



Article HLGNN-MDA: Heuristic Learning Based on Graph Neural Networks for miRNA–Disease Association Prediction

Liang Yu *^D, Bingyi Ju and Shujie Ren

School of Computer Science and Technology, Xidian University, Xi'an 710071, China * Correspondence: lyu@xidian.edu.cn; Tel.: +86-137-5992-1156

Abstract: Identifying disease-related miRNAs can improve the understanding of complex diseases. However, experimentally finding the association between miRNAs and diseases is expensive in terms of time and resources. The computational screening of reliable miRNA-disease associations has thus become a necessary tool to guide biological experiments. "Similar miRNAs will be associated with the same disease" is the assumption on which most current miRNA-disease association prediction methods rely; however, biased prior knowledge, and incomplete and inaccurate miRNA similarity data and disease similarity data limit the performance of the model. Here, we propose heuristic learning based on graph neural networks to predict microRNA-disease associations (HLGNN-MDA). We learn the local graph topology features of the predicted miRNA-disease node pairs using graph neural networks. In particular, our improvements to the graph convolution layer of the graph neural network enable it to learn information among homogeneous nodes and among heterogeneous nodes. We illustrate the performance of HLGNN-MDA by performing tenfold cross-validation against excellent baseline models. The results show that we have promising performance in multiple metrics. We also focus on the role of the improvements to the graph convolution layer in the model. The case studies are supported by evidence on breast cancer, hepatocellular carcinoma and renal cell carcinoma. Given the above, the experiments demonstrate that HLGNN-MDA can serve as a reliable method to identify novel miRNA-disease associations.

Keywords: miRNA-disease association; graph neural network; graph classification; heuristics learning

1. Introduction

Since the discovery of microRNAs, an increasing number of researchers have investigated these molecules [1–8]. In particular, the discovery of a regulatory role for microRNAs in cellular activity suggests that these molecules are inextricably linked to many diseases [9–13]. Uncovering microRNA–disease associations has important implications for understanding disease mechanisms and assisting disease treatment [14–20]. However, due to the long time period required for biological experiments and the high resource costs, the use of computational methods to predict miRNA–disease associations has now become an important means for guiding traditional biological experiments, and it has greatly improved the efficiency of discovering disease-related miRNAs.

Some miRNA–disease association prediction models derive from combinatorial optimization theory and metric learning ideas, such as matrix-related operations and score estimation. MCMDA (matrix completion for miRNA–disease association) [21] performs matrix completion by applying a singular value thresholding algorithm on known miRNA– disease associations, and ILRMR (improved low-rank matrix recovery) [22] improves lowrank matrix recovery by referencing a weight matrix to enhance the prediction accuracy. MDMF (miRNA–disease Based on Matrix Factorization) [23] uses matrix factorization with disease similarity constraints to identify potential miRNA–disease associations. MDHGI (Decomposition and Heterogeneous Graph Inference) [24] discovers new miRNA–diseases associations by integrating the predicted association probability obtained from matrix



Citation: Yu, L.; Ju, B.; Ren, S. HLGNN-MDA: Heuristic Learning Based on Graph Neural Networks for miRNA–Disease Association Prediction. *Int. J. Mol. Sci.* 2022, 23, 13155. https://doi.org/10.3390/ ijms232113155

Academic Editors: Samuel De Visser and Lorenzo Lo Muzio

Received: 16 August 2022 Accepted: 26 October 2022 Published: 29 October 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/). decomposition through the sparse learning method. IMIPMF (inferring miRNA–disease interactions using probabilistic matrix factorization) [25] is a novel method for predicting miRNA–disease associations using probabilistic matrix factorization. WBSMDA (Within and Between Score for MiRNA–Disease Association prediction) [26] calculates a "Within and Between Score" for each miRNA–disease pair to predict the association between them. MLMD (Metric Learning for predicting miRNA–Disease) [27] is a novel computational model of metric learning for predicting miRNA–disease associations. It aims at learning miRNA–disease metrics to unravel not only novel disease-related miRNAs but also miRNA–miRNA and disease–disease similarities. DBNMDA (deep-belief network for predicting miRNA–disease associations) [28] constructs feature vectors to pre-train restricted Boltzmann machines for all miRNA–disease pairs and applies positive samples and the same number of selected negative samples to fine-tune a deep-belief network to obtain the final predicted scores.

Machine learning is also a class of methods [29,30] that are commonly applied to predict miRNA–disease associations [20,31–37]. RBMMMDA (restricted Boltzmann machine for multiple types of miRNA–disease associations) [38] proposes the restricted Boltzmann machine model to predict various types of miRNA–disease associations.

With the development of graph neural networks and the accumulation of large-scale graph data, in addition to traditional machine learning algorithms, DGCNN (multi-view multi-layer convolutional neural network) [39] and other deep learning [40-45] models have also been developed to deal with similar tasks. DGCNN focuses on large-scale and irregular network structures and adapts to the dynamic structure of local regions in the graph by flexibly designing convolutional filters. DeepMDA (predict miRNA–disease associations using deep learning) [46] uses a stacked self-encoder to obtain low-dimensional features from two high-dimensional feature vectors of miRNAs and diseases. A three-layer deep neural network [47] has then been developed to train classifiers of miRNA-disease feature pairs. MDA-CNN [48] constructs a three-layer miRNA-gene-disease association (MDA) network, and the network-based features of miRNAs and diseases are extracted using genes as the intermediate medium. The features are then downscaled using a selfencoder. Convolutional neural networks (CNNs) [49] are then used to further learn features from the miRNA-disease feature pairs. MDA-GCNFTG [50] predicts associations based on graph convolutional networks via graph sampling through the feature and topology graph to improve the training efficiency and accuracy. Instead of using heterogeneous graphs, MDA-GCNFTG constructs a homogeneous graph with MDPs (miRNA-disease pairs) as the nodes, which is the biggest difference with respect to our method. Although both use the GCN algorithm, the background graphs and the model focused on are completely different. For this task, the models focus on solving different problems.

However, there are still shortcomings in these recently proposed excellent computational models. Most of the current methods for predicting miRNA–disease associations are based on a strong assumption of similarity data. However, different models have different definitions of similarity, which makes the prediction results inaccurate. In addition, the miRNA functional similarity is incomplete and derived from known associations. Additionally, inconsistencies and incompleteness further lead to inaccurate prediction results. In this article, a heuristic learning method based on graph neural networks for miRNA–disease association prediction (HLGNN-MDA) is proposed. Inputting the whole miRNA–disease association network into the graph neural network for the model training causes high computational costs. To overcome this, we choose to train the graph neural network on enclosing subgraphs.

Figure 1 shows the overall framework of HLGNN-MDA. Our HLGNN-MDA model improves the graph neural network so that it can learn the information between miRNA and disease nodes and the topological relationships among homogeneous nodes at the same time. Compared with the previous computational models, our proposed method does not require too many similarity data and improves the applicability of graph neural networks

a. Extract enclosing subgraph and node labeling miRNA negative sample b. Predict miRNA-disease association by graph neural network $D^{-1}A \cdot X$ $D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \cdot X$ A2 . X Concat wise Fully Con Graph convolution layers h Verification and analysis of Graph pooling layers prediction results 000000000 00000000 HMDD v3.2 dbDEMC 2.0 0000000 miR2Disease Base sort a 1 -P Known miRNA-dis Potential miRNA-disease association ully-0

in bipartite graph networks. More importantly, under such conditions, HLGNN-MDA can also achieve more accurate predictions.

Figure 1. Flowchart of HLGNN-MDA. (**a**) The enclosing subgraph of all pairs of nodes are extracted, and all the nodes in each enclosing subgraph are labeled. (**b**) The enclosing subgraphs are input into the graph neural network. The graph convolution layers adopt the three-layer structure model as shown in Figure 2 where each sub-figure of three in the second part corresponds to the processing results of three convolution modules. As shown in Figure 3, each convolution layer, the graph pooling layer, 1D convolution and fully connected layers. Finally, the predicted results are verified against databases.



Figure 2. Three-layer graph convolutional layer structure of HLGNN-MDA.



Figure 3. Design of the graph convolution block for HLGNN-MDA.

2. Results and Discussion

2.1. Performance Analysis of HLGNN-MDA Mode

In this section, we compare HLGNN-MDA with other related methods using tenfold cross-validation [20]. The algorithms selected for comparison were BLHARMDA (bipartite local models and hubness-aware regression for miRNA–disease association prediction) [51], BNPMDA (bipartite network projection for miRNA–disease association prediction) [52], IMCMDA (inductive matrix completion for miRNA–disease association prediction) [52], LFEMDA (predict miRNA–disease associations by latent features extraction) [53] and MKR-MDA (multiple kernel learning-based Kronecker regularized least squares for miRNA–disease association prediction) [54]. All models for performance analysis experiments adopted the association data from HMDD v2.0. We also performed a performance comparison with the DGCNN model in the subsequent experiments of graph convolutional layer analysis, which is the baseline of our HLGNN-MDA convolutional module.

The miRNA functional similarity and disease semantic similarity were downloaded directly from IMCMDA. HLGNN-MDA used 5430 known miRNA–disease associations as positive samples and a random sample of the same number of candidate associations as negative samples. The dataset with the positive and negative samples together was then randomly divided into ten parts. One copy of each round was selected as the test set, and the remaining nine copies were used as the training set. The tenfold cross-validation was completed in turn. In each round, HLGNN-MDA removed the positive samples in the test set from the adjacency matrix. All five comparison algorithms completed tenfold cross-validation on the miRNA–disease association matrix. There were six evaluation metrics to analyze the model: ACC, precision, recall, AUROC, AUPR and MCC. The results are shown in Table 1, and their corresponding ROC curves are shown in Figure 4. The corresponding precision–recall (PR) curves are shown in Figure 5.

Model	ACC	Precision	Recall	AUROC	AUPR	MCC
BNPMDA	0.79088	0.87069	0.68324	0.85648	0.88275	0.59574
IMCMDA	0.77274	0.80102	0.72578	0.84004	0.84989	0.54791
LFEMDA	0.84751	0.85590	0.83573	0.90039	0.91289	0.69522
BLHARMDA	0.85442	0.85619	0.85193	0.92838	0.92699	0.70885
MKRMDA	0.84549	0.87610	0.80479	0.89658	0.91971	0.69328
HLGNN-MDA-hop1	0.85442	0.86263	0.84309	0.92974	0.92779	0.70902
HLGNN-MDA-hop2	0.85976	0.85917	0.86059	0.92833	0.92927	0.71952
HLGNN-MDA-hop3	0.85635	0.86745	0.84125	0.92863	0.93007	0.71303
HLGNN-MDA-hop4	0.85912	0.86709	0.84825	0.93086	0.93247	0.71840

Table 1. Performance analysis of HLGNN-MDA and state-of-the-art miRNA-disease association

 prediction algorithms. The bold part indicates the maximum value in the corresponding column.



Figure 4. ROC curve of HLGNN-MDA and state-of-the-art miRNA–disease association prediction algorithms.



Figure 5. PR curve of HLGNN-MDA and state-of-the-art miRNA-disease association prediction algorithms.

In Table 1, HLGNN-MDA-hopx represents the HLGNN-MDA model in which the enclosing subgraph's hop is x (x = 1, 2, 3 or 4). Hops were used to extract subgraphs. The extraction process is described in Section 3.2.1. Compared with other methods, HLGNN-MDA-hopx had high performance in five of the six indicators. The minimum AUROC of HLGNN-MDA-hopx was greater than those of BNPMDA, IMCMDA, LFEMDA and MKRMDA. The minimum value of the AUPR of HLGNN-MDA-hopx was greater than the results of all other algorithms.

From the ROC curve in Figure 4, HLGNN-MDA-hop4 could cover the curves of BNPMDA, IMCMDA and MKRMDA. The maximum AUROC was 0.25% larger than the AUROC of BLHARMDA. Similarly, in the PR curve of Figure 5, HLGNN-MDA-hop4 could cover the PR curves of BNPMDA, IMCMDA and MKRMDA, which were roughly the same as that of BLHARMDA. The maximum AUPR value of HLGNN-MDA-hop4 was 0.55% larger than the maximum value of BLHARMDA.

Compared with the other algorithms, HLGNN-MDA had a relatively large advantage. Moreover, HLGNN-MDA used less similarity information to obtain better prediction results.

2.2. Influence of Different Hops in the Enclosing Subgraph

In this section, we discuss how HLGNN-MDA was affected by different enclosing subgraph hops. The required experimental data were miRNA–disease associations. The positive samples represented the known associations, and a random sample of the same number of candidates represented the negative samples. The test set was one-tenth the size of the entire dataset.

First, we evaluated the results of the HLGNN-MDA model using different hops in the enclosed subgraph. Each model was trained with the same training set and evaluated on the same test set. The range of the number of hops in the enclosing subgraph was 1 to 4. Their corresponding results are shown in Table 2. The ROC, PR and accuracy curves under different thresholds are shown in Figures 6–8.

Table 2. Influence of different hops in enclosing subgraphs on HLGNN-MDA. The bold part indicates the maximum value in the corresponding column.

Model	ACC	Precision	Recall	AUROC	AUPR	MCC
HLGNN-MDA-hop1	0.88122	0.90430	0.85267	0.93535	0.93281	0.76368
HLGNN-MDA-hop2	0.92726	0.93939	0.91344	0.97212	0.97564	0.85484
HLGNN-MDA-hop3	0.93831	0.95946	0.91529	0.97266	0.97744	0.87754
HLGNN-MDA-hop4	0.96869	0.96690	0.97053	0.99178	0.99332	0.93739



Figure 6. ROC curve of HLGNN-MDA in different enclosing subgraphs.



Figure 7. PR curve of HLGNN-MDA in different enclosing subgraphs.



Figure 8. Accuracy curve of HLGNN-MDA using different enclosing subgraph hops.

As shown in Table 2, HLGNN-MDA obtained the best value for all six metrics when the hops were 4. In Figures 6–8, it can be seen that the curves with less hops were always covered by curves with larger hops. In particular, it should be noted that the results of HLGNN-MDA-hop2 had a larger increase than the results of HLGNN-MDA-hop1. The results of HLGNN-MDA-hop4 had a larger increase than the results of HLGNN-MDA-hop3. There was only a slight increase between HLGNN-MDA-hop3 and HLGNN-MDA-hop2, as shown in the three figures. These findings show that using an even larger hop in enclosing subgraphs for miRNA–disease association prediction could obtain better results.

2.3. Analysis of the Improved Graph Convolutional Layer

To show that our improvement in the graph neural network was effective, in this section, we present the analysis of the improved graph convolutional layer. After considering the information transfer between homogeneous nodes and heterogeneous nodes, HLGNN-MDA aggregated four propagation functions in the graph convolutional layer: A, A^2 , $D^{-1}A$ and $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$.

Next, we explored the role of each propagation function in graph neural networks. If we deleted a certain propagation function from the current HLGNN-MDA but obtained a

better result, it showed that the propagation function had a negative effect on the graph neural network. If the result of deleting a certain propagation function was worse, it had a positive effect on the graph neural network. Therefore, we discussed the role of the four propagation functions in turn in this way. The HLGNN-MDA model that removed propagation function *A* was marked as HLGNN-MDA-a. In this order, the HLGNN-MDA model deleting A^2 was marked as HLGNN-MDA-b; the model without $D^{-1}A$ was marked as HLGNN-MDA-d.

Then, we compared the HLGNN-MDA model with its four variations: HLGNN-MDA-a, HLGNN-MDA-b, HLGNN-MDA-c and HLGNN-MDA-d. As shown in Figure 9, the horizontal axis is the hop in the enclosing subgraph, and the vertical axis is the AUROC. The red line indicates HLGNN-MDA, and the yellow line indicates HLGNN-MDA-b, wherein A^2 was deleted from HLGNN-MDA; these variations were the most disparate. These results showed that A^2 had a greater impact on the model. The difference between the green fold of HLGNN-MDA-a and the red fold of HLGNN-MDA was minimal. However, when the hop number was 1, 2 or 4, HLGNN-MDA gave better results than HLGNN-MDA-a. Meanwhile, HLGNN-MDA-a and HLGNN-MDA-b all reached a maximum with hop = 3 and decreased slightly with hop = 4. Both propagation functions A and A^2 played an effective role in improving the predictive performance of the HLGNN-MDA model. From the figures, we could find that $D^{-1}A$ (purple line) and $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ (blue line) had similar roles in graph neural networks and complemented each other. Meanwhile, if the evaluation metrics in Figure 9 were replaced with the AUPR and the ACC, a similar trend could be obtained.



Figure 9. AUC comparison between HLGNN-MDA and its four variations under different hops.

Furthermore, since the graph neural network [55,56] of HLGNN-MDA is improved with respect to DGCNN, in order to show the effectiveness of HLGNN-MDA model, we also compared the performance of four variants of HLGNN-MDA with DGCNN with different hops. From Table 3, it could be concluded that DGCNN and HLGNN-MDA-b performed similarly. HLGNN-MDA-b was slightly higher than DGCNN when the enclosing subgraph hops were 1, 2 and 3. Compared with other HLGNN-MDA variant models, HLGNN-MDA-b, i.e., the A^2 -deleted model, had the worst performance among all models. As the value of hops increased, the six measures of HLGNN-MDA-b also increased slowly, and when *hop* = 3, HLGNN-MDA-b reached a maximum. This finding was consistent with the results presented in the above figures; that is, A^2 played a large role in the HLGNN-MDA model. Overall, HLGNN-MDA performed better than DGCNN in predicting miRNA-disease associations.

Model	ACC	Precision	Recall	AUROC	AUPR	MCC
HLGNN-MDA-a-hop1	0.85820	0.91121	0.79374	0.92795	0.92303	0.72242
HLGNN-MDA-a-hop2	0.90055	0.90503	0.89503	0.94681	0.94629	0.80115
HLGNN-MDA-a-hop3	0.94843	0.94516	0.95212	0.98538	0.98626	0.89689
HLGNN-MDA-a-hop4	0.93186	0.95183	0.90976	0.97535	0.97945	0.86456
HLGNN-MDA-b-hop1	0.86096	0.90164	0.81031	0.93369	0.93412	0.72565
HLGNN-MDA-b-hop2	0.87845	0.92371	0.82505	0.93721	0.94106	0.76126
HLGNN-MDA-b-hop3	0.89042	0.91569	0.86004	0.94265	0.94575	0.78229
HLGNN-MDA-b-hop4	0.88858	0.90734	0.86556	0.94257	0.94877	0.77799
HLGNN-MDA-c-hop1	0.85635	0.84127	0.87845	0.92729	0.92241	0.71340
HLGNN-MDA-c-hop2	0.92265	0.91652	0.93002	0.96673	0.96729	0.84540
HLGNN-MDA-c-hop3	0.88398	0.92464	0.83610	0.93691	0.94300	0.77150
HLGNN-MDA-c-hop4	0.93923	0.96311	0.91344	0.97610	0.97901	0.87962
HLGNN-MDA-d-hop1	0.86280	0.89879	0.81768	0.92999	0.92586	0.72857
HLGNN-MDA-d-hop2	0.93831	0.94238	0.93370	0.98012	0.98081	0.87665
HLGNN-MDA-d-hop3	0.87569	0.92324	0.81952	0.93266	0.94060	0.75617
HLGNN-MDA-d-hop4	0.94015	0.95437	0.92449	0.98585	0.98702	0.88073
DGCNN-hop1	0.85820	0.88822	0.81952	0.92889	0.92889	0.71854
DGCNN-hop2	0.87201	0.90400	0.83241	0.93509	0.93831	0.74636
DGCNN-hop3	0.88582	0.90522	0.86188	0.94241	0.94083	0.88302
DGCNN-hop4	0.89411	0.91961	0.86372	0.95250	0.95707	0.89079

Table 3. Comparison between the four variations of HLGNN-MDA and DGCNN with different hops.

In summary, the four propagation functions in HLGNN-MDA all played a positive role, thus leading HLGNN-MDA to achieve good predictive performance. Of these, A and A^2 had a stronger role in improving the predictive performance of the model, while $D^{-1}A$ and $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ enabled the model to be stable in its results with different hops of the enclosing subgraph. The combined use of these four propagation functions allowed HLGNN-MDA to perform well in the task of predicting miRNA–disease associations.

2.4. Validation of Prediction Results

In this section, all the known correlations in HMDD v2.0 were the training set for the model, and as many potential associations as possible were sampled as negative samples. According to the above analysis, HLGNN-MDA-hop4 gave the best predictions, so training was performed on it. All potential association relationships in HMDD v2.0 were then extracted and predicted on the trained HLGNN-MDA-hop4.

The prediction results were validated using the following databases: HMDD v3.0 [57], dbDEMC [58] and miR2Disease [59]. One association was considered to be validated if it was found in at least one database.

The final validation results were as follows: 10 out of the top 10 predictions were validated; a total of 49 out of the top 50 predictions were verified; a total of 97 out of the top 100 predictions were verified; and 169 out of the top 180 predictions were verified. The results demonstrated the effectiveness of HLGNN-MDA and its ability to predict potential novel associations.

2.5. Case Study

2.5.1. Breast Cancer

Breast cancer is one of the most dangerous malignancies to human health, especially for women. Globally, breast cancer accounts for 2.088 million new cases and 627,000 deaths per year, making it the number one malignancy in women [60]. The top 50 miRNAs predicted to be associated with breast cancer are listed in Table 4. In total, 49 of them were found in the relevant validation database. Only hsa-mir-362 (validation = no) does not present a record related to breast cancer in the three databases at present. Based on the literature validation [61], the hERG potassium channel, which enhances tumor

aggressiveness and breast cancer proliferation, is transcriptionally regulated by hsa-miR-362-3p and thus associated with breast cancer growth. Another study [62] compared MDA-MB-231 and MCF7 breast cancer cell lines to the control CCD-1095Sk cell line, where hsa-miR-362-5p showed significant upregulation. The inhibition of hsa-miR-362-5p was found to significantly inhibit the diffusion, migration and invasion of MCF7 human breast cancer cells.

Rank	MicroRNA	Validation	Rank	MicroRNA	Validation
1	hsa-mir-211	yes <h, d=""></h,>	26	hsa-mir-30e	yes <h, d=""></h,>
2	hsa-mir-186	yes <d></d>	27	hsa-mir-494	yes <h, d=""></h,>
3	hsa-mir-744	yes <h, d=""></h,>	28	hsa-mir-421	yes <h, d=""></h,>
4	hsa-mir-138	yes <h, d=""></h,>	29	hsa-mir-501	yes <h, d=""></h,>
5	hsa-mir-154	yes <d></d>	30	hsa-mir-99b	yes <h, d=""></h,>
6	hsa-mir-216b	yes <h, d=""></h,>	31	hsa-mir-196b	yes <h, d=""></h,>
7	hsa-mir-106a	yes <h, d=""></h,>	32	hsa-mir-185	yes <h, d=""></h,>
8	hsa-mir-432	yes <h, d=""></h,>	33	hsa-mir-484	yes <h, d=""></h,>
9	hsa-mir-32	yes <h, d=""></h,>	34	hsa-mir-144	yes <h, d=""></h,>
10	hsa-mir-381	yes <h, d=""></h,>	35	hsa-mir-592	yes <h, d=""></h,>
11	hsa-mir-142	yes <h, d=""></h,>	36	hsa-mir-130a	yes <h, d=""></h,>
12	hsa-mir-150	yes <h, d=""></h,>	37	hsa-mir-542	yes <h, d=""></h,>
13	hsa-mir-491	yes <h, d=""></h,>	38	hsa-mir-1224	yes <h, d=""></h,>
14	hsa-mir-449a	yes <h, d=""></h,>	39	hsa-mir-376a	yes <h, d=""></h,>
15	hsa-mir-362	no	40	hsa-mir-451	yes <h, d,="" m=""></h,>
16	hsa-mir-28	yes <h, d=""></h,>	41	hsa-mir-433	yes <h, d=""></h,>
17	hsa-mir-378a	yes <h, d=""></h,>	42	hsa-mir-483	yes <h, d=""></h,>
18	hsa-mir-212	yes <h, d=""></h,>	43	hsa-mir-1207	yes <h, d=""></h,>
19	hsa-mir-98	yes <h, d,="" m=""></h,>	44	hsa-mir-33b	yes <h, d=""></h,>
20	hsa-mir-92b	yes <h, d=""></h,>	45	hsa-mir-15b	yes <h, d=""></h,>
21	hsa-mir-455	yes <h, d=""></h,>	46	hsa-mir-630	yes <h, d=""></h,>
22	hsa-mir-590	yes <h, d=""></h,>	47	hsa-mir-622	yes <h, d=""></h,>
23	hsa-mir-330	yes <h, d=""></h,>	48	hsa-mir-1271	yes <h, d=""></h,>
24	hsa-mir-675	yes <h, d=""></h,>	49	hsa-mir-424	yes <h, d=""></h,>
25	hsa-mir-217	yes <h, d=""></h,>	50	hsa-mir-95	yes <h, d=""></h,>

Table 4. Top 50 miRNAs predicted by HLGNN-MDA to be associated with breast cancer.

Note: H <HMDD v3.0>, D <dbDEMC> and M <miR2Disease > represent the databases in which the relations could be validated.

2.5.2. Hepatocellular Carcinoma

Primary liver cancer is the fifth most common cancer worldwide, mainly including hepatocellular carcinoma (HCC) [63,64]. A total of 49 of the top 50 miRNAs predicted to be related to hepatocellular carcinoma could be validated in three validation databases. The results are shown in Table 5. At present, no clear association between hsa-mir-495 and hepatocellular carcinoma could be found in these databases. However, a previous study [65] reported that hsa-mir-495 expression was frequently downregulated in hepatocellular carcinoma tissues and cell lines. Its expression levels were significantly correlated with tumor size, tumor lymph node metastasis (TNM) stage and lymph node metastasis in patients with hepatocellular carcinoma [65].

Rank	MicroRNA	Validation	Rank	MicroRNA	Validation
1	hsa-mir-143	yes <h, d,="" m=""></h,>	26	hsa-mir-23b	yes <h, d,="" m=""></h,>
2	hsa-mir-196b	yes <h, d=""></h,>	27	hsa-mir-574	yes <h, d=""></h,>
3	hsa-mir-137	yes <h, d,="" m=""></h,>	28	hsa-mir-26b	yes <h, d,="" m=""></h,>
4	hsa-mir-520c	yes <h, d=""></h,>	29	hsa-mir-495	no
5	hsa-mir-376c	yes <h, d=""></h,>	30	hsa-mir-328	yes <h, d,="" m=""></h,>
6	hsa-mir-184	yes <h, d=""></h,>	31	hsa-mir-452	yes <h, d=""></h,>
7	hsa-mir-215	yes <h, d,="" m=""></h,>	32	hsa-mir-204	yes <h, d=""></h,>
8	hsa-mir-302a	yes <h, d=""></h,>	33	hsa-mir-135b	yes <h, d=""></h,>
9	hsa-mir-34b	yes <h, d=""></h,>	34	hsa-mir-95	yes <h, d=""></h,>
10	hsa-mir-339	yes <h, d=""></h,>	35	hsa-mir-185	yes <h, d,="" m=""></h,>
11	hsa-mir-708	yes <h, d=""></h,>	36	hsa-mir-206	yes <h, d=""></h,>
12	hsa-mir-193	yes <h, d=""></h,>	37	hsa-mir-449a	yes <h, d=""></h,>
13	hsa-mir-30e	yes <h, d,="" m=""></h,>	38	hsa-mir-520a	yes <h, d=""></h,>
14	hsa-mir-488	yes <h, d=""></h,>	39	hsa-mir-194	yes <h, d,="" m=""></h,>
15	hsa-mir-200	yes <h, m=""></h,>	40	hsa-mir-451	yes <h, d=""></h,>
16	hsa-mir-342	yes <h, d=""></h,>	41	hsa-mir-149	yes <h, d=""></h,>
17	hsa-mir-367	yes <h, d=""></h,>	42	hsa-mir-153	yes <h, d=""></h,>
18	hsa-mir-302d	yes <h, d=""></h,>	43	hsa-mir-299	yes <h, d=""></h,>
19	hsa-mir-494	yes <h, d=""></h,>	44	hsa-mir133a	yes <h, d,="" m=""></h,>
20	hsa-mir-128	yes <h, d,="" m=""></h,>	45	hsa-mir-633	yes <d></d>
21	hsa-mir-340	yes <h, d=""></h,>	46	hsa-mir-132	yes <h, d,="" m=""></h,>
22	hsa-mir-33b	yes <h, d=""></h,>	47	hsa-mir-27b	yes <h, d=""></h,>
23	hsa-mir-625	yes <h, d=""></h,>	48	hsa-mir-935	yes <h, d=""></h,>
24	hsa-mir-424	yes <h, d=""></h,>	49	hsa-mir-32	yes <h, d=""></h,>
25	hsa-mir-151b	yes <h, d=""></h,>	50	hsa-mir-186	yes <h, d,="" m=""></h,>

Table 5. Top 50 miRNAs predicted by HLGNN-MDA to be associated with hepatocellular carcinoma.

Note: H <HMDD v3.0>, D <dbDEMC> and M <miR2Disease > represent the databases in which the relations could be validated.

2.5.3. Renal Cell Carcinoma

Approximately 270,000 kidney cancer cases and 116,000 deaths are diagnosed annually worldwide [66]. Ninety percent of kidney cancers are tumors originating from the kidney epithelium and renal cell carcinoma [67]. The miRNAs predicted to be associated with the top 50 renal cell carcinomas are listed in Table 6. A total of 47 of the top 50 miRNAs could be validated clearly.

Table 6. Top 50 miRNAs predicted by HLGNN-MDA to be associated with renal cell carcinoma.

Rank	MicroRNA	Validation	Rank	MicroRNA	Validation
1	hsa-mir-20a	yes <h, d,="" m=""></h,>	26	hsa-mir-181a	yes <h, d=""></h,>
2	hsa-mir-17	yes <h, d,="" m=""></h,>	27	hsa-mir-192	yes <h, d=""></h,>
3	hsa-mir-27b	yes <h, d=""></h,>	28	hsa-mir-22	yes <h, d=""></h,>
4	hsa-mir-221	yes <h, d,="" m=""></h,>	29	hsa-mir-182	yes <h, d,="" m=""></h,>
5	hsa-mir-223	yes <h, d,="" m=""></h,>	30	hsa-mir-29b	yes <h, d,="" m=""></h,>

Rank	MicroRNA	Validation	Rank	MicroRNA	Validation
6	hsa-mir-31	yes <h, d=""></h,>	31	hsa-mir-15a	yes <h, d,="" m=""></h,>
7	hsa-mir-29a	yes <h, d,="" m=""></h,>	32	hsa-mir-375	yes <h, d=""></h,>
8	hsa-mir-125b	yes <h, d=""></h,>	33	hsa-mir-486	yes <d></d>
9	hsa-mir-133a	yes <h, d,="" m=""></h,>	34	hsa-mir-15b	yes <h, d=""></h,>
10	hsa-mir-125a	yes <h, d=""></h,>	35	hsa-mir-107	yes <h, d=""></h,>
11	hsa-mir-18a	yes <h, d=""></h,>	36	hsa-mir-328	yes <d></d>
12	hsa-mir-1	yes <h, d=""></h,>	37	hsa-mir-23a	yes <d></d>
13	hsa-mir-30a	yes <h, d,="" m=""></h,>	38	hsa-mir-194	yes <h, d=""></h,>
14	hsa-mir-181b	yes <h, d=""></h,>	39	hsa-mir-193b	yes <h, d=""></h,>
15	hsa-mir-19b	yes <h, d,="" m=""></h,>	40	hsa-mir-196b	yes <d></d>
16	hsa-mir-214	yes <h, d,="" m=""></h,>	41	hsa-mir-137	yes <h, d=""></h,>
17	hsa-mir-130a	yes <h, d=""></h,>	42	hsa-mir-191	yes <h, d,="" m=""></h,>
18	hsa-mir-222	yes <h, d=""></h,>	43	hsa-mir-302a	no
19	hsa-mir-148a	yes <h, d=""></h,>	44	hsa-mir-135b	yes <d></d>
20	hsa-mir-25	yes <d></d>	45	hsa-mir-451b	no
21	hsa-mir-133b	yes <h, d=""></h,>	46	hsa-mir-342	yes <d, m=""></d,>
22	hsa-mir-183	yes <h, d=""></h,>	47	hsa-mir-30b	yes <h, d=""></h,>
23	hsa-mir-106a	yes <h, d,="" m=""></h,>	48	hsa-mir-373	no
24	hsa-mir-24	yes <d></d>	49	hsa-mir-212	yes <d></d>
25	hsa-mir-132	ves <d></d>	50	hsa-mir-193a	ves <h, d=""></h,>

Table 6. Cont.

Note: H <HMDD v3.0>, D <dbDEMC> and M <miR2Disease > represent the databases in which the relations could be validated.

yes <D>

3. Materials and Methods

3.1. Data Resources

We collected human miRNA-disease associations from the HMDD v2.0 database [68] and obtained 5430 miRNA-disease associations between 495 miRNAs and 383 diseases. Therefore, many miRNA-disease associations were organized into an adjacency matrix $Y \in N^{nm \times nd}$, where nm and nd represent the number of miRNAs and the number of diseases, respectively. If an association between miRNA m_i and disease d_i was recorded in HMDD v2.0, then Y(i, j) equaled 1; otherwise, it equaled 0.

3.2. Methods

3.2.1. Extraction of the Enclosing Subgraph of Node Pair

The design of our HLGNN-MDA model is inspired to the SEAL (learning from subgraphs, embedding and attributes for link prediction) [69] framework, which uses graph neural networks for link prediction. SEAL proves that most high-order heuristics can be approximated by learning from local enclosing subgraphs. Therefore, the first step of HLGNN-MDA is to extract the enclosing subgraph of all pairs of nodes. The enclosing subgraph is composed of two nodes, which are marked as central nodes, and their surrounding nodes. Then, the *h*-hop enclosing subgraph of central nodes *m* and *d* consists of all nodes whose distance from node *m* or node *d* is less than *h* steps.

HLGNN-MDA inputs the enclosing subgraphs into an "end-to-end" graph neural network for training. The association information of the central node pair of the enclosing subgraph is used as the supervisory label of the graph neural network. In association matrix Y, miRNA–disease associations equal to 1 are considered known, while those equal to 0 are treated as potential. The training set extracts known associations as positive samples and randomly samples an equal number of potential associations as negative samples. Subsequently, the enclosing subgraphs of all samples in the training set are extracted. To prevent the leakage of supervised labels, the edges between the central node pairs and the positive samples is removed during the extraction of the enclosing subgraph. The enclosing subgraph is denoted with *A*.

3.2.2. Label Nodes

In this section, we label all the nodes in each enclosing subgraph. Node labeling is the process of assigning an integer to a node in an enclosing subgraph and can be defined as $f_l : V \rightarrow N$. Node labeling uses the DRNL (double-radius node labeling) method proposed in SEAL. First, we label the central two nodes with label 1. Other nodes with the same distance from the central two nodes are labeled with the same value. The farther the distance is, the greater the value is. The label of a node *i* can be derived from the following hash function:

$$f_l(i) = 1 + \min(d_x, d_y) + (d/2)[(d/2) + (d\%2) - 1]$$
(1)

where d_x and d_y represent the distances from node *i* to central nodes *x* and *y*, respectively, with $d = d_x + d_y$; and d/2 and d%2 indicate division and remainder, respectively. When a node is disconnected from the central nodes, it is labeled with 0.

In this way, we can label all the nodes in the enclosing subgraphs. Then, before being input into the graph neural network, the label of each node is expanded into one-hot encoding as its feature. This one-hot feature represents the position information of the node in the enclosing subgraph.

3.2.3. Construct Graph Neural Network

After extracting the enclosing subgraphs and labeling nodes in each enclosing subgraph, we can input the labeled enclosing subgraphs into the graph neural network for predicting miRNA–disease associations. At present, most of the proposed graph neural networks are applicable to homogeneous networks, whereas the miRNA–disease associations we use are in a bipartite graph network. Therefore, we improve the graph neural network DGCNN (deep graph convolutional neural network) to obtain better performance in heterogeneous networks.

Graph convolution layers. The role of the graph convolution layer is to learn the node representations [70,71]. A graph is input into the graph convolution layer through multilayer convolution, and the vector representation of each node can be extracted. The vector contains local substructural features of the graph. The process of graph convolution is the aggregation of feature information from the neighbors around each node.

Given the adjacency matrix of a graph $A \in \mathbb{R}^{n \times n}$, its information matrix is $X \in \mathbb{R}^{n \times d}$, where *n* indicates the number of nodes and *d* represents the dimension of the features. Graph convolution can be represented as follows:

$$Z = \sigma(f(A) \cdot X \cdot W) \tag{2}$$

where $W \in \mathbb{R}^{d \times c}$ indicates the parameters to be trained, which converts the d-dimensional signal into c-dimensional signals; and f(A) represents the propagation function of adjacency matrix A. Usually, $f(A) = \widetilde{D}^{-1}\widetilde{A}$ or $f(A) = \widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}$, where $\widetilde{A} = A + I$ and \widetilde{D} are the degree matrices of \widetilde{A} . Moreover, $f(A) \cdot X \cdot W$ indicates that the feature vector of each node is aggregated in the manner of a propagation function and then undergoes information conversion. In addition, $\sigma(\cdot)$ is the activation function.

In the bipartite undirected network, miRNAs are only connected with diseases. That is, after a two-step jump, one disease node can only reach another disease, which is also true for miRNA nodes. Therefore, we use a propagation function for second-order topological information that allows information between homogeneous nodes to be aggregated together directly, i.e., defining a propagation function $f(A) = A^2$.

With different propagation functions selected, the topological characteristics of the graph learned by the graph convolutional network are slightly different [72]. By splicing the neighbor information of nodes aggregated by different propagation functions, the graph convolutional layer can learn better graph topological features. The graph convolutional layer of our HLGNN-MDA model is shown in Figure 3 and can be represented by Formula (3):

$$Z^{t+1} = \sigma\left(\left[A \cdot Z^t, A^2 \cdot Z^t, D^{-1}A \cdot Z^t, D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \cdot Z^t\right] \cdot W^t\right)$$
(3)

where $Z^0 = X$ and $Z^t \in \mathbb{R}^{n \times d_t}$ represent the output of the *t*-th graph convolutional layer; d_t is the output dimension of the *t*-th layer graph convolution; and $[\cdot]$ represents the splicing of row vectors, which splices the node vectors obtained through different propagation functions. The propagation functions used in our model are A, A^2 , $D^{-1}A$ and $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. After splicing, the topological features captured by these several propagation functions can be considered at the same time. $W^{4d_t \times d_{t+1}}$ is the parameter to be trained, which maps the spliced node features from the $4d_t$ dimension to the d_{t+1} dimension. Dimension $4d_t$ is used here because there are four propagation functions involved in the graph convolution process.

Enclosing subgraph *A* and its node features *X* go through *t* graph convolution layers to produce output Z^t , t = 1, ..., T. The overall graph convolution layer uses a global structure where the results of each layer are stitched together at the end, resulting in a result denoted as $Z = [Z^1, ..., Z^T]$. Each row of $Z \in \mathbb{R}^{n \times \sum_{i=1}^{T} d_i}$ is the $\sum_{i=1}^{T} d_i$ -dimensional embedding vector representation of a node that contains rich topological information about that node in that graph. Figure 2 shows the overall architecture of the graph convolutional layer.

Graph pooling layers. A graph convolutional layer is used to learn a latent vector representation for each node. Here, the graph pooling layer of HLGNN-MDA selects the *k* most important nodes from the nodes of the graph to represent the graph. The importance of the node is evaluated using the result of the graph convolutional layer. The last layer of graph convolution maps the result of the previous layer to 1 dimension. That is, the parameter, W^T , of the last layer of graph convolution maps dimension $4d_{T-1}$ to 1 dimension. In this way, a value is obtained for each node, and this value indicates how important the node is in the graph.

We sort *Z* in descending order according to its last dimension. If two nodes appear to be equal in the last dimension of *Z*, we compare their penultimate dimension and so on, until the two nodes can be separated. The graph pooling layer takes the top *k* nodes in the ranking as its output, which helps subsequent conventional neural network layers to obtain a tensor with a fixed specification. When the number of nodes, *n*, is less than *k*, (n - k) zero vectors are added after the sorted nodes.

Convolution layers and fully connected layer. After the graph pooling layer, a tensor $Z = k \times \sum_{1}^{T} d_t$ is obtained. In the remaining layers, we first use the traditional onedimensional convolutional neural network combined with the max pooling layer to further refine the graph representation features and then make the final prediction using a fully connected layer.

To train a one-dimensional convolutional neural network on tensor *Z*, it first needs to be reshaped into a one-dimensional vector. To apply the filter for each node feature, the filter size and step length of the first one-dimensional convolutional neural network are set to $\sum_{1}^{T} d_t$; that is, each node feature is convolved first. Then, after a max pooling layer, a one-dimensional convolutional neural layer is used to further learn the graph to represent the local features in the sequence features. Finally, it is connected to the fully connected layer. We use NLLLoss as the loss function, which adds up the predicted values of all predicted samples under the true labels. The predicted value here is a negative number in the logarithmic form of a normalized exponential function (softmax), mapping the prediction range from (0, 1) to $(0, +\infty)$. If all prediction samples are predicted correctly,

NLLLoss is closer to 0. Finally, the fully connected layer outputs the probability of miRNA and disease node pair connections.

3.2.4. Evaluation Metrics

In order to verify the model performance, we choose the following metrics: AUROC (Receiver Operating Characteristic curve), ACC (accuracy), precision, recall, AUPR (area under the precision–recall curve) and MCC (Matthews correlation coefficient). The relevant definitions are as follows:

$$Precision = \frac{TP}{TP + FP}$$
(4)

$$\operatorname{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{TN}}$$
(5)

$$FPR = \frac{FP}{FP + TN}$$
(6)

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$
(7)
$$TR = TN \times FR \times FN$$

$$MCC = \frac{\Pi - \Pi \times FT \times FI}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$
(8)

TP (true positive), FP (false positive), TN (true negative) and FN (false negative) were all derived from the confusion matrix. The ROC curve has the FPR (false positive rate) and TPR (recall or true positive rate) as the horizontal and vertical coordinates, respectively, and the area under the ROC curve is the AUC value. The area under the curve with recall and precision as the horizontal and vertical coordinates is the AUPR value.

4. Conclusions

Understanding the relationship between miRNAs and disease has important implications for disease prevention, detection and treatment. This paper proposes the HLGNN-MDA method, which is a heuristic for learning miRNA-disease association prediction from known miRNA-disease relationships based on graph neural networks. HLGNN-MDA first extracts the enclosing subgraphs around each miRNA-disease pair to be predicted to obtain the local network structure. Each node in the enclosing subgraphs is then labeled. The labeled subgraphs are then input into the graph neural network for classification. In particular, second-order topological information is added to the convolutional layer of the graph neural network to enable it to learn information between similar nodes. Second, different combinations of propagation functions are designed to improve the accuracy and stability of the graph neural network. We compared the model with the same type of miRNA–disease association prediction model using tenfold cross-validation. The results showed that HLGNN-MDA was able to obtain better performance than most miRNAdisease association prediction models. After discussing the effect of hop count on extracting closed subgraphs, we successively evaluated each propagation function combined in the model. Finally, we used the trained HLGNN-MDA model to make predictions and performed case studies on breast cancer, hepatocellular carcinoma and renal cell carcinoma. A total of 49 of the top 50 predicted miRNAs for breast cancer could be found in the validation database. The remaining hsa-mir-362 was also found to be associated with breast cancer, as supported by the literature. Similarly, 49 of the top 50 miRNAs predicted for hepatocellular carcinoma could be validated against the database. The remaining hsa-mir-495 was also found to be related to hepatocellular carcinoma, as supported by the literature. Finally, 47 of the top 50 miRNAs associated with renal cell carcinoma could be validated.

Generally, HLGNN-MDA has the following advantages: First, HLGNN-MDA can select arbitrary links for prediction without having to predict all potential miRNA–disease associations in the adjacency matrix. In particular, when predicting individual diseases or miRNAs, HLGNN-MDA can directly obtain the corresponding results. Second, HLGNN-

MDA does not strictly require the corresponding similarity data because it can learn information through the topology of the network.

However, there are still some aspects for improvement. Valid miRNA and disease signatures are also important for prediction. Therefore, adding valid miRNA and disease signatures to HLGNN-MDA should be further investigated. Second, HLGNN-MDA is an end-to-end supervised learning algorithm framework. One of the problems with miRNA–disease association prediction is that there is no definite negative sample. For the potential associations, only a small proportion of them are truly associated, and most are unassociated. In this paper, the prediction of miRNA–disease association is approximated as a supervised learning model [73] with insufficient samples. Therefore, a new linkage prediction heuristic represents a future researchable direction.

Author Contributions: Conceptualization, L.Y. and B.J.; methodology, B.J.; software, B.J.; validation, B.J. and S.R.; formal analysis, B.J.; investigation, L.Y.; resources, L.Y.; data curation, B.J.; writing—original draft preparation, B.J.; writing—review and editing, S.R.; visualization, B.J. and S.R.; supervision, L.Y.; project administration, L.Y.; funding acquisition, L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Natural Science Foundation of China (grants Nos. 62072353 and 62132015).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Code of the model, datasets and results can be downloaded from GitHub (https://github.com/LiangYu-Xidian/HLGNN-MDA).

Acknowledgments: Thanks to all those who maintain excellent databases and to all experimentalists who enabled this work by making their data publicly available.

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

References

- 1. Lee, R.C.; Feinbaum, R.L.; Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **1993**, *75*, 843–854. [CrossRef]
- Perez-Rodriguez, D.; Lopez-Fernandez, H.; Agis-Balboa, R.C. Application of miRNA-seq in neuropsychiatry: A method-ological perspective. *Comput. Biol. Med.* 2021, 135, 104603. [CrossRef]
- Cui, F.; Zhou, M.; Zou, Q. Computational biology and chemistry Special section editorial: Computational analyses for miRNA. Comput. Biol. Chem. 2021, 91, 107448. [CrossRef]
- Shaker, F.; Nikravesh, A.; Arezumand, R.; Aghaee-Bakhtiari, S.H. Web-based tools for miRNA studies analysis. *Comput. Biol. Med.* 2020, 127, 104060. [CrossRef]
- Wang, G.; Wang, Y.; Teng, M.; Zhang, D.; Li, L.; Liu, Y. Signal transducers and activators of transcription-1 (STAT1) regulates microRNA transcription in interferon gamma-stimulated HeLa cells. *PLoS ONE* 2010, 5, e11794.
- Zhao, Y.; Wang, F.; Juan, L. MicroRNA Promoter Identification in *Arabidopsis* using Multiple Histone Markers. *BioMed Res. Int.* 2015, 2015, 861402. [CrossRef]
- Zhao, Y.; Wang, F.; Chen, S.; Wan, J.; Wang, G. Methods of MicroRNA Promoter Prediction and Transcription Factor Mediated Regulatory Network. *BioMed Res. Int.* 2017, 2017, 7049406. [CrossRef]
- Wang, X.; Yang, Y.; Liu, J.; Wang, G. The stacking strategy-based hybrid framework for identifying non-coding RNAs. Brief. Bioinform. 2021, 22, bbab023. [CrossRef]
- Tétreault, N.; De Guire, V. miRNAs: Their discovery, biogenesis and mechanism of action. *Clin. Biochem.* 2013, 46, 842–845. [CrossRef]
- 10. Tian, L.; Wang, S.L. Exploring miRNA Sponge Networks of Breast Cancer by Combining miRNA-disease-lncRNA and miRNA-target Networks. *Curr. Bioinform.* **2021**, *16*, 385–394. [CrossRef]
- 11. Han, W.; Lu, D.; Wang, C.; Cui, M.; Lu, K. Identification of Key mRNAs, miRNAs, and mRNA-miRNA Network Involved in Papillary Thyroid Carcinoma. *Curr. Bioinform.* **2021**, *16*, 146–153. [CrossRef]
- Sarkar, J.P.; Saha, I.; Sarkar, A.; Maulik, U. Machine learning integrated ensemble of feature selection methods followed by survival analysis for predicting breast cancer subtype specific miRNA biomarkers. *Comput. Biol. Med.* 2021, 131, 104244. [CrossRef] [PubMed]
- 13. Liao, Z.; Li, D.; Wang, X.; Li, L.; Zou, Q. Cancer Diagnosis Through IsomiR Expression with Machine Learning Method. *Curr. Bioinform.* **2018**, *13*, 57–63. [CrossRef]

- Calin, G.A.; Dumitru, C.D.; Shimizu, M.; Bichi, R.; Zupo, S.; Noch, E.; Aldler, H.; Rattan, S.; Keating, M.; Rai, K.; et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. USA* 2002, 99, 15524–15529. [CrossRef] [PubMed]
- Lawrie, C.H.; Gal, S.; Dunlop, H.M.; Pushkaran, B.; Liggins, A.P.; Pulford, K.; Banham, A.H.; Pezzella, F.; Boultwood, J.; Wainscoat, J.S.; et al. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *Br. J. Haematol.* 2008, 141, 672–675. [CrossRef] [PubMed]
- 16. Reddy, K.B. MicroRNA (miRNA) in cancer. Cancer Cell Int. 2015, 15, 38. [CrossRef] [PubMed]
- 17. Khan, A.; Zahra, A.; Mumtaz, S.; Fatmi, M.Q.; Khan, M.J. Integrated In-silico Analysis to Study the Role of microRNAs in the Detection of Chronic Kidney Diseases. *Curr. Bioinform.* **2020**, *15*, 144–154. [CrossRef]
- 18. Porta, C.; Figlin, R.A. MiR-193a-3p and miR-224 mediate renal cell carcinoma progression by targeting al-pha-2,3-sialyltransferase IV and the phosphatidylinositol 3 kinase/Akt pathway. *Mol. Carcinog.* **2019**, *58*, 1926–1927.
- Zhao, Z.; Zhang, C.; Li, M.; Yu, X.; Liu, H.; Chen, Q.; Wang, J.; Shen, S.; Jiang, J. Integrative Analysis of miRNA-mediated Competing Endogenous RNA Network Reveals the lncRNAs-mRNAs Interaction in Glioblastoma Stem Cell Differentiation. *Curr. Bioinform.* 2020, 15, 1187–1196. [CrossRef]
- Zhu, Q.; Fan, Y.; Pan, X. Fusing Multiple Biological Networks to Effectively Predict miRNA-disease Associations. *Curr. Bioinform.* 2021, 16, 371–384. [CrossRef]
- Li, J.Q.; Rong, Z.H.; Chen, X.; Yan, G.Y.; You, Z.H. MCMDA: Matrix completion for MiRNA-disease association prediction. Oncotarget 2017, 8, 21187–21199. [CrossRef] [PubMed]
- 22. Peng, L.; Peng, M.; Liao, B.; Huang, G.; Liang, W.; Li, K. Improved low-rank matrix recovery method for predicting miRNA-disease association. *Sci. Rep.* 2017, 7, 6007. [CrossRef] [PubMed]
- 23. Ha, J. MDMF: Predicting miRNA-Disease Association Based on Matrix Factorization with Disease Similarity Constraint. J. Pers. Med. 2022, 12, 885. [CrossRef]
- 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] [PubMed]
- Ha, J.; Park, C.; Park, S. IMIPMF: Inferring miRNA-disease interactions using probabilistic matrix factorization. J. Biomed. Inform. 2020, 102, 103358. [CrossRef] [PubMed]
- 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]
- Ha, J.; Park, C. MLMD, MLMD: Metric Learning for predicting miRNA-Disease associations. *IEEE Access* 2021, *9*, 78847–78858. [CrossRef]
- Chen, X.; Li, T.H.; Zhao, Y.; Wang, C.C.; Zhu, C.C. Deep-belief network for predicting potential miRNA-disease associations. Brief. Bioinform. 2021, 22, bbaa186.
- 29. Zhang, Z.-Y.; Sun, Z.-J.; Yang, Y.-H.; Lin, H. Towards a better prediction of subcellular location of long non-coding RNA. *Front. Comput. Sci.* **2022**, *16*, 165903. [CrossRef]
- Yang, H.; Luo, Y.; Ren, X.; Wu, M.; He, X.; Peng, B.; Deng, K.; Yan, D.; Tang, H.; Lin, H. Risk Prediction of Diabetes: Big data mining with fusion of multi-farious physical examination indicators. *Inf. Fusion* 2021, 75, 140–149. [CrossRef]
- Lu, X.; Gao, Y.; Zhu, Z.; Ding, L.; Wang, X.; Liu, F.; Li, J. A Constrained Probabilistic Matrix Decomposition Method for Predicting miRNA-disease Associations. *Curr. Bioinform.* 2021, 16, 524–533. [CrossRef]
- 32. Liu, Z.P. Predicting lncRNA-protein Interactions by Machine Learning Methods: A Review. *Curr. Bioinform.* **2020**, *15*, 831–840. [CrossRef]
- 33. Zhang, Y.; Duan, G.; Yan, C.; Yi, H.; Wu, F.X.; Wang, J. MDAPlatform: A Component-based Platform for Constructing and Assessing miRNA-disease Association Prediction Methods. *Curr. Bioinform.* **2021**, *16*, 710–721. [CrossRef]
- Zhang, X.; Zou, Q.; Rodriguez-Paton, A.; Zeng, X. Meta-Path Methods for Prioritizing Candidate Disease miRNAs. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2019, 16, 283–291. [CrossRef] [PubMed]
- 35. Zeng, X.; Liu, L.; Lü, L.; Zou, Q. Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics* 2018, 34, 2425–2432. [CrossRef] [PubMed]
- 36. Dai, Q.; Chu, Y.; Li, Z.; Zhao, Y.; Mao, X.; Wang, Y.; Xiong, Y.; Wei, D.Q. MDA-CF: Predicting MiRNA-Disease associations based on a cascade forest model by fusing multi-source information. *Comput. Biol. Med.* **2021**, *136*, 104706. [CrossRef]
- 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]
- Chen, X.; Clarence Yan, C.; Zhang, X.; Li, Z.; Deng, L.; Zhang, Y.; Dai, Q. RBMMMDA: Predicting multiple types of diseasemicroRNA associations. *Sci. Rep.* 2015, *5*, 13877. [CrossRef]
- Phan, A.V.; Le Nguyen, M.; Nguyen, Y.L.; Bui, L.T. Dgcnn: A convolutional neural network over large-scale labeled graphs. *Neural. Netw.* 2018, 108, 533–543. [CrossRef]
- 40. Zhang, Y.; Yan, J.; Chen, S.; Gong, M.; Gao, D.; Zhu, M.; Gan, W. Review of the Applications of Deep Learning in Bioinformatics. *Curr. Bioinform.* **2021**, *15*, 898–911. [CrossRef]
- 41. Ayachit, G.; Shaikh, I.; Pandya, H.; Das, J. Salient Features, Data and Algorithms for MicroRNA Screening from Plants: A Review on the Gains and Pitfalls of Machine Learning Techniques. *Curr. Bioinform.* **2021**, *15*, 1091–1103. [CrossRef]

- 42. Chen, L.; Zhou, J.-P. Identification of Carcinogenic Chemicals with Network Embedding and Deep Learning Methods. *Curr. Bioinform.* **2021**, *15*, 1017–1026. [CrossRef]
- 43. Wang, D.; Zhang, Z.; Jiang, Y.; Mao, Z.; Wang, D.; Lin, H.; Xu, D. DM3Loc: Multi-label mRNA subcellular localization prediction and analysis based on multi-head self-attention mechanism. *Nucleic Acids Res.* **2021**, *49*, e46. [CrossRef] [PubMed]
- Lv, H.; Dao, F.Y.; Zulfiqar, H.; Lin, H. DeepIPs: Comprehensive assessment and computational identification of phosphorylation sites of SARS-CoV-2 infection using a deep learning-based approach. *Brief Bioinform.* 2021, 22, bbab244. [CrossRef] [PubMed]
- 45. Lv, H.; Dao, F.-Y.; Zulfiqar, H.; Su, W.; Ding, H.; Liu, L.; Lin, H. A sequence-based deep learning approach to predict CTCFmediated chromatin loop. *Brief. Bioinform.* **2021**, *22*, bbab031. [CrossRef]
- 46. Fu, L.; Peng, Q. A deep ensemble model to predict miRNA-disease association. Sci. Rep. 2017, 7, 14482. [CrossRef]
- Geete, K.; Pandey, M. Robust Transcription Factor Binding Site Prediction Using Deep Neural Networks. *Curr. Bioinform.* 2020, 15, 1137–1152. [CrossRef]
- Peng, J.; Hui, W.; Li, Q.; Chen, B.; Hao, J.; Jiang, Q.; Shang, X.; Wei, Z. A learning-based framework for miRNA-disease association identification using neural networks. *Bioinformatics* 2019, 35, 4364–4371. [CrossRef]
- Wu, B.; Zhang, H.; Lin, L.; Wang, H.; Gao, Y.; Zhao, L.; Chen, Y.-P.P.; Chen, R.; Gu, L. A Similarity Searching System for Biological Phenotype Images Using Deep Convolutional Encoder-decoder Architecture. *Curr. Bioinform.* 2019, 14, 628–639. [CrossRef]
- Chu, Y.; Wang, X.; Dai, Q.; Wang, Y.; Wang, Q.; Peng, S.; Wei, X.; Qiu, J.; Salahub, D.R.; Xiong, Y.; et al. MDA-GCNFTG: Identifying miRNA-disease associations based on graph convolutional networks via graph sampling through the feature and topology graph. *Brief. Bioinform.* 2021, 22, bbab165. [CrossRef]
- Chen, X.; Cheng, J.-Y.; Yin, J. Predicting microRNA-disease associations using bipartite local models and hubness-aware regression. RNA Biol. 2018, 15, 1192–1205. [CrossRef] [PubMed]
- 52. 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] [PubMed]
- Che, K.; Guo, M.; Wang, C.; Liu, X.; Chen, X. Predicting MiRNA-Disease Association by Latent Feature Extraction with Positive Samples. *Genes* 2019, 10, 80. [CrossRef]
- Chen, X.; Niu, Y.W.; Wang, G.H.; Yan, G.Y. MKRMDA: Multiple kernel learning-based Kronecker regularized least squares for MiRNA-disease association prediction. J. Transl. Med. 2017, 15, 251. [CrossRef]
- 55. Zhang, M.; Cui, Z.; Neumann, M.; Chen, Y. An End-to-End Deep Learning Architecture for Graph Classification. *Proc. Conf. AAAI Artif. Intell.* **2018**, *32*, 4438–4445. [CrossRef]
- 56. Du, X.; Yao, Y. ConvsPPIS: Identifying Protein-protein Interaction Sites by an Ensemble Convolutional Neural Network with Feature Graph. *Curr. Bioinform.* **2020**, *15*, 368–378. [CrossRef]
- 57. Huang, Z.; Shi, J.; Gao, Y.; Cui, C.; Zhang, S.; Li, J.; Zhou, Y.; Cui, Q. HMDD v3.0: A database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.* **2019**, *47*, D1013–D1017. [CrossRef]
- 58. 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.* 2016, 45, D812–D818. [CrossRef]
- Jiang, Q.; Wang, Y.; Hao, Y.; Juan, L.; Teng, M.; Zhang, X.; Li, M.; Wang, G.; Liu, Y. miR2Disease: A manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* 2009, *37*, D98–D104. [CrossRef]
- 60. Tao, Z.; Shi, A.; Lu, C.; Song, T.; Zhang, Z.; Zhao, J. Breast Cancer: Epidemiology and Etiology. *Cell Biophys.* **2014**, *72*, 333–338. [CrossRef]
- 61. Assiri, A.A.; Mourad, N.; Shao, M.; Kiel, P.; Liu, W.; Skaar, T.C.; Overholser, B.R. MicroRNA 362-3p Reduces hERG-related Current and Inhibits Breast Cancer Cells Proliferation. *Cancer Genom. Proteom.* **2019**, *16*, 433–442. [CrossRef] [PubMed]
- 62. NI, F.; Gui, Z.; Guo, Q.; Hu, Z.; Wang, X.; Chen, D.; Wang, S. Downregulation of miR-362-5p inhibits proliferation, migration and invasion of human breast cancer MCF7 cells. *Oncol. Lett.* **2015**, *11*, 1155–1160. [CrossRef] [PubMed]
- El-Serag, B.H.; Rudolph, L. Hepatocellular carcinoma: Epidemiology and molecular carcinogenesis. *Gastroenterology* 2007, 132, 2557–2576. [CrossRef] [PubMed]
- 64. Jeyaram, C.; Philip, M.; Perumal, R.C.; Benny, J.; Jayakumari, J.M.; Ramasamy, M.S. A Computational Approach to Identify Novel Potential Precursor miRNAs and their Targets from Hepatocellular Carcinoma Cells. *Curr. Bioinform.* **2019**, *14*, 24–32. [CrossRef]
- Ye, Y.; Zhuang, J.; Wang, G.; He, S.; Zhang, S.; Wang, G.; Ni, J.; Wang, J.; Xia, W. MicroRNA-495 suppresses cell proliferation and invasion of hepatocellular carcinoma by directly targeting insulin-like growth factor receptor-1. *Exp. Ther. Med.* 2018, 15, 1150–1158.
- 66. Ljungberg, B.; Campbell, S.C.; Cho, H.Y.; Jacqmin, D.; Lee, J.E.; Weikert, S.; Kiemeney, L.A. The Epidemiology of Renal Cell Carcinoma. *Eur. Urol.* **2011**, *60*, 1317. [CrossRef]
- 67. Hsieh, J.J.; Purdue, M.P.; Signoretti, S.; Swanton, C.; Albiges, L.; Schmidinger, M.; Heng, D.Y.; Larkin, J.; Ficarra, V. Renal cell carcinoma. *Nat. Rev. Dis. Prim.* 2017, *3*, 17009. [CrossRef]
- 68. Li, Y.; Qiu, C.; Tu, J.; Geng, B.; Yang, J.; Jiang, T.; Cui, Q. HMDD v2.0: A database for experimentally supported human mi-croRNA and disease associations. *Nucleic Acids Res.* 2014, 42, D1070–D1074. [CrossRef]
- Zhang, M.; Chen, Y. Link prediction based on graph neural networks. In Advances in Neural Information Processing Systems 31, Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montreal, QC, Canada, 3–8 December 2018; Neural Information Processing Systems Foundation, Inc.: San Diego, CA, USA, 2018.
- 70. Thomas, N.; Kipf, M.W. Semi-Supervised Classification with Graph Convolutional Networks. arXiv 2016, arXiv:1609.02907.

- Hamilton, W.; Ying, Z.; Leskovec, J. Inductive Representation Learning on Large Graphs. In Advances in Neural Information Processing Systems 30, Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Neural Information Processing Systems Foundation, Inc.: San Diego, CA, USA, 2017; pp. 1025–1035.
- 72. Dehmamy, N.; Barabási, A.L.; Yu, R. Understanding the Representation Power of Graph Neural Networks in Learning Graph Topology. In Advances in Neural Information Processing Systems 32, Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019; Neural Information Processing Systems Foundation, Inc.: San Diego, CA, USA, 2019.
- 73. Lan, Y.; Li, Q. Supervised Learning in Spiking Neural Networks with Synaptic Delay Plasticity: An Overview. *Curr. Bioinform.* **2020**, *15*, 854–865. [CrossRef]