



Article

CNNDLP: A Method Based on Convolutional Autoencoder and Convolutional Neural Network with Adjacent Edge Attention for Predicting lncRNA–Disease Associations

Ping Xuan ¹, Nan Sheng ¹, Tiangang Zhang ², Yong Liu ¹ and Yahong Guo ^{3,*}

¹ School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China

² School of Mathematical Science, Heilongjiang University, Harbin 150080, China

³ School of Information Science and Technology, Heilongjiang University, Harbin 150080, China

* Correspondence: guoyahong_hlju@163.com; Tel.: +86-188-4618-9060

Received: 27 June 2019; Accepted: 28 August 2019; Published: 30 August 2019



Abstract: It is well known that the unusual expression of long non-coding RNAs (lncRNAs) is closely related to the physiological and pathological processes of diseases. Therefore, inferring the potential lncRNA–disease associations are helpful for understanding the molecular pathogenesis of diseases. Most previous methods have concentrated on the construction of shallow learning models in order to predict lncRNA–disease associations, while they have failed to deeply integrate heterogeneous multi-source data and to learn the low-dimensional feature representations from these data. We propose a method based on the convolutional neural network with the attention mechanism and convolutional autoencoder for predicting candidate disease-related lncRNAs, and refer to it as CNNDLP. CNNDLP integrates multiple kinds of data from heterogeneous sources, including the associations, interactions, and similarities related to the lncRNAs, diseases, and miRNAs. Two different embedding layers are established by combining the diverse biological premises about the cases that the lncRNAs are likely to associate with the diseases. We construct a novel prediction model based on the convolutional neural network with attention mechanism and convolutional autoencoder to learn the attention and the low-dimensional network representations of the lncRNA–disease pairs from the embedding layers. The different adjacent edges among the lncRNA, miRNA, and disease nodes have different contributions for association prediction. Hence, an attention mechanism at the adjacent edge level is established, and the left side of the model learns the attention representation of a pair of lncRNA and disease. A new type of lncRNA similarity and a new type of disease similarity are calculated by incorporating the topological structures of multiple bipartite networks. The low-dimensional network representation of the lncRNA–disease pairs is further learned by the autoencoder based convolutional neural network on the right side of the model. The cross-validation experimental results confirm that CNNDLP has superior prediction performance compared to the state-of-the-art methods. Case studies on stomach cancer, breast cancer, and prostate cancer further show the ability of CNNDLP for discovering the potential disease lncRNAs.

Keywords: lncRNA–disease association prediction; feature learning based on convolutional autoencoder; convolutional neural networks; attention at adjacent edge level; similarity calculation based on multiple bipartite networks

1. Introduction

For the past few years, genetic information has been thought to be stored only in protein-coding genes, while non-coding RNAs (ncRNAs) are only byproducts of the transcription process [1,2].

However, accumulating evidences indicate that ncRNAs play important roles in various biological processes, especially long non-coding RNAs (lncRNAs), with lengths > 200 nucleotides [3,4].

The previous methods have been presented for predicting the lncRNA-disease associations, and they are classified into three categories. The methods in the first category utilize machine learning methods to identify the candidate associations. Chen et al. develop a semi-supervised learning model, LRLSLDA, which uses Laplacian regularized least squares to identify possible associations between lncRNA and disease [5]. A model based on the Bayesian classifier was developed for predicting candidate disease lncRNAs [6]. However, most of the methods in this category fail to achieve the good performances for the lncRNAs with no any known associated diseases.

The second category of methods takes use of the biological premise that lncRNAs with similar functions tend to be associated with similar diseases [7]. First, the similarity between two lncRNAs is calculated by the diseases associated with the lncRNAs, and a network composed of lncRNA is constructed by using the similarities between lncRNAs [8]. Several methods are presented for predicting the lncRNAs related to a given disease based on the lncRNA network, for instance, via random walks on the lncRNA network [9,10] or by utilizing the information of neighboring nodes of lncRNA [11]. These methods are ineffective for the new diseases with no known related lncRNAs, as they rely on a set of seed lncRNAs that have been observed to be related to the disease. Some methods attempt to introduce additional information about diseases to solve this shortcoming. Disease information is incorporated with the lncRNA network to create a heterogeneous lncRNA-disease network that contains information of lncRNA similarities, that of disease similarities and that of lncRNA-disease associations. Several methods exploit the information, but they construct different models within the heterogeneous network to estimate the association scores between the lncRNAs and the diseases. For instance, the association scores are derived by random walks in the lncRNA-disease network [10,12], or by matrix factorization of lncRNA-disease associations [13,14]. Since lncRNAs are often involved in disease processes along with miRNAs, it is necessary to integrate the interactions and associations about the miRNAs. Nevertheless, most of the previous methods overlook these information related to the miRNAs.

The third category of methods integrates multiple biological data sources about the lncRNA, the miRNA, the proteins. MFLDA integrates various information about the genes and the miRNAs interacted with lncRNAs, and about the diseases associated with lncRNAs. The method constructs a matrix factorization based on data fusion model for predicting disease lncRNAs [15]. Zhang et al. introduce the protein information to establish the lncRNA-protein-disease network and predict the candidate associations between lncRNAs and diseases based on propagating information streams in the network [16]. The diverse information available about the lncRNAs, diseases, genes, and proteins reflect the associations of lncRNAs and diseases from the different perspectives. However, it is difficult for these methods to deeply integrate heterogeneous data from multiple sources. Therefore, we present a novel prediction method based on dual convolutional neural networks to learn the latent representations of lncRNA-disease pairs from the multiple-source data.

2. Experimental Evaluations and Discussions

2.1. Evaluation Metrics

Five-fold cross-validation is used to evaluate the prediction performances of CNNDLP and several state-of-the-art methods for predicting lncRNA-disease associations. All the known lncRNA-disease associations are regarded as positive samples, and the unobserved associations are taken as negatives samples. We randomly divided all the positive samples into five subsets, and four of them are used to training the model. As the number of positive samples is far less than that of the negative samples (ratio of positive samples to negative samples is nearly 1:36 in our study), during the training process, we select the negative samples randomly whose number match to the number of the positive training samples, and these negative samples are also used for training the model. The positive samples in

the remaining subset and all the negative samples are considered as the testing samples. The number of positive samples and that of negative samples during the cross-validation process are listed in Supplement Table S1. In particular, during each cross-validation, the positive samples used for testing are removed and the lncRNA similarities are recalculated by using the remaining positive samples.

We obtain the association scores of testing samples and prioritize them by their scores. The positive samples are ranked higher, which indicate that the prediction performance is better. The lncRNA-disease node pairs whose scores are greater than a classification threshold θ are identified as positive samples, and the ones that have lower scores are determined as negative samples. The true positive rates (TPRs) and the false positive rates (FPRs) at various θ values are calculated as follows:

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \quad (1)$$

where TP and TN are the numbers of positive and negative samples that are identified correctly, while FN and FP are the numbers of misidentified positive and negative samples. The receiver operating characteristic (ROC) curve can be drawn according to the TPRs and FPRs at each various θ , while the area under the ROC curve (AUC) is usually used to evaluate the overall performance of a prediction method [17].

A serious imbalance between the positive samples and the negative ones appears since their ratio is 1:36. For such imbalanced cases, precision-recall (PR) curve is confirmed to be more informative than ROC curve [18]. Therefore, the PR curve is used as another important measurement for the prediction performance of each method. Precision and recall are calculated as follows:

$$precision = \frac{TP}{TP + FP}, recall = \frac{TP}{TP + FN} \quad (2)$$

where *precision* is the proportion of the real positive samples among the samples that are identified as the positive ones, while *recall* is the proportion of the real positive samples to the total actual positive ones. The area under the P-R curve (AUPR) is also utilized to evaluate the performance of lncRNA-disease association prediction [19].

In addition, the top candidate lncRNAs are usually selected by the biologists for further experimental verification of their associations with an interested disease. Therefore, we demonstrate the recall rates of the top 30, 60, and 240 candidates, which demonstrates how many of the positive samples are identified correctly within the ranking list of top k .

2.2. Comparison with Other Methods

To assess the prediction performance of CNNDLP, we compare it with several state-of-the-art methods for predicting disease lncRNAs: SIMCLDA [20], Ping's method [21], MFLDA [15], LDAP [22] and CNNDLP [8]. CNNDLP and the other four methods have specific hyperparameters for fine-tuning to achieve their best association prediction performance. We choose the values of CNNDLP's hyperparameters, α , β and λ , from {0.1, ..., 0.9}. CNNDLP achieved the best performance of five-fold cross-validation, when $\alpha = 0.9$, $\beta = 0.8$ and $\lambda = 0.3$. The prediction performances of CNNDLP at different values of α , β , and λ on CNNDLP in the Supplementary Table S2, Supplementary Table S3, and Supplementary Table S4. In addition, the window size of all convolutional layers and pooling layers in CNNDLP is set as 2×2 . The number of filters in the first and the second convolutional layers n_{conv1} and n_{conv2} are set to 16 and 32 respectively. CNNDLP has a great many parameters, which is easy to make the model overfit all the training samples. Therefore, we adopt dropout strategy and batch normalization to prevent the overfitting. To make a fair comparison, we set the hyperparameters of other methods to the optimal values that are recommended by their respective literatures (i.e., $\alpha_l = 0.8$, $\alpha_d = 0.6$ and $\lambda = 1$ for SIMCLDA, $\alpha = 0.6$ for Ping's method, $\alpha = 105$ for MFLDA, *Gap open* = 10. *Gap extend* = 0.5 for LDAP).

As shown in Figure 1a, CNNDLP yields the highest average performance on all of the 405 diseases (AUC = 0.969). In particular, its performance is increased SIMCLDA by 21.2%, Ping’s method by 9.3%, MFDDA by 34.4%, LDAP by 10.7%, and CNNLDA by 1.7%. The AUCs of the five methods on 10 well-characterized diseases are also listed in Table 1, and CNNDLP achieves the best performance in all of the 10 diseases. The AUC of CNNDLP is slightly better than CNNLDA, but the AUPR of the former is 3.5% higher than the latter. The possible reason for this is that CNNLDA did not learn the low-dimensional network representation of a lncRNA-disease pair. Ping’s method and the LDAP achieved decent performance as they take advantage of the various similarities of different types of lncRNAs and diseases. MFLDA does not perform as well as the other four methods. A possible reason is that it ignored the lncRNA similarity and the disease similarity, while are exploited by the other methods. The improvement of CNNDLP over the compared methods is primarily due to the fact that it deeply learns the attention representation and low-dimensional network-level representation of the lncRNA-disease node pairs.

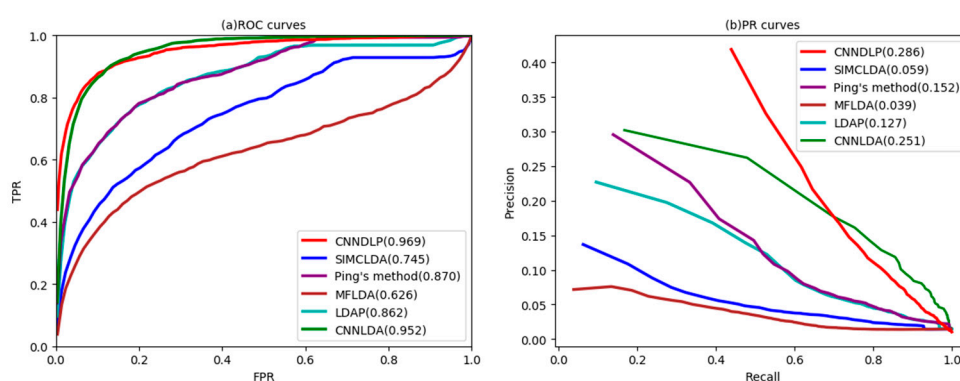


Figure 1. ROC curves and PR curves of CNNDLP and other methods for all diseases.

Table 1. AUCs of CNNDLP and other methods on all the diseases and 10 well-characterized diseases.

Disease Name	CNNDLP	Ping’s Method	AUC LDAP	SIMCLDA	MFLDA	CNNLDA
Prostate cancer	0.951	0.826	0.710	0.874	0.553	0.897
Stomach cancer	0.947	0.930	0.928	0.864	0.467	0.958
Lung cancer	0.976	0.911	0.882	0.790	0.676	0.940
Breast cancer	0.956	0.872	0.830	0.742	0.517	0.836
Reproduce organ cancer	0.943	0.818	0.742	0.707	0.740	0.922
Ovarian cancer	0.970	0.913	0.857	0.786	0.558	0.942
Hematologic cancer	0.989	0.908	0.903	0.828	0.716	0.934
Kidney cancer	0.984	0.979	0.977	0.728	0.677	0.956
Liver cancer	0.956	0.910	0.898	0.799	0.634	0.918
Thoracic cancer	0.921	0.860	0.792	0.792	0.649	0.890
Average AUC of 405 diseases	0.969	0.870	0.745	0.745	0.626	0.952

The bold values indicate the higher AUCs.

As shown in Figure 1b, CNNDLP’s average AUPR is also higher than other methods on 405 diseases (AUPR = 0.286). Its average AUPR is 22.7%, 13.4%, 24.7%, 15.9%, and 3.5% higher SIMCLDA, Ping’s method, MFLDA, LDAP and CNNLDA, respectively. In addition, CNNDLP performs the best performance among nine of the ten well-characterized diseases (Table 2).

Table 2. AUPRs of CNNDLP and other methods on all the diseases and 10 well-characterized diseases.

Disease Name	CNNDLP	Ping's Method	AUPR LDAP	SIMCLDA	MFLDA	CNNLDA
Prostate cancer	0.538	0.333	0.297	0.176	0.092	0.390
Stomach cancer	0.373	0.364	0.094	0.138	0.008	0.286
Lung cancer	0.666	0.437	0.363	0.131	0.171	0.058
Breast cancer	0.485	0.403	0.396	0.047	0.031	0.964
Reproduce organ cancer	0.498	0.281	0.240	0.130	0.103	0.091
Ovarian cancer	0.552	0.483	0.427	0.027	0.023	0.526
Hematologic cancer	0.667	0.403	0.370	0.216	0.121	0.523
Kidney cancer	0.569	0.663	0.462	0.030	0.034	0.584
Liver cancer	0.630	0.498	0.511	0.140	0.110	0.666
Thoracic cancer	0.399	0.383	0.364	0.155	0.102	0.890
Average AUC of 405 diseases	0.286	0.152	0.127	0.059	0.039	0.251

The bold values indicate the higher AUPRs.

A higher recall value in the top k of ranking list indicates that more real lncRNA-disease associations are identified correctly. Figure 2 shows CNNDLP outperforms the other methods at different top k cutoffs, and ranks 88.6% in top 30, ranks 94.6% in top 60, ranks 97.5% in top 90, and ranks 98.3% in top 120. Most of the recall rates of Ping's method are very close to LDAP. The former ranked 68.9%, 81.3%, 87.5% and 92.7% in top 30, 60, 90 and 120, respectively, and the latter ranked 68.5%, 81.7%, 88.0% and 93.3%. MFLDA is still worse than the other methods, and it ranked 42.0%, 53.9%, 61.0% and 65.6%.

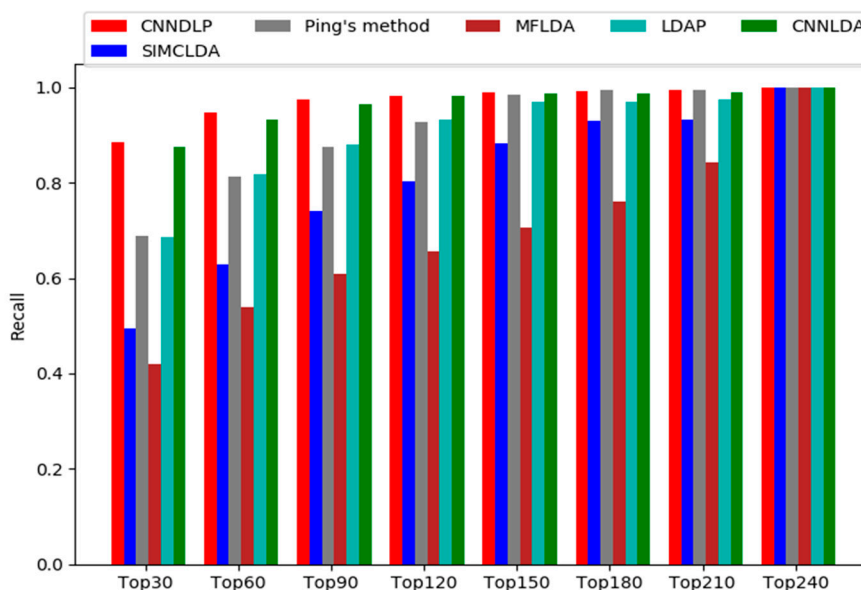


Figure 2. Recall values of top k candidates of CNNDLP and other four methods.

In addition, a paired Wilcoxon test is conducted to confirm whether CNNDLP's prediction performance is significantly greater than the other methods. The statistical results in Table 3 show that CNNDLP yields better performance than the other methods in terms of not only AUCs but AUPRs, as well for the threshold p -value of 0.05.

Table 3. Comparing of different methods based on AUCs with the paired Wilcoxon test.

	SIMCLDA	Ping's Method	MFLDA	LDAP	CNNLDA
<i>p</i> -value of ROC curve	9.2454×10^{-6}	0.00048	5.9940×10^{-7}	0.00121	0.00773
<i>p</i> -value of PR curve	8.3473×10^{-7}	0.04174	3.5037×10^{-8}	0.00126	0.00024

2.3. Case Studies: Stomach Cancer, Breast Cancer and Prostate Cancer

To further demonstrate the capability of CNNDLP to discover potential disease-related candidate lncRNAs, we construct the case studies on stomach cancer, breast cancer, and prostate cancer. For each of these three diseases, we prioritize the candidate lncRNA-disease associations based on their association scores and gather their respective 15 candidates.

Stomach cancer is currently the fourth most common malignant tumor in the world and the second leading cause of cancer-related death [23]. First, Lnc2Cancer is a manually curated database that are verified associations between the lncRNAs and the human cancers by the biological experiments [24]. Twelve of 15 candidates are included by Lnc2Cancer (Table 4), which indicates that these lncRNAs are indeed associated with the disease.

Table 4. The top 15 stomach cancer-related candidate lncRNAs.

Rank	lncRNA Name	Description	Rank	lncRNA Name	Description
1	SPRY4-IT1	Lnc2Cancer, LncRNADisease	9	CDKN2B-AS1	LncRNADisease
2	TINCR	Lnc2Cancer, LncRNADisease	10	CCAT1	Lnc2Cancer, LncRNADisease
3	H19	Lnc2Cancer, LncRNADisease	11	HOTAIR	Lnc2Cancer, LncRNADisease
4	TUSC7	Lnc2Cancer, LncRNADisease	12	GACAT2	LncRNADisease
5	BANCR	Lnc2Cancer, LncRNADisease	13	UCA1	Lnc2Cancer, LncRNADisease
6	MEG3	Lnc2Cancer, LncRNADisease	14	PVT1	Lnc2Cancer, LncRNADisease
7	GAS5	Lnc2Cancer, LncRNADisease	15	MEG8	literature
8	GHET1	Lnc2Cancer, LncRNADisease			

Second, LncRNADisease records more than 4564 lncRNA-disease associations that are obtained from experiments, the published literatures or computation, and then the dysregulation of lncRNAs are manually confirmed [25]. There are 14 candidates contained by the LncRNADisease, indicating they are upregulated or downregulated in stomach cancer tissues. In addition, one candidate labeled by “literature” is supported by the literature, and it is confirmed to have dysregulation in the cancer when compared with the normal tissues [26].

Among the top 15 candidates for breast cancer, 11 candidates are reported in Lnc2Cancer with abnormal expression in breast cancer. (Table 5) LncRNADisease contains 12 candidates, which confirms the associations between these candidates and the disease. The remaining 2 candidates are confirmed by the literatures to have desregulation in the breast cancer [27,28].

Table 5. The top 15 breast cancer-related candidate lncRNAs.

Rank	lncRNA Name	Description	Rank	lncRNA Name	Description
1	SOX2-OT	Lnc2Cancer, LncRNADisease	9	CCAT1	Lnc2Cancer, LncRNADisease
2	HOTAIR	Lnc2Cancer, LncRNADisease	10	GAS5	Lnc2Cancer, LncRNADisease
3	LINC00472	Lnc2Cancer, LncRNADisease	11	MIR124-2HG	literature
4	BCYRN1	LncRNADisease	12	XIST	Lnc2Cancer, LncRNADisease
5	LINC-PINT	literature	13	LINC-ROR	Lnc2Cancer, LncRNADisease
6	MALAT1	Lnc2Cancer, LncRNADisease	14	PANDAR	Lnc2Cancer, LncRNADisease
7	CDKN2B-AS1	LncRNADisease	15	AFAP1-AS1	Lnc2Cancer
8	SPRY4-IT1	Lnc2Cancer, LncRNADisease			

The top 15 prostate cancer-related candidates and the corresponding evidences are listed in Table 6. Fourteen candidates are included by Lnc2Cancer and 14 ones are contained by LncRNADisease, which indicates that they truly are related to the disease. All the case studies confirm that CNNDLP is effective and impactful for discovering potential candidate disease lncRNAs.

Table 6. The top 15 prostate cancer-related candidate lncRNAs.

Rank	lncRNA Name	Description	Rank	lncRNA Name	Description
1	CDKN2B-AS1	LncRNADisease	9	HOTAIR	Lnc2Cancer, LncRNADisease
2	PCGEM1	Lnc2Cancer, LncRNADisease	10	LINC00963	Lnc2Cancer, LncRNADisease
3	PVT1	Lnc2Cancer, LncRNADisease	11	H19	Lnc2Cancer, LncRNADisease
4	GAS5	Lnc2Cancer, LncRNADisease	12	MEG3	Lnc2Cancer, LncRNADisease
5	HOTTIP	Lnc2Cancer, LncRNADisease	13	TUG1	Lnc2Cancer, LncRNADisease
6	NEAT1	Lnc2Cancer, LncRNADisease	14	PCA3	Lnc2Cancer, LncRNADisease
7	PCAT5	Lnc2Cancer	15	DANCR	Lnc2Cancer, LncRNADisease
8	PRINS	Lnc2Cancer, LncRNADisease			

2.4. Prediction of Novel Disease lncRNAs

After five-fold cross validation and case studies to confirm its prediction performance, we further apply CNNDLP to 405 diseases. All the known lncRNA-disease associations are used for training CNNDLP's to predict potential disease-related lncRNAs. The top 50 potential candidates for each of 405 diseases are demonstrated in Supplementary Table S5.

3. Materials and Methods

3.1. Datasets for lncRNA-Disease Association Prediction

We obtained thousands of lncRNA-disease associations, lncRNA-miRNA interactions and miRNA-disease associations from a published work [15]. The human lncRNA-disease database (LncRNADisease) consists of 2687 lncRNA-disease associations that were verified by the biological

experiments, covering 240 lncRNAs and 405 diseases [29]. The disease similarities were calculated based on directed acyclic graphs (DAGs) and the DAGs were constructed based on the disease terms from the U.S. National Library of Medicine (MeSH). The 1002 lncRNA-miRNA interactions were originally extracted from starBase2.0 and they have been confirmed by biological experiments [30], and were involved 495 miRNAs. The 13,559 miRNA-disease associations were obtained from HMDD database [31].

3.2. Bipartite Graphs about the lncRNAs, Diseases, miRNAs, and Representations

We firstly construct a bipartite graph composed of lncRNAs and diseases by connecting them according to the observed lncRNA-disease associations (Figure 3a). The graph is represented by matrix $A = [A_{ij}] \in R^{N_l \times N_d}$, where N_l and N_d are the number of lncRNAs and that of diseases, respectively. Each of rows corresponds to a lncRNA while each of columns represent a disease. If a lncRNA l_i has been observed to be associated a disease d_j , the A_{ij} in A is set to 1, otherwise A_{ij} is 0.

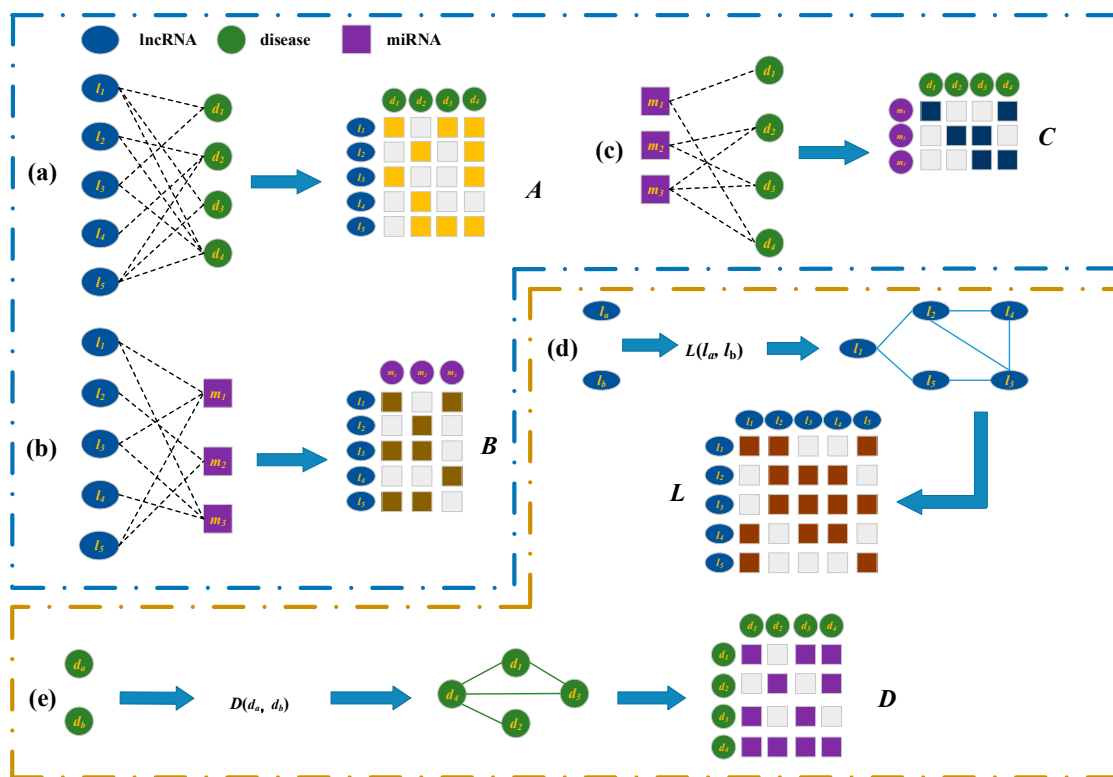


Figure 3. Construction and representation of multiple bipartite graphs. (a) Construct a lncRNA-disease association bipartite graph based on the known associations between lncRNAs and diseases, and its' matrix representation A . (b) Construct lncRNA-miRNA interactions bipartite graph based on the known lncRNA-miRNA interactions, and its' matrix representation B . (c) Construct miRNA-disease association bipartite graph based on known miRNA-disease associations, and its' matrix representation C . (d) Calculate the lncRNA similarity, and construct the matrix representation L . (e) Calculate the disease similarity, and construct the matrix representation D .

There are a great many interactions between the lncRNAs and miRNAs that have been confirmed by the biological experiments [32]. A bipartite graph composed of lncRNA and miRNA nodes is established when there are known interactions between them (Figure 3b). $B = [B_{ij}] \in R^{N_l \times N_m}$ is used to represent interaction matrix, the graph including N_l lncRNAs and N_m miRNAs. If it is known that lncRNA l_i is interacted with miRNA m_j , $B_{ij} = 1$, or $B_{ij} = 0$ when their interaction has not been observed.

An edge is added to connect a miRNA and a disease, when they are observed to have association (Figure 3c). $C = [C_{ij}] \in R^{N_m \times N_d}$ is a matrix representing a bipartite graph with N_m miRNAs and N_d diseases. We set C_{ij} to 1 if miRNA m_i is associated with disease d_j , or 0 when no such association is observed.

3.3. lncRNA-Disease Association Prediction Model Based on CNN

In this section, we describe the prediction model based on convolutional neural networks and attention mechanism for learning the latent representation and predicting the association score of lncRNA l_i and disease d_j . The embedding layer is firstly constructed by incorporating the associations, the similarities, the interactions about lncRNAs, diseases, miRNAs. A novel prediction model is constructed and it is composed of the left and right parts. The left side of the model learns the attention representation of l_i and d_j , while the network representation of l_i and d_j is learned in the right side of model. Each of the two representations goes through a fully connected layer and a softmax layer and the associated possibility between l_i and d_j is obtained and it is regarded as their association score. The final score is the weighted sum of two association scores.

3.3.1. Embedding Layer on the Left

lncRNA Functional Similarity Measurement

On the basis of the biological premise that lncRNAs with similar functions are more possibly to be associated with similar diseases, the similarity of two lncRNAs is measured by their associated diseases. For instance, lncRNA l_a is associated with disease d_1, d_2 and d_4 and lncRNA l_b is associated with diseases d_2, d_4 and d_5 . The similarity between $E_a = \{d_1, d_2, d_4\}$ and $E_b = \{d_2, d_4, d_5\}$ is regarded as the functional similarity of l_a and l_b . The lncRNA similarity that are used by us is calculated according to the Xuan's method [8]. Matrix $L = [L_{ij}] \in R^{N_l \times N_l}$ is the lncRNA similarity matrix (Figure 3d), where L_{ij} is the similarity of lncRNAs l_i and l_j , L_{ij} value changes between 0 and 1.

Disease Similarity Measurement

All semantic terms related to a disease form its directed acyclic graph (DAG). The semantic similarities between the diseases are successfully calculated by Wang et al. based on their DAGs [33]. We calculate the disease similarities according to Wang's method, and the similarities can be represented by matrix $D = [D_{ij}] \in R^{N_d \times N_d}$, where D_{ij} is the similarity of disease d_i and d_j (Figure 3e). The similarity of two diseases also varies between 0 and 1.

The Left Embedding Layer for Integrating the Original Information

If a lncRNA and a disease have similarity relationships and association relationships with the more common lncRNAs, they are more likely to associated with each other. We take the lncRNA l_1 and the disease d_2 as an example to explain the process of constructing the embedding layer on the left. As shown in Figure 4, let L_1 represents the first row of L which records the similarities between l_1 and all the lncRNAs. The second row of A^T, A_2^T , contains the associations between d_2 and all the lncRNAs. For example, as l_1 is similar to l_2 and l_5 , and d_2 has been associated with l_2, l_4 and l_5 , l_1 is possibly related to d_2 . We stack L_1 and A_2^T together as the first part of the embedding layer. Similarly, l_1 and d_2 are more likely to associate when l_1 and d_2 have the similarity and association connections with more common diseases. Therefore, we stack A_1 and D_2 as the second part of the embedding layer. In addition, when a lncRNA and a disease have interaction and association relationships with the common miRNAs, they are more possibly to have association. For instance, there is a possible association between l_1 and d_2 , since l_1 interacts with miRNAs m_1 and m_3 , and disease d_2 is associated with m_2 and m_3 . The first row of B and the second row G^T are stacked as the third part of the embedding layer. The final embedding layer matrix between l_1 and d_2 is denoted as $X \in R^{2 \times (N_l + N_m + N_d)}$.

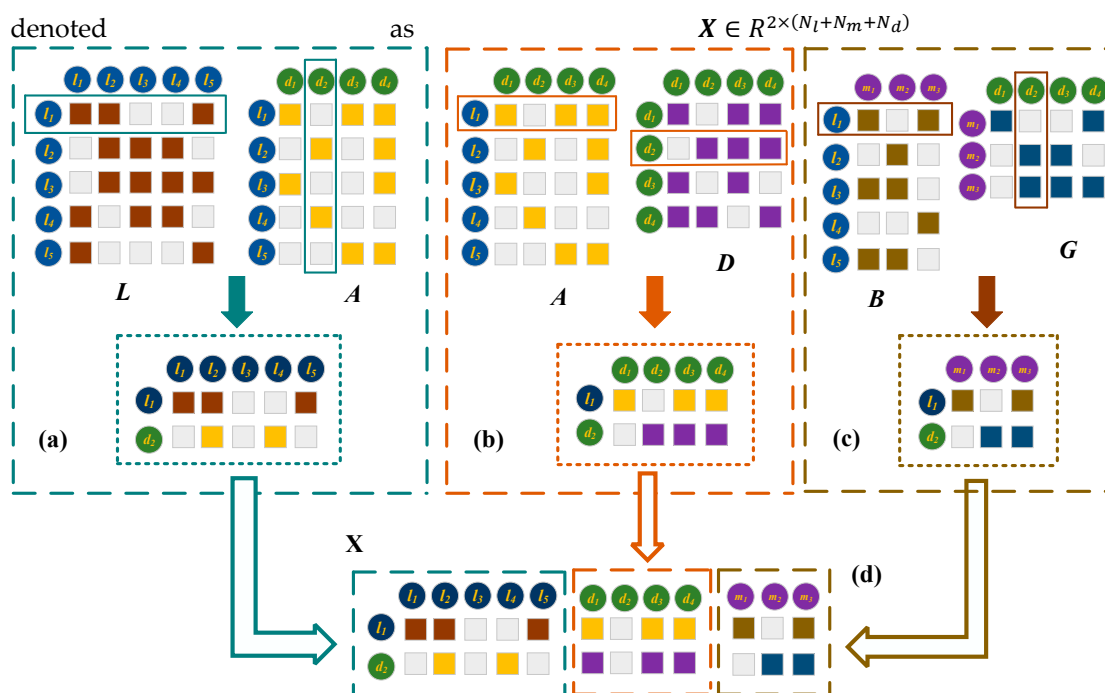


Figure 4. Construction of the left embedding layer matrix of l_1 and d_2 , X . (a) Construct the first part of X by exploiting the IncRNA similarities and the IncRNA-disease associations. (b) Construct the second part of X by integrating the IncRNA-disease associations and the disease similarities. (c) Construct the third part of X by incorporating the IncRNA-miRNA interactions and the miRNA-disease associations. (d) Concatenate the three parts of X .

Attention at the Adjacent Edge Level

For a lncRNA node or a disease node, not all the adjacent edges of the node have equal contributions for learning the representation of a pair of lncRNA-disease. In order to solve the issue, we establish the attention mechanism to enhance the adjacent edges that are important for predicting the lncRNA-disease associations. In the embedding layer matrix X , X_{ij} represents the connection case between the i -th node and the j -th node, and X_{ij} is assigned an attention weight α_{ij} , which is defined as follows,

$$F_{ij} = \tan h(WX_{ij} + b) \tag{3}$$

$$\alpha_{ij} = \frac{\exp(F_{ij}^T u_e)}{\sum_j \exp(F_{ij}^T u_e)} \tag{4}$$

$$\hat{X}_{ij} = \alpha_{ij} X_{ij} \tag{5}$$

where W and b are a weight matrix and a context vector respectively, and u_e is a bias vector. F_{ij} is an attention score that represents the importance of X_{ij} . α_{ij} is a normalized importance X_{ij} . \hat{X} is the enhanced embedding layer matrix after the attention mechanism at the adjacent edge level is applied for X .

3.3.2. Embedding Layer on the Right

First, it is known that two lncRNA nodes are similar if they are associated with some common disease nodes [22]. In the bipartite network of lncRNA-disease (Figure 5a), lncRNA l_1 and l_3 are associated with a common disease node d_2 , so l_1 and l_3 are similar. l_3 and l_5 are also similar because they are related to a disease node d_1 (Figure 5b). Similarly, d_1 is similar to d_2 as they are associated with common lncRNA node l_3 (Figure 5c). Second, if two lncRNA nodes have no common neighboring

nodes, while they are related to some similar disease nodes, they are also similar to each other [22]. As shown in Figure 5d, l_1 and l_5 are similar, because their neighboring nodes d_1 and d_2 are similar. Similarly, d_2 and d_3 are similar as they are associated with similar neighboring nodes l_3 and l_5 (Figure 5e). Ping et al. successfully measured the lncRNA similarities and the disease similarities by utilizing the lncRNA-disease bipartite network.

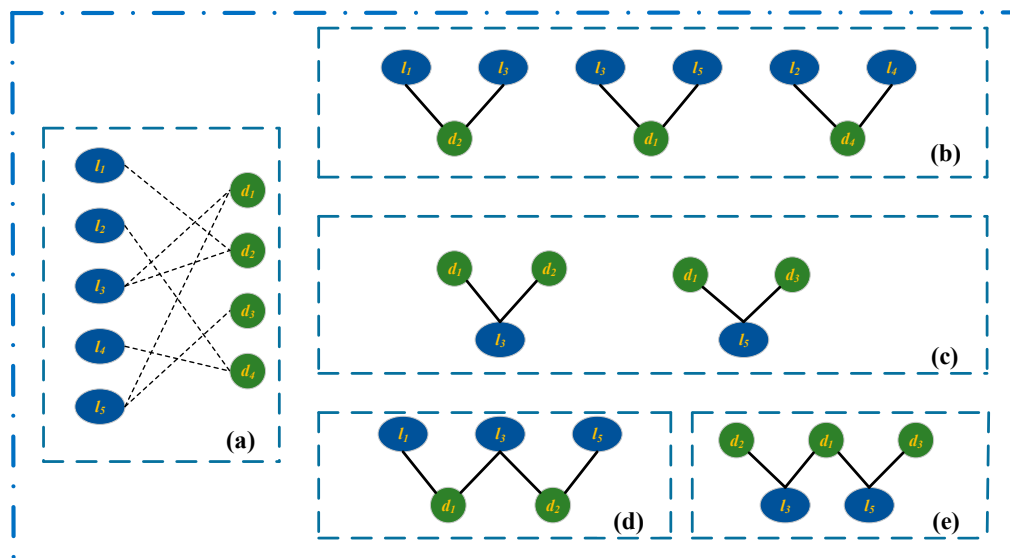


Figure 5. Calculation of the first type of lncRNA similarity and the first type of disease similarity. (a) The lncRNA-disease association bipartite network. (b) Calculate the lncRNA similarities based on the common associated disease nodes. (c) Compute the disease similarities based on their common related lncRNA nodes. (d) Calculate the lncRNA similarities according to their associated similar disease nodes. (e) The disease similarity calculation based on their related similar lncRNA nodes.

Unlike Ping’s method that focused on a single bipartite network, multiple kinds of lncRNA similarities and disease similarities are calculated by utilizing the bipartite networks from different sources about lncRNA-disease associations, lncRNA-miRNA interactions and miRNA-disease associations. The first kind of lncRNA similarity $L^{(1)} = [L_{ij}^{(1)}] \in R^{N_l \times N_l}$, and the first kind of disease similarity $D^{(1)} = [D_{ij}^{(1)}] \in R^{N_d \times N_d}$ are calculated according to Ping’s method. The second kind of lncRNA similarity is measured by exploiting the information of common miRNA nodes and similar ones interacting with two lncRNA nodes in the lncRNA-miRNA bipartite network, and it is denoted as $L^{(2)} = [L_{ij}^{(2)}] \in R^{N_l \times N_l}$. Finally, the second kind of disease similarity $D^{(2)} = [D_{ij}^{(2)}] \in R^{N_d \times N_d}$ is calculated based on the miRNA-disease bipartite network.

In order to incorporate two kinds of lncRNA similarities $L^{(1)}$ and $L^{(2)}$, the final lncRNA similarity $L^{(c)}$ is defined as follows,

$$L^{(c)} = \alpha L^{(1)} + (1 - \alpha)L^{(2)} \tag{6}$$

where α is the parameter utilized to control the contributions of $L^{(1)}$ and $L^{(2)}$.

Similarly, the final disease similarity $D^{(c)}$ is the weighted sum of $D^{(1)}$ and $D^{(2)}$, as follows,

$$D^{(c)} = \beta D^{(1)} + (1 - \beta)D^{(2)} \tag{7}$$

where β is a parameter for regulating the weights of $D^{(1)}$ and $D^{(2)}$.

The right embedding layer for integrating the second kinds of lncRNA and disease similarities. The establishment of the right embedding layer matrix $Y \in R^{2 \times (N_l + N_m + N_d)}$ is similar to the left embedding layer matrix X. First, we stack the first row of $L^{(c)}$, $L_1^{(c)}$, and A_2^T together as the first part

of the embedding layer. Second, A_1 and $D_2^{(c)}$ are stacked as the second part of the embedding layer. Finally, the first row of B and the second row G^T are stacked as the third part of the embedding layer.

3.4. Convolutional Module on the Left

In this section, we describe a novel model based on convolutional neural networks with adjacent attention for learning latent representations of lncRNA-disease node pairs. The overall architecture is shown in Figure 6. We describe the left convolutional module in detail. Left module includes convolution and activation layer, max-pooling layer, fully connected layer. The embedding matrix $\hat{X} \in R^{2 \times (N_l + N_m + N_d)}$ is inputted the convolutional module to learn an original representation of a pair of lncRNA-disease node.

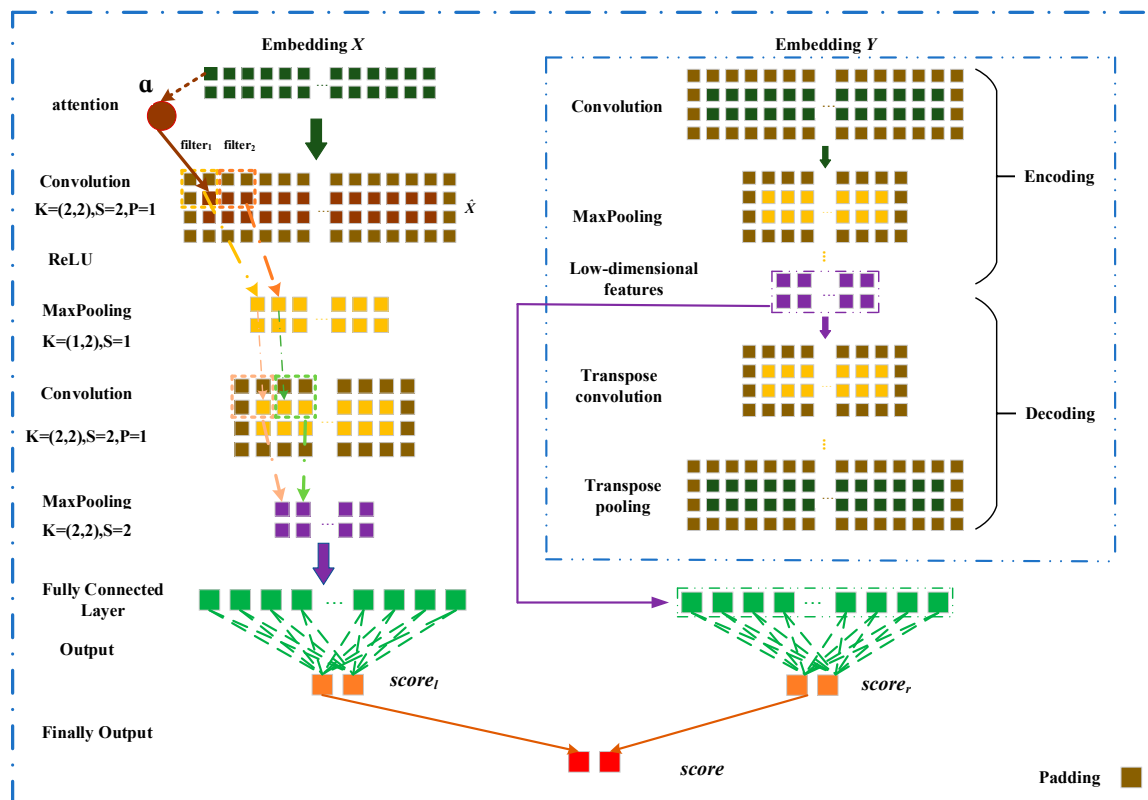


Figure 6. Construction of the prediction model based on the convolutional neural network and convolutional autoencoder for learning the attention representation and the low-dimensional network representation.

For a convolutional layer, the length and the width of a filter are set to w and h respectively, which means the filter is applied on $w \times h$ features. In order to learn the marginal information of the embedding matrix \hat{X} , we pad zeros around \hat{X} . Let the number of filters is n_{conv} . The convolution filters $W_{conv} \in R^{w \times h \times n_{conv}}$ are applied to the embedding matrix \hat{X} , and obtain the feature maps $Z \in R^{n_{conv} \times (4-w+1) \times (2+N_l+N_m+N_d-h+1)}$. \hat{X}_{ij} is the element at the i -th row and j -th column of \hat{X} . $\hat{X}_{k,i,j}$ represent a region in a filter when the k th filter slides the position \hat{X}_{ij} .

$$\hat{X}_{k,i,j} = \hat{X}(i : i + w, j : j + h) \tag{8}$$

$$Z_k(i, j) = g(W_{k,i,j} * \hat{X}_{k,i,j} + b(k)) \tag{9}$$

$$i \in [1, 4 - w + 1] j \in [1, 2 + N_l + N_m + N_d - h + 1], k \in [1, n_{conv}]$$

$Z_k(i, j)$ is the element at the i -th row and the j -th column of the k -th feature map. g is relu function that it is a nonlinear activation function [34], W_k is the weight matrix of the k -th filter and b is a bias vector.

The max-pooling layer is used to down-sample the features of the feature maps $Z_k(i, j)$, and it outputs the most important feature in each feature map. Given an input $Z_k(i, j)$, the output of pooling layer is shown as follows,

$$V_k(i, j) = \max(Z_k(i : i + w_p, j + h_p)) \quad (10)$$

$$i \in [1, 5 - w - w_p + 1], j \in [1, 3 + N_l + N_m + N_d - h - h_p + 1], k \in [1, n_{conv}]$$

where w_p is the length of a filter of pooling layer and h_p is the width. $V_k(i, j)$ is the element at the i -th row and the j -th column in the k th feature map. \hat{X} goes through two convolutional and two max-pooling layers, and we obtain the original representation Z_{left} of l_1 and d_2 from the left convolutional module.

Finally, Z_{left} is flattened to a vector z , which z is feed to fully connected layer. A softmax layer is used to normalized the output of the fully connected layer and we have

$$score_l = softmax(Wz + b) \quad (11)$$

where W is the weight matrix, and b is a bias vector. $score_l$ is an associated probability distribution of C class ($C = 2$). $score_l$ is the probability that the lncRNA l_1 is associated with the disease d_2 and $score_l^0$ is the probability that l_1 and d_2 have no association relationship.

Similarly, the embedding matrix $Y \in R^{2 \times (N_l + N_m + N_d)}$ is feed to the convolutional module on the right side of the prediction model for learning the network representation Z_{right} of l_1 and d_2 . $score_r$ of l_1 and d_2 are obtained when Z_{right} is feed to the full connection layer and the softmax layer.

3.5. Convolutional Autoencoder Module on the Right

The matrices about lncRNA-disease association, lncRNA-miRNA interaction, and miRNA-disease association are very sparse, resulting in many 0 elements are contained in the embedding matrix Y . An autoencoder based convolutional neural network is constructed to learn important and low-dimensional feature representations of lncRNA-disease pair on the right side of CNNDLP. The encoding and decoding strategies are given as follows,

3.5.1. Encoding Strategy

The embedding layer matrix $Y \in R^{2 \times (N_l + N_d + N_m)}$ is mapped into the low-dimensional feature space through encoding based on convolutional neural network. $Y_{encode,k}^{(n-1)}$ is inputted to the n -th convolution layer to obtain $Z_{encode,k}^{(n)}$. $Y_{encode,k}^{(n)}$ is formed after $Z_{encode,k}^{(n)}$ passes the n -th max-pooling layer. They are defined as follows,

$$Z_{encode,k}^{(n)}(i, j) = g(W_{encode,k}^{(n)} * Y_{encode,k}^{(n-1)}(i, j) + b_{encode}^{(n)}(k)) \quad (12)$$

$$Y_{encode,k}^{(n)}(i, j) = \max(Z_{encode,k}^{(n)}(i : i + w_e, j + h_e)) \quad (13)$$

$$i \in [1, 4 - w_e + 1], j \in [1, 2 + N_l + N_m + N_d - h_e + 1], k \in [1, n_{encode}], n \in [1, H_e]$$

where H_e is the total number of encoding layers, and $Y_{encode}^{(0)} = Y$. k represents the k -th filter and n_{encode} is the number of filters during encoding process. $Z_{encode,k}^{(n)}(i, j)$ and $Y_{encode,k}^{(n)}(i, j)$ are the elements at the i -th row and the j -th column of the k -th feature map, respectively. $W_{encode,k}^{(n)}$ is a weight matrix and $b_{encode}^{(n)}(k)$ is a bias vector.

3.5.2. Decoding Strategy

The output of the H_e -th encoding layer $Y_{encode}^{(H_e)}$ is used as the input of the decoder. It is a matrix that is similar to Y by decoding. The decoding process includes both the transpose convolution layer and transpose pooling layer, and they are respectively defined as,

$$Z_{decode,k}^{(n)}(i, j) = g\left(W_{decode,k}^{(n)} * Y_{decode,k,i,j}^{(n-1)} + b_{decode}^{(n)}(k)\right) \quad (14)$$

$$Y_{decode,k}^{(n)}(i, j) = \text{Maxunpool}\left(W_{decode,k}^{(n)} * Z_{decode,k}^{(n)}(i, j) + b_{decode}^{(n)}(k)\right) \quad (15)$$

$$k \in [1, n_{decode}], n \in [1, H_d]$$

$Y_{decode,k}^{(n)}$ and $Z_{decode,k}^{(n)}$ are the outputs of the n -th transpose convolution layer and transpose max-pooling layer, respectively. H_d is the total number of decoding layers, and n_{decode} is the number of filters for decoding. As $Y_{decode}^{(H_d)}$ should be consistent with Y , we defined the loss function as follows,

$$\text{loss}_{auto} = \frac{\sum_{i=1}^T (Y_i - (Y_{decode})_i)^2}{T} \quad (16)$$

where Y_{decode} is the output of decoding and Y is the input of encoding; Y_i is corresponding to the i -th training sample (lncRNA-disease pair), and T is the number of training samples. The $score_l$ of l_1 and d_2 is obtained after the Y_{encode} is feed to the full connection layer and the softmax layer.

3.6. Combined Strategy

In our model, the cross-entropy is used as the loss function, for the left and right parts of the prediction model loss functions are defined as follows,

$$\text{loss}_1 = - \sum_{i=1}^T \left[y_{label} \log(score_l) + (1 - y_{label}) \log(1 - score_l) \right] \quad (17)$$

$$\text{loss}_2 = - \sum_{i=1}^T \left[y_{label} \log(score_r) + (1 - y_{label}) \log(1 - score_r) \right] \quad (18)$$

where y_{label} denotes the actual association label between a lncRNA and a disease. y_{label} is 1 when the lncRNA is indeed associated with the disease, otherwise y_{label} is 0. T is the number of training samples.

The final score of our model is a weighted sum of $score_l$ and $score_r$ as follows,

$$score = \lambda score_l + (1 - \lambda) score_r \quad (19)$$

where the parameter $\lambda \in (0, 1)$ is used to adjust the importance of $score_l$ and $score_r$.

4. Conclusions

A novel method based on the convolutional neural network with adjacent edge attention and convolutional autoencoder, entitled CNNDLP, is developed for inferring potential candidate lncRNA-disease associations. Two embedding layers are constructed from the biological perspective for integrating heterogeneous data about lncRNAs, diseases, and miRNAs from multiple sources. We construct the attention mechanism at the adjacent edge level to discriminate the different contributions of edges and the latent representation of a lncRNA-disease pair is learned from the more informative edges by the left side of CNNDLP's prediction model. On the basis of calculating the new type of lncRNA similarity and that of disease similarity, the right side of CNNDLP's model captures the complex relationships among these similarities and the lncRNA-disease associations, as well as the

topological structures of multiple heterogeneous networks. The novel prediction model based on the convolutional neural network learns the attention representation and the low-dimensional network one of the lncRNA-disease pair. The experimental results demonstrated that CNNDLP outperforms the other methods in terms of not only AUCs but AUPRs as well. In particular, CNNDLP is more beneficial for the biologists as the top part of its ranking list may retrieve more real lncRNA-disease associations. Case studies on three diseases further confirm that CNNDLP is able to discover the potential candidate disease-related lncRNAs. CNNDLP may serve as a powerful prioritization tool that screens prospective candidates for the subsequent discovery of actual lncRNA-disease associations through wet-lab experiments.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/17/4260/s1>.

Author Contributions: Author Contributions: P.X. and N.S. conceived the prediction method, and they wrote the paper. N.S. and Y.L. developed the computer programs. T.Z. and Y.G. analyzed the results and revised the paper.

Funding: The work was supported by the Natural Science Foundation of China (61972135), the Natural Science Foundation of Heilongjiang Province (LH2019F049, LH2019A029), the China Postdoctoral Science Foundation (2019M650069), the Heilongjiang Postdoctoral Scientific Research Staring Foundation (BHL-Q18104), the Fundamental Research Foundation of Universities in Heilongjiang Province for Technology Innovation (KJCX201805), the Innovation Talents Project of Harbin Science and Technology Bureau (2017RAQXJ094), and the Fundamental Research Foundation of Universities in Heilongjiang Province for Youth Innovation Team (RCYJTD201805).

Acknowledgments: We would like to thank Editage (www.editage.com) for English language editing.

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

References

1. Fu, X.-D. Non-coding RNA: A new frontier in regulatory biology. *Natl. Sci. Rev.* **2014**, *1*, 190–204. [[CrossRef](#)] [[PubMed](#)]
2. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D.G.; et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **2012**, *22*, 1775–1789. [[CrossRef](#)]
3. Guttman, M.; Rinn, J.L. Modular regulatory principles of large non-coding RNAs. *Nature* **2012**, *482*, 339–346. [[CrossRef](#)] [[PubMed](#)]
4. Wapinski, O.; Chang, H.Y. Long noncoding RNAs and human disease. *Trends Cell Biol.* **2011**, *21*, 354–361. [[CrossRef](#)] [[PubMed](#)]
5. Chen, X.; Yan, G.-Y. Novel human lncRNA-disease association inference based on lncRNA expression profiles. *Bioinformatics* **2013**, *29*, 2617–2624. [[CrossRef](#)] [[PubMed](#)]
6. Zhao, T.; Xu, J.; Liu, L.; Bai, J.; Xu, C.; Xiao, Y.; Li, X.; Zhang, L. Identification of cancer-related lncRNAs through integrating genome, regulome and transcriptome features. *Mol. Biosyst.* **2015**, *11*, 126–136. [[CrossRef](#)] [[PubMed](#)]
7. Lan, W.; Huang, L.; Lai, D.; Chen, Q. Identifying Interactions Between Long Noncoding RNAs and Diseases Based on Computational Methods. *Methods Mol. Biol.* **2018**, *1754*, 205–221.
8. Xuan, P.; Cao, Y.; Zhang, T.; Kong, R.; Zhang, Z. Dual Convolutional Neural Networks with Attention Mechanisms Based Method for Predicting Disease-Related lncRNA Genes. *Front. Genet.* **2019**, *10*, 416. [[CrossRef](#)] [[PubMed](#)]
9. Yu, G.; Fu, G.; Lu, C.; Ren, Y.; Wang, J. BRWLDA: Bi-random walks for predicting lncRNA-disease associations. *Oncotarget* **2017**, *8*, 60429–60446. [[CrossRef](#)]
10. Chen, X.; You, Z.-H.; Yan, G.-Y.; Gong, D.-W. IRWLDA: Improved random walk with restart for lncRNA-disease association prediction. *Oncotarget* **2016**, *7*, 57919–57931. [[CrossRef](#)]
11. Xiao, X.; Zhu, W.; Liao, B.; Xu, J.; Gu, C.; Ji, B.; Yao, Y.; Peng, L.; Yang, J. BPLDA: Predicting lncRNA-Disease Associations Based on Simple Paths with Limited Lengths in a Heterogeneous Network. *Front. Genet.* **2018**, *9*, 411. [[CrossRef](#)] [[PubMed](#)]

12. Sun, J.; Shi, H.; Wang, Z.; Zhang, C.; Liu, L.; Wang, L.; He, W.; Hao, D.; Liu, S.; Zhou, M. Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network. *Mol. Biosyst.* **2014**, *10*, 2074–2081. [[CrossRef](#)] [[PubMed](#)]
13. Ding, L.; Wang, M.; Sun, D.; Li, A. TPGLDA: Novel prediction of associations between lncRNAs and diseases via lncRNA-disease-gene tripartite graph. *Sci. Rep.* **2018**, *8*, 1065. [[CrossRef](#)] [[PubMed](#)]
14. Xuan, Z.; Li, J.; Yu, J.; Feng, X.; Zhao, B.; Wang, L. A Probabilistic Matrix Factorization Method for Identifying lncRNA-disease Associations. *Genes* **2019**, *10*, 126. [[CrossRef](#)] [[PubMed](#)]
15. Fu, G.; Wang, J.; Domeniconi, C.; Yu, G. Matrix factorization-based data fusion for the prediction of lncRNA-disease associations. *Bioinformatics* **2017**, *34*, 1529–1537. [[CrossRef](#)] [[PubMed](#)]
16. Zhang, J.; Zhang, Z.; Chen, Z.; Deng, L. Integrating Multiple Heterogeneous Networks for Novel lncRNA-Disease Association Inference. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 396–406. [[CrossRef](#)] [[PubMed](#)]
17. Xuan, P.; Ye, Y.; Zhang, T.; Zhao, L.; Sun, C. Convolutional Neural Network and Bidirectional Long Short-Term Memory-Based Method for Predicting Drug-Disease Associations. *Cells* **2019**, *8*, 705. [[CrossRef](#)] [[PubMed](#)]
18. Saito, T.; Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* **2015**, *10*, e0118432. [[CrossRef](#)] [[PubMed](#)]
19. Xuan, P.; Shen, T.; Wang, X.; Zhang, T.; Zhang, W. Inferring disease-associated microRNAs in heterogeneous networks with node attributes. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2018**. [[CrossRef](#)] [[PubMed](#)]
20. Lu, C.; Yang, M.; Luo, F.; Wu, F.-X.; Li, M.; Pan, Y.; Li, Y.; Wang, J. Prediction of lncRNA-disease associations based on inductive matrix completion. *Bioinformatics* **2018**, *34*, 3357–3364. [[CrossRef](#)]
21. Ping, P.; Wang, L.; Kuang, L.; Ye, S.; Iqbal, M.F.B.; Pei, T. A Novel Method for lncRNA-Disease Association Prediction Based on an lncRNA-Disease Association Network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 688–693. [[CrossRef](#)] [[PubMed](#)]
22. Lan, W.; Li, M.; Zhao, K.; Liu, J.; Wu, F.-X.; Pan, Y.; Wang, J. LDAP: A web server for lncRNA-disease association prediction. *Bioinformatics* **2016**, *33*, 458–460. [[CrossRef](#)] [[PubMed](#)]
23. Ferlay, J.; Shin, H.-R.; Bray, F.; Forman, D.; Mathers, C.; Maxwell Parkin, D. Estimates of worldwide burden of cancer in 2008: GLOBOCAN. *Int. J. Cancer* **2010**, *127*, 2893–2917. [[CrossRef](#)] [[PubMed](#)]
24. Ning, S.; Zhang, J.; Wang, P.; Zhi, H.; Wang, J.; Liu, Y.; Gao, Y.; Guo, M.; Yue, M.; Wang, L.; et al. lnc2Cancer: A manually curated database of experimentally supported lncRNAs associated with various human cancers. *Nucleic Acids Res.* **2016**, *44*, D980–D985. [[CrossRef](#)] [[PubMed](#)]
25. Bao, Z.; Yang, Z.; Huang, Z.; Zhou, Y.; Cui, Q.; Dong, D. lncRNADisease 2.0: An updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.* **2019**, *47*, D1034–D1037. [[CrossRef](#)] [[PubMed](#)]
26. He, Y.; Luo, Y.; Liang, B.; Ye, L.; Lu, G.; He, W. Potential applications of MEG3 in cancer diagnosis and prognosis. *Oncotarget* **2017**, *8*, 73282–73295. [[CrossRef](#)]
27. Xu, Y.; Chen, M.; Liu, C.; Zhang, X.; Li, W.; Cheng, H.; Zhu, J.; Zhang, M.; Chen, Z.; Zhang, B. Association Study Confirmed Three Breast Cancer-Specific Molecular Subtype-Associated Susceptibility Loci in Chinese Han Women. *Oncologist* **2017**, *22*, 890–894. [[CrossRef](#)]
28. Lv, X.-B.; Jiao, Y.; Qing, Y.; Hu, H.; Cui, X.; Lin, T.; Song, E.; Yu, F. miR-124 suppresses multiple steps of breast cancer metastasis by targeting a cohort of pro-metastatic genes in vitro. *Chin. J. Cancer* **2011**, *30*, 821–830. [[CrossRef](#)]
29. Chen, G.; Wang, Z.; Wang, D.; Qiu, C.; Liu, M.; Chen, X.; Zhang, Q.; Yan, G.; Cui, Q. lncRNADisease: A database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.* **2013**, *41*, D983–D986. [[CrossRef](#)]
30. Li, J.H.; Liu, S.; Zhou, H.; Qu, L.H.; Yang, J.H. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **2014**, *42*, D92–D97. [[CrossRef](#)]
31. 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](#)] [[PubMed](#)]
32. Paraskevopoulou, M.D.; Hatzigeorgiou, A.G. Analyzing MiRNA-lncRNA Interactions. *Methods Mol. Biol.* **2016**, *1402*, 271–286. [[PubMed](#)]

33. 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](#)] [[PubMed](#)]
34. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814.



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).