, H.; Wang, X.; Zhang, T.; Pan, S. Inferring the Disease-Associated miRNAs Based on Network Representation Learning and Convolutional Neural Networks. *Int. J. Mol. Sci.* **2019**, *20*, 3648. [[CrossRef](#)]

29. Ji, C.; Gao, Z.; Ma, X.; Wu, Q.; Ni, J.; Zheng, C. AEMDA: Inferring miRNA-disease associations based on deep autoencoder. *Bioinformatics* **2021**, *37*, 66–72. [[CrossRef](#)]
30. Jin, S.; Zeng, X.; Xia, F.; Huang, W.; Liu, X. Application of deep learning methods in biological networks. *Brief. Bioinform.* **2021**, *22*, 1902–1917. [[CrossRef](#)]
31. Zhao, T.; Hu, Y.; Cheng, L. Deep-DRM: A computational method for identifying disease-related metabolites based on graph deep learning approaches. *Brief. Bioinform.* **2021**, *22*, bbaa212. [[CrossRef](#)]
32. Sun, M.; Zhao, S.; Gilvary, C.; Elemento, O.; Zhou, J.; Wang, F. Graph convolutional networks for computational drug development and discovery. *Brief. Bioinform.* **2020**, *21*, 919–935. [[CrossRef](#)]
33. Yue, X.; Wang, Z.; Huang, J.; Parthasarathy, S.; Moosavinasab, S.; Huang, Y.; Lin, S.M.; Zhang, W.; Zhang, P.; Sun, H. Graph embedding on biomedical networks: Methods, applications and evaluations. *Bioinformatics* **2020**, *36*, 1241–1251. [[CrossRef](#)]
34. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint* **2016**, arXiv:1609.02907.
35. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv preprint* **2017**, arXiv:1710.10903.
36. Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; Yu, P.S. Heterogeneous graph attention network. In Proceedings of the 26th International World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 2022–2032.
37. Xie, G.; Fan, Z.; Sun, Y.; Wu, C.; Ma, L. WBNPMD: Weighted bipartite network projection for microRNA-disease association prediction. *J. Transl. Med.* **2019**, *17*, 322. [[CrossRef](#)]
38. Yan, C.; Wang, J.; Ni, P.; Lan, W.; Wu, F.-X.; Pan, Y. DNRLMF-MDA: Predicting microRNA-disease associations based on similarities of microRNAs and diseases. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**, *16*, 233–243. [[CrossRef](#)]
39. Ding, Y.; Tian, L.-P.; Lei, X.; Liao, B.; Wu, F.-X. Variational graph auto-encoders for miRNA-disease association prediction. *Methods* **2021**, *192*, 25–34. [[CrossRef](#)]
40. Yang, Z.; Wu, L.; Wang, A.; Tang, W.; Zhao, Y.; Zhao, H.; Teschendorff, A.E. dbDEMC 2.0: Updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res.* **2017**, *45*, D812–D818. [[CrossRef](#)]
41. Mohammadian, M.; Mahdaviifar, N.; Mohammadian-Hafshejani, A.; Salehiniya, H. Liver cancer in the world: Epidemiology, incidence, mortality and risk factors. *World Cancer Res. J.* **2018**, *5*, 8.
42. El-Serag, H.B.; Rudolph, K.L. Hepatocellular carcinoma: Epidemiology and molecular carcinogenesis. *Gastroenterology* **2007**, *132*, 2557–2576. [[CrossRef](#)]
43. Pineau, P.; Volinia, S.; McJunkin, K.; Marchio, A.; Battiston, C.; Terris, B.; Mazzaferro, V.; Lowe, S.W.; Croce, C.M.; Dejean, A. miR-221 overexpression contributes to liver tumorigenesis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 264–269. [[CrossRef](#)]
44. Han, Z.-B.; Chen, H.-Y.; Fan, J.-W.; Wu, J.-Y.; Tang, H.-M.; Peng, Z.-H. Up-regulation of microRNA-155 promotes cancer cell invasion and predicts poor survival of hepatocellular carcinoma following liver transplantation. *J. Cancer Res. Clin. Oncol.* **2012**, *138*, 153–161. [[CrossRef](#)]
45. Inamura, K. Lung cancer: Understanding its molecular pathology and the 2015 WHO classification. *Front. Oncol.* **2017**, *7*, 193. [[CrossRef](#)]
46. Fan, L.; Qi, H.; Teng, J.; Su, B.; Chen, H.; Wang, C.; Xia, Q. Identification of serum miRNAs by nano-quantum dots microarray as diagnostic biomarkers for early detection of non-small cell lung cancer. *Tumor Biol.* **2016**, *37*, 7777–7784. [[CrossRef](#)]
47. Zhang, H.; Mao, F.; Shen, T.; Luo, Q.; Ding, Z.; Qian, L.; Huang, J. Plasma miR-145, miR-20a, miR-21 and miR-223 as novel biomarkers for screening early-stage non-small cell lung cancer. *Oncol. Lett.* **2017**, *13*, 669–676. [[CrossRef](#)]
48. Florean, C.; Schneckeburger, M.; Grandjennette, C.; Dicato, M.; Diederich, M. Epigenomics of leukemia: From mechanisms to therapeutic applications. *Epigenomics* **2011**, *3*, 581–609. [[CrossRef](#)]
49. Ning, F.; Zhou, Q.; Chen, X. miR-200b promotes cell proliferation and invasion in t-cell acute Lymphoblastic leukemia through NOTCH1. *J. Biol. Regul. Homeost. Agents* **2018**, *32*, 1467–1471.
50. Sanghvi, V.R.; Mavrikakis, K.J.; Van der Meulen, J.; Boice, M.; Wolfe, A.L.; Carty, M.; Mohan, P.; Rondou, P.; Socci, N.D.; Benoit, Y. Characterization of a set of tumor suppressor microRNAs in T cell acute lymphoblastic leukemia. *Sci. Signal.* **2014**, *7*, ra111. [[CrossRef](#)]
51. Xiao, Y.; Su, C.; Deng, T. miR-223 decreases cell proliferation and enhances cell apoptosis in acute myeloid leukemia via targeting FBXW7. *Oncol. Lett.* **2016**, *12*, 3531–3536. [[CrossRef](#)]
52. Li, Y.; Qiu, C.; Tu, J.; Geng, B.; Yang, J.; Jiang, T.; Cui, Q. HMDD v2.0: A database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* **2014**, *42*, D1070–D1074. [[CrossRef](#)]
53. Davis, A.P.; Grondin, C.J.; Johnson, R.J.; Sciaky, D.; Wieggers, J.; Wieggers, T.C.; Mattingly, C.J. Comparative toxicogenomics database (CTD): Update 2021. *Nucleic Acids Res.* **2021**, *49*, D1138–D1143. [[CrossRef](#)]
54. Hsu, S.-D.; Lin, F.-M.; Wu, W.-Y.; Liang, C.; Huang, W.-C.; Chan, W.-L.; Tsai, W.-T.; Chen, G.-Z.; Lee, C.-J.; Chiu, C.-M. miRTarBase: A database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **2011**, *39*, D163–D169. [[CrossRef](#)]
55. Van Laarhoven, T.; Nabuurs, S.B.; Marchiori, E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* **2011**, *27*, 3036–3043. [[CrossRef](#)]
56. Wang, D.; Wang, J.; Lu, M.; Song, F.; Cui, Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* **2010**, *26*, 1644–1650. [[CrossRef](#)]
57. Cheng, L.; Li, J.; Ju, P.; Peng, J.; Wang, Y. SemFunSim: A new method for measuring disease similarity by integrating semantic and gene functional association. *PLoS ONE* **2014**, *9*, e99415. [[CrossRef](#)]

58. Sun, Y.; Han, J.; Yan, X.; Yu, P.S.; Wu, T. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *Proc. VLDB Endow.* **2011**, *4*, 992–1003. [[CrossRef](#)]
59. Xiao, Q.; Luo, J.; Liang, C.; Cai, J.; Ding, P. A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics* **2018**, *34*, 239–248. [[CrossRef](#)]
60. Chen, X.; Yin, J.; Qu, J.; Huang, L. MDHGI: Matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction. *PLoS Comput. Biol.* **2018**, *14*, e1006418. [[CrossRef](#)]
61. Chen, X.; Sun, L.-G.; Zhao, Y. NCMCMDA: miRNA-disease association prediction through neighborhood constraint matrix completion. *Brief. Bioinform.* **2021**, *22*, 485–496. [[CrossRef](#)]
62. Chen, X.; Wang, L.; Qu, J.; Guan, N.-N.; Li, J.-Q. Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* **2018**, *34*, 4256–4265. [[CrossRef](#)]