



Research article

HRGCNLDA: Forecasting of lncRNA-disease association based on hierarchical refinement graph convolutional neural network

Li Peng^{1,2,*}, Yujie Yang¹, Cheng Yang¹, Zejun Li³ and Ngai Cheong^{4,*}

¹ College of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China

² Hunan Key Laboratory for Service Computing and Novel Software Technology, Hunan University of Science and Technology, Xiangtan 411201, China

³ School of Computer Science and Engineering, Hunan Institute of Technology, Hengyang 421002, China

⁴ Faculty of Applied Sciences, Macao Polytechnic University, Macau 999078, China

* **Correspondence:** Email: plpeng@hnu.edu.cn, ncheong@mpu.edu.mo.

Abstract: Long non-coding RNA (lncRNA) is considered to be a crucial regulator involved in various human biological processes, including the regulation of tumor immune checkpoint proteins. It has great potential as both a cancer biomolecular biomarker and therapeutic target. Nevertheless, conventional biological experimental techniques are both resource-intensive and laborious, making it essential to develop an accurate and efficient computational method to facilitate the discovery of potential links between lncRNAs and diseases. In this study, we proposed HRGCNLDA, a computational approach utilizing hierarchical refinement of graph convolutional neural networks for forecasting lncRNA-disease potential associations. This approach effectively addresses the over-smoothing problem that arises from stacking multiple layers of graph convolutional neural networks. Specifically, HRGCNLDA enhances the layer representation during message propagation and node updates, thereby amplifying the contribution of hidden layers that resemble the ego layer while reducing discrepancies. The results of the experiments showed that HRGCNLDA achieved the highest AUC-ROC (area under the receiver operating characteristic curve, AUC for short) and AUC-PR (area under the precision versus recall curve, AUPR for short) values compared to other methods. Finally, to further demonstrate the reliability and efficacy of our approach, we performed case studies on the case of three prevalent human diseases, namely, breast cancer, lung cancer and gastric cancer.

Keywords: lncRNA-disease associations prediction; graph convolutional neural network; hierarchical refinement

1. Introduction

LncRNAs are RNA transcripts that typically exceed 200 nucleotides in length and generally do not serve as templates for protein or peptide synthesis. Instead, they execute diverse biological functions through the modulation of gene expression and activity, spanning the transcriptional, translational, and post-translational dimensions [1]. According to recent research, approximately 74.7% of genes within the human genome are involved in transcription processes, yet only around 1.5% of these genes are responsible for encoding proteins [2]. In recent years, functional genomics research has discovered numerous lncRNAs, with lncRNAs often emerging as the principal regulators of gene expression. LncRNAs are also directly related to the modulation of transcription factors, including lncRNA NFAT, a calcium-sensitive nuclear factor that activates T cells, and the fly homolog Ultrabithorax gene (Ubx) [3]. LncRNA LUCAT1, by regulating microRNA-7-5p and reducing its expression level, is considered a latent cancer treatment biological marker for breast cancer [4]. Hence, the identification of potential lncRNA-disease links aids in identifying the exact function of lncRNAs and gains insights into the potential disease mechanisms at the molecular level of lncRNAs.

As high-throughput sequencing technology advances rapidly and multiple disease databases are being established, a large amount of biological information has been generated, and mining the potential value of biological big data is of great importance to the field of intelligent medicine [5–7]. Researchers have developed a keen interest in various biological problems, such as inference of gene regulatory networks [8, 9], identification of biological entity-associated non-coding RNAs (CircRNA [10, 11], lncRNA [12, 13], micro-RNA, miRNA for short [14–19], etc.), and the discovery of potential targets for drugs [20]. The generation of large amounts of lncRNA sequence information and semantic information of diseases has made it possible to comprehensively analyze the development of diseases at the molecular level of lncRNAs. Thus, more and more researchers are devoting themselves to predicting potential associations between lncRNAs and diseases (for example, cancer, COVID-19, etc. [21, 22]) and have created many databases to document known associations between lncRNAs and diseases validated through experiments, such as the LncRNADisease [23] and Lnc2Cancer [24] databases. Among them, the LncRNADisease v2.0 collects and collates approximately 205,959 experimentally supported lncRNA-disease association entries, including 529 diseases and 19,166 lncRNAs. The Lnc2Cancer v3.0 records 9254 known lncRNA-cancer associations by reviewing more than 1500 published papers associations, including 2659 human lncRNAs and 216 human cancer subtypes. Although some new lncRNA-disease known associations have been added to the newly released databases, a large number of potential lncRNA-disease associations remain undiscovered. However, traditional biological experimental methods are not only costly but also time consuming when it comes to uncovering potential lncRNA-disease associations, and, hence, it becomes crucial to create computational models that are both effective and efficient in predicting potential lncRNA-disease associations [25–27].

In recent years, some researchers have developed many methods for predicting lncRNA-disease associations, aimed at identifying lncRNAs linked to specific diseases [28–30]. These methods encompass machine learning approaches such as Bayes, which is defined as the probability estimation of a relationship given prior information [31, 32], support vector machine, which is a supervised machine learning method for binary classification [33], matrix decomposition and completion methods, network propagation techniques, and deep learning strategies [34]. Among them, the utilization of graph neural networks (GNN) for predicting lncRNA-disease correlations has garnered growing interest [35]. For

instance, Xuan et al. [36] introduced a computational model, called GCNLDA, which utilized graph convolutional neural networks (GCN) and convolutional neural networks (CNNs—the training of CNNs was proved to be faster than long-short term memory [37]) for capturing both global and local representations illustrating the associations between lncRNAs and diseases. Considering the complexity of the GCNLDA framework and the large number of parameters that need to be adjusted, Wu et al. [38] proposed an approach that utilized graph auto-encoders (GAE) in combination with random forest (RF) to identify disease-related lncRNAs. The method employs GAE to learn vectors that represent the low-dimensional features of nodes extracted from the network, thereby reducing the dimensionality and heterogeneity of biological data. Sheng et al. [39] encompassed both consistent and diverse data within the lncRNA-miRNA-disease network. They proposed an auto-encoder model based on a multichannel graph attention network (GAT) for capturing a wide spectrum of nuanced insights from complex graphs, inter-graphs, and intra-graphs pertaining lncRNA and disease nodes. Ultimately, an RF classifier was employed to forecast plausible associations between lncRNAs and diseases.

The GCN-based approach represents a computational model tailored for graph data, proficient in extracting node feature details, consolidating information from adjacent nodes, and leveraging the network's topological structure [40, 41]. However, these types of models are prone to over-smoothing when too many stacked network layers are used, resulting in a degradation of model prediction performance. Within this paper, we introduce a computational model, namely, the hierarchical refinement graph convolutional neural network (HRGCNLDA), shown in Figure 1, for predicting associations between long non-coding RNAs and diseases. This model enhances the contribution of each layer to the final feature embedding of a node by determining the weight coefficient of the current layer through cosine similarity calculation with respect to the ego layer. In brief, the implementation of the model can be divided into four steps as follows:

- Apply affine transformations to the initial attributes related to lncRNAs and diseases to align their initial feature dimensions.
- Refine the aggregated feature embeddings for each layer by applying hierarchical refinement-based GCN.
- Select the appropriate READOUT function to aggregate the feature embedding of all layers of the node to obtain the final node feature representation.
- Calculate the predicted association score for each lncRNA-disease pair.

Our contributions are listed below:

- To the best of our knowledge, we are the first to apply GCNs with hierarchical refinement mechanisms to predict disease-associated lncRNAs.
- We compared our model with multiple other methods and verified the superiority of our model's performance.
- We validated the model on large and small datasets, respectively. Ablation experiments were also performed to verify the effectiveness of the hierarchical refinement mechanism.

2. Materials and methods

2.1. Dataset

In this research, the dataset1 employed was originated in a prior study conducted by Fu et al. [42], which aimed to predict lncRNA-disease associations. This dataset comprises 240 lncRNAs, 405 diseases and 495 miRNAs, along with 2687 known lncRNA-disease associations that have been validated in LncRNADisease, Lnc2Cancer, and GeneRIF [43] databases. In addition, the dataset1 also includes 13,559 known miRNA-disease associations from the HDMM and 1002 miRNA-lncRNA interactions from the starBase [44]. The dataset2 was from Lan et al. [45]. After data processing, the dataset included 573 lncRNAs, 46 diseases and 526 miRNAs. The dataset3 was obtained from Guo et al. [46], which contains 769 lncRNAs, 2062 diseases, and 1023 miRNAs. The details are shown in Table 1 (LDA–lncRNA-disease known associations, MDA–miRNA-disease known associations, LMI–lncRNA-miRNA known interactions).

Table 1. The details of the used datasets.

Datasets	lncRNAs	diseases	miRNAs	LDA	MDA	LMI
dataset1	240	405	495	2687	13,559	1002
dataset2	573	46	526	1013	660	308
dataset3	769	2062	1023	1264	16,427	8374

2.1.1. Semantic similarity of diseases

Diseases' semantic similarity is determined by utilizing directed acyclic graphs (DAGs) related to diseases in the MeSH database according to Wang's method [47]. The calculation of semantic similarity between two diseases is performed by taking diseases A and B as examples. The DAG of A is represented as $DAG_A = (A, T_A, E_A)$, with T_A representing the set of all ancestral nodes of A , including node A , while E_A encompasses the edges connecting these nodes. The contribution of disease d to A within DAG_A is characterized by its D value concerning A , which is symbolized as $D_A(d)$, as shown in Eq (2.1).

$$\begin{cases} D_A(d) = 1, & \text{if } d = A; \\ D_A(d) = \max[\Delta * D_A(d') \mid d' \in d], & \text{otherwise,} \end{cases} \quad (2.1)$$

where Δ is the semantic contribution factor of the edge E_A that connects disease d to its child d' . Based on the above formula, the representation of the semantic value for A is expressed as $SV(A)$.

$$SV(A) = \sum_{d \in T_A} D_A(d). \quad (2.2)$$

Assuming that diseases sharing significant portions of the DAG are more likely to demonstrate heightened semantic similarity, we use the relative positions of the two diseases' DAG in the MeSH database to measure the semantic similarity between them, and the calculation formula is defined as:

$$DSS(A, B) = \frac{\sum_{d \in T_A \cap T_B} (D_A(d) + D_B(d))}{SV(A) + SV(B)}. \quad (2.3)$$

2.1.2. Functional similarity of lncRNAs

Within this study, we have employed an approach akin to the LFSCM model introduced by Chen et al. [48] to calculate lncRNAs' functional similarity. First, disease semantic similarity is determined through the utilization of both disease DAGs and disease MeSH descriptors. Second, the calculation of miRNA functional similarity relies on the interplay between disease semantic similarity and disease-miRNA associations. In conclusion, the lncRNA's functional similarity is determined by considering miRNA functional similarity and lncRNA-miRNA interaction, and we use *LFS* to represent the matrix of function similarity for lncRNAs.

2.1.3. GIP similarity of diseases and lncRNAs

GIP (gaussian interaction profile) similarity takes full account of the graph's topological structure information, and the approach assumes that diseases with similar characteristics are inclined to be associated with lncRNAs that have related functions and vice versa [49]. We consider $I(d(u))$ as the representation of disease $d(u)$'s interaction profile with all lncRNAs, which aligns with the u th column in the lncRNA-disease correlation matrix LD . As for $d(u)$ and $d(v)$, the GIP similarity matrix KD is formulated as:

$$KD(d(u), d(v)) = \exp\left(-\alpha_d \|I(d(u)) - I(d(v))\|^2\right), \quad (2.4)$$

where the parameter α_d is utilized to fine-tune the kernel bandwidth, and it is derived by normalizing the original parameter α'_d with the following equation:

$$\alpha_d = \alpha'_d / \left(\frac{1}{n_d} \sum_{i=1}^{n_d} \|I(d(i))\|^2 \right). \quad (2.5)$$

Likewise, we calculate the GIP similarity matrix KL of lncRNAs using the ensuing expressions:

$$KL(l(u), l(v)) = \exp\left(-\alpha_l \|I(l(u)) - I(l(v))\|^2\right), \quad (2.6)$$

$$\alpha_l = \alpha'_l / \left(\frac{1}{n_l} \sum_{j=1}^{n_l} \|I(l(j))\|^2 \right). \quad (2.7)$$

Here, the vector $I(l(u))$ (or $I(l(v))$) is the u th row (v th row) in the lncRNA-disease association matrix LD .

2.1.4. Fusion of similarity

We first standardize the various similarity matrices so that they have the same absolute scale [50], then the fusion similarity matrices for lncRNA, miRNA, and disease are represented using the adjacency matrices: $SML \in R^{(n_l \times n_l)}$, $SMM \in R^{(n_m \times n_m)}$, and $SMD \in R^{(n_d \times n_d)}$, respectively. The lncRNA similarity fusion strategy is as follows:

$$SML_{i,j} = \begin{cases} LFS_{i,j}, & \text{if } LFS_{i,j} \neq 0; \\ KL_{i,j}, & \text{others,} \end{cases} \quad (2.8)$$

where *LFS* denotes the lncRNA functional similarity matrix.

Analogously, the strategy for merging disease similarity is as outlined below:

$$SMD_{i,j} = \begin{cases} DSS_{i,j}, & \text{if } DSS_{i,j} \neq 0; \\ KD_{i,j}, & \text{others,} \end{cases} \quad (2.9)$$

where DSS denotes the semantic similarity matrix of the disease.

2.2. Methods

We use $G_l = (L, D, E_{ld})$ to denote the known associations graph of the lncRNA-disease. The notation used here includes sets and edges, with L denoting the set of nodes containing n_l lncRNAs $\{l_1, l_2, \dots, l_{n_l}\}$, D representing the set of nodes containing n_d diseases $\{d_1, d_2, \dots, d_{n_d}\}$, and E_{ld} denoting the set of edges corresponding to lncRNA-disease interactions $\{(l_i, d_j)\}$. Similarly, we use $G_m = (M, D, E_{md})$ and $G_{lm} = (L, M, E_{lm})$ to denote the miRNA-disease known associations graph and lncRNA-miRNA known interactions graph, respectively. Their associations can be represented by the adjacency matrices $LD \in \mathbb{R}^{n_l \times n_d}$, $MD \in \mathbb{R}^{n_m \times n_d}$ and $LM \in \mathbb{R}^{n_l \times n_m}$, respectively. If a connection exists between nodes of two distinct types, the element value at the corresponding position in the adjacency matrix is set to 1, otherwise, it is set to 0. We then integrate disease semantic similarity, lncRNA functional similarity, and GIP kernel similarity between diseases and lncRNAs as feature representations of corresponding nodes.

2.2.1. Construction of heterogeneous network graph

The incorporation of heterogeneous information attributes for lncRNAs and diseases effectively enhances the model's accuracy. Based on the previously mentioned association matrices LD , MD , and LM , along with the calculated similarity matrices SML , SMM and SMD , we construct a heterogeneous network encompassing lncRNA, miRNA and disease, comprising three distinct node types (lncRNA, miRNA and disease) and two categories of connections (edges connecting homologous nodes and edges linking heterogeneous nodes). We use G_{lmd} to represent the graph structure of this heterogeneous network, and the adjacency matrix A of G_{lmd} can be expressed as:

$$A = \begin{pmatrix} SML & LD & LM \\ LD^T & SMD & MD^T \\ LM^T & MD & SMM \end{pmatrix}, \quad (2.10)$$

where LD^T , LM^T , and MD^T denote the transpose of LD , LM , and MD matrices, respectively.

2.2.2. Traditional GCN

The concept of GCN was introduced in 2017 by Kipf et al. [51]. It is a deep learning model tailored for graph data, capable of acquiring lncRNA and disease node representations by leveraging information from the graph structure data [52]. The embedding for the $(l + 1)$ th layer, denoted as $X^{(l+1)}$, is expressed as follows:

$$X^{(l+1)} = \sigma(\tilde{A}X^lW^l), \quad (2.11)$$

$$\tilde{A} = \tilde{D}^{1/2}(A + E)\tilde{D}^{1/2}, \quad (2.12)$$

where $\sigma(\cdot)$ is a nonlinear activation function. E is a homotopy unity matrix to the adjacency matrix A . \tilde{D} stands as the degree matrix for $(A + E)$. \tilde{A} is called the regularization matrix of A .

The existing GCN-based models for predicting lncRNA-disease associations acquire the ultimate node embeddings through message passing and aggregation at each layer of CNN, and then solve the downstream association prediction tasks based on the final nodes embedding.

2.2.3. Hierarchical refinement-based GCN

Although the lncRNA-disease association prediction methods based on traditional GCN can achieve certain results, it faces many problems. For example, the learned node features are not accurate enough and relatively simple, and when the number of network layers is stacked more, it leads to over-smoothing of the model [53]. Therefore, we introduce an association prediction model between lncRNAs and diseases based on HRGCNLDA. The details of HRGCNLDA are as follows:

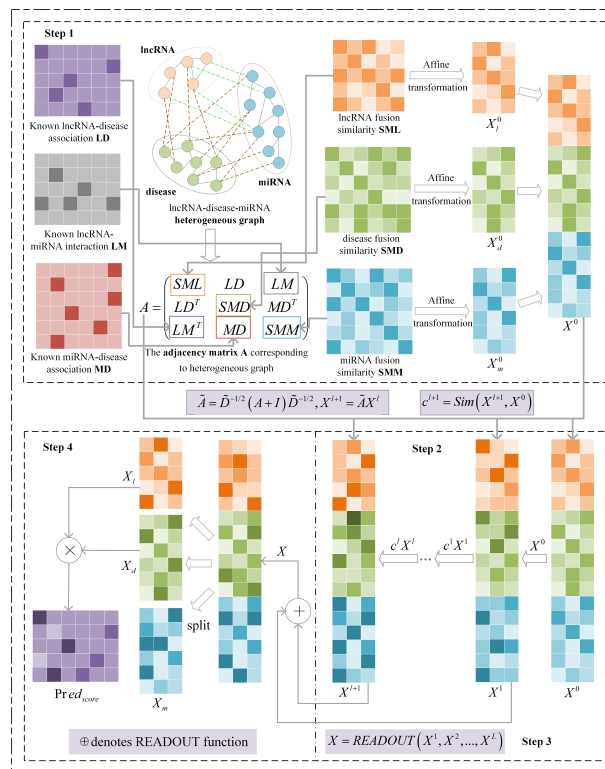


Figure 1. flowchart of HRGCNLDA. Step 1. Apply affine transformations to the initial attributes related to lncRNAs and diseases to align their initial feature dimensions; Step 2. Refine the aggregated feature embeddings for each layer by applying HRGCNLDA; Step 3. Select the appropriate READOUT function; Step 4. Calculate the predicted association score.

2.2.4. Step1: Generate ego embeddings using an affine transformation

Due to the disparity in the counts of nodes for lncRNA, diseases and miRNAs, the dimensionality of node features is also different. Therefore, we align the nodes feature dimension of lncRNA, disease, and miRNA and construct ego embeddings. The affine transformation formula is given by:

$$\begin{cases} X_l^0 = SML \cdot W_1 + b_1, \\ X_d^0 = SMD \cdot W_2 + b_2, \\ X_m^0 = SMM \cdot W_3 + b_3, \end{cases} \quad (2.13)$$

where (\cdot) denotes matrix multiplication, $W_1 \in R^{n_l \times d}$, $W_2 \in R^{n_d \times d}$, and $W_3 \in R^{n_m \times d}$ are trainable weight matrices, d is the affine transformed dimension, and b_1 , b_2 , and b_3 are trainable bias parameters.

Eventually, the ego embeddings $X^0 \in R^{(n_l+n_d+n_m) \times d}$ used for training are constructed.

$$X^0 = [X_l^0; X_d^0; X_m^0]. \quad (2.14)$$

In the ego embeddings, not only are the lncRNA and disease nodes' attributes included, but we also add the node feature of miRNA that is closely related to lncRNA and disease. This operation makes our node feature information richer and also enables our model to learn more complex feature representations.

2.2.5. Step2: Hierarchical refinement of aggregated feature embedding

As traditional GCNs suffer from over-smoothing when stacking more network layers, Chen et al. [54] mitigated this phenomenon by simulating a jump connection in ResNet that combines the smoothed feature embedding in the $(l + 1)$ th layer with the feature embedding in the ego layer. Inspired by this, we propose a hierarchical refinement of GCN that can dynamically extract feature information of nodes from the ego layer during the message propagation phase. The message propagation equation is defined as:

$$X^{(l+1)} = \tilde{A}X^lW, \quad (2.15)$$

$$c^{(l+1)} = Sim(X^{(l+1)}, X^0), \quad (2.16)$$

$$X^{(l+1)} = (c^{(l+1)} + \varepsilon)X^{l+1}, \quad (2.17)$$

where $X^{(l+1)}$ denotes the feature embedding at $(l + 1)$ th layer, $Sim(\cdot)$ denotes the similarity function, $c^{(l+1)} \in R^{n_l+n_d+n_m}$ denotes the cosine similarity vector between the feature embedding at $(l + 1)$ th layer and the feature embedding at the ego layer, and ε is a very small, isotype vector with $c^{(l+1)}$, in order to prevent $c^{(l+1)}$ from being a 0 vector.

In this process, we use the similarity vector between the feature embedding of the $(l + 1)$ th layer and the ego layer as the weight of the contribution of the $(l + 1)$ th layer to the final feature embedding. This amplifies the contribution of layers that are more similar to the ego layer and shrinks the contribution of the more different.

2.2.6. Step3: Select the READOUT function

The contribution of each hidden layer to the final feature embeddings has been taken into account by incorporating ego embeddings in our calculations. Consequently, the computation of the final feature embedding excludes consideration of the ego layer, which is defined as follows:

$$X = READOUT(X^1, X^2, \dots, X^L), \quad (2.18)$$

where $READOUT(\cdot)$ indicates the average aggregation function and L means the total number of layers.

2.2.7. Step4: Predict lncRNA-disease pair association scores

In the previous step, we obtain the final nodal feature embedding X , which contains all the nodes. We extract only the feature embeddings X_l and X_d for lncRNA and disease. Finally, the predicted

association score matrix for the lncRNA-disease pair is given by:

$$Pred_{score} = X_l \cdot X_d^T, \quad (2.19)$$

where X_d^T denotes the transpose of X_d and (\cdot) represents the matrix multiplication.

3. Results and discussion

In this section, five-fold cross-validation (5-CV) is applied to assess the effectiveness of HRGC-NLDA regarding the potential relationships between lncRNA-disease pairs. We compare HRGCNLDA with other methods on some metrics and also implement ablation experiments to confirm the effectiveness of hierarchical refinement. Finally, case studies further validate the model's reliability.

3.1. Performance evaluation

To validate the accuracy of the HRGCNLDA model in the association prediction task, we implement the 5-CV experiment on the model. The results are shown in Figures 2–4. Concretely, all known positive instances and an equivalent quantity of randomly chosen nonpositive instances are randomly divided into five groups. One of these groups is reserved for testing, while the four others are allocated for the training phase. To enhance the accuracy and reduce HRGCNLDA's computational complexity, we repeat 400 epochs for each fold experiment and set the network with 4 layers. Based on this experimental setup, the accuracy of HRGCNLDA was evaluated on six metrics: AUC, accuracy, precision, recall, F1-score, and AUPR, respectively. The corresponding values are 0.9287, 0.8409, 0.8239, 0.8693, 0.8453 and 0.9333; 0.8854, 0.7769, 0.7270, 0.8924, 0.7999, and 0.8841 and 0.9708, 0.9181, 0.9246, 0.9114, 0.9174, and 0.9679 on dataset 1, dataset 2 and dataset 3, respectively, which illustrate the accuracy of HRGCNLDA on prediction task (see Tables 2–4).

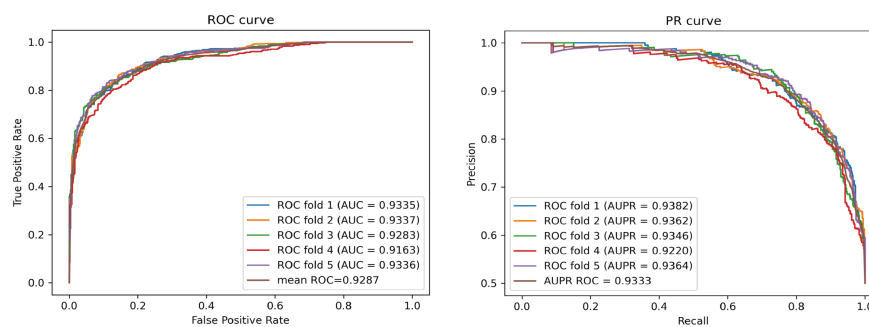


Figure 2. Performance of HRGCNLDA in lncRNA-disease associations prediction under 5-CV on dataset1.

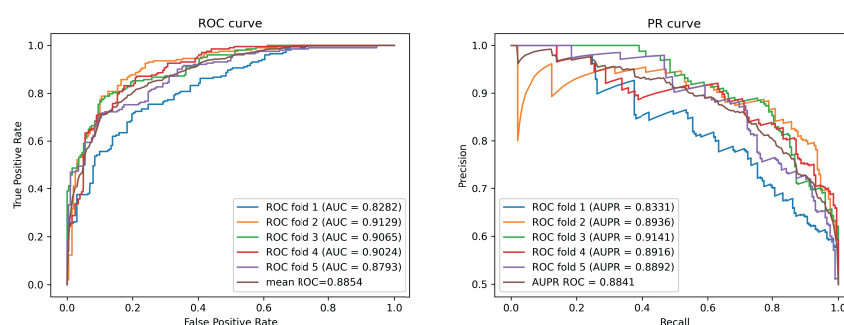


Figure 3. Performance of HRGCNLDA in lncRNA-disease associations prediction under 5-CV on dataset2.

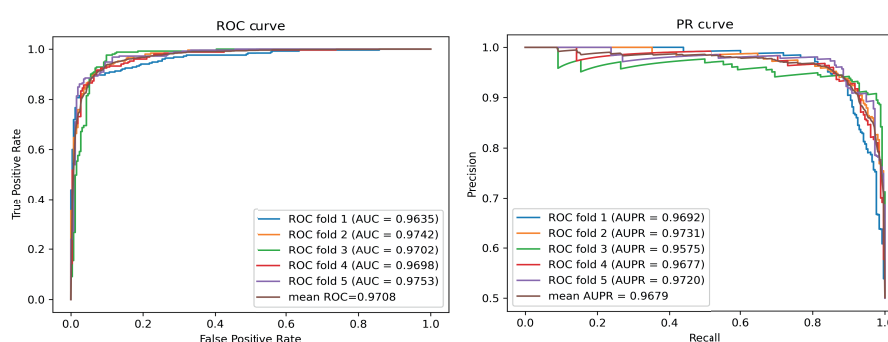


Figure 4. Performance of HRGCNLDA in lncRNA-disease associations prediction under 5-CV on dataset3.

Table 2. Comparison of method performances assessed using 5-CV on dataset1.

models	AUC	Accuracy	Precision	Recall	F1-score	AUPR
HRGCNLDA	0.9287	0.8409	0.8239	0.8693	0.8453	0.9333
KATZLDA	0.9047	0.8524	0.9517	0.1208	0.2136	0.8080
SIMCLDA	0.9073	0.7551	0.7005	0.9611	0.8104	0.9008
DMFLDA	0.8222	0.8141	0.8634	0.7460	0.8003	0.8878
GAMCLDA	0.9112	0.3915	0.3915	0.9132	0.5553	0.8096
IPCARFLDA	0.9067	0.7912	0.8946	0.6602	0.7596	0.9028

* Bolded values are the highest values, and the same applies to the following.

Table 3. Comparison of method performances assessed using 5-CV on dataset2.

models	AUC	Accuracy	Precision	Recall	F1-score	AUPR
HRGCNLDA	0.8854	0.7769	0.7270	0.8924	0.7999	0.8841
KATZLDA	0.6805	0.8705	0.9844	0.2265	0.3669	0.5758
SIMCLDA	0.8423	0.7305	0.8542	0.5558	0.8403	0.8403
DMFLDA	0.6838	0.7700	0.7606	0.7868	0.7729	0.8122
GAMCLDA	0.8730	0.4592	0.8749	0.4592	0.5898	0.7681
IPCARFLDA	0.7412	0.6757	0.8194	0.4522	0.5818	0.7434

Table 4. Comparison of method performances assessed using 5-CV on dataset3.

models	AUC	Accuracy	Precision	Recall	F1-score	AUPR
HRGCNLDA	0.9708	0.9181	0.9246	0.9114	0.9174	0.9679
KATZLDA	0.7372	0.8585	0.9674	0.1555	0.2677	0.8350
SIMCLDA	0.6605	0.6218	0.5803	0.8806	0.6995	0.7774
DMFLDA	0.5528	0.5012	0.6000	0.0024	0.0048	0.7506
GAMCLDA	0.8850	0.8299	0.8299	0.8892	0.8231	0.8979
IPCARFLDA	0.9305	0.8275	0.8785	0.7603	0.8139	0.8811

3.2. Comparison with other methods

For a further evaluation of HRGCNLDA's accuracy when predicting potential associations between lncRNAs and diseases, we also perform contrast experiments on five new methods, which are KATZLDA [55], SIMCLDA [56], GAMCLDA [57], DMFLDA [58] and IPCARF [59]. The central ideas and characteristics of the contrast methods are summarized as follows:

- **KATZLDA:** The lncRNA-disease associations were predicted using the KATZ metric, which aims to predict associations between diseases and lncRNAs, particularly in cases where there are no previously established associations. However, the KATZ metric exhibits limitations in terms of prediction accuracy.
- **SIMCLDA:** Initially, it employs principal component analysis to extract distinct principal feature vectors for both lncRNAs and diseases. Subsequently, when dealing with a new lncRNA or disease, it utilizes its neighbors' interaction profiles instead. Finally, inductive matrix completion is employed for predicting lncRNA-disease associations.
- **GAMCLDA:** For learning the latent feature vectors for both lncRNAs and diseases, the method initially employs a GCN for capturing local graph structures and node attributes. Utilized as a decoder, the dot product of feature vectors for lncRNAs and diseases reconstructs the association matrix of lncRNA and diseases.
- **DMFLDA:** The model utilizes a sequence of nonlinear hidden layers to acquire potential feature representations of nodes for lncRNAs and diseases, and is therefore capable of learning more complex, nonlinear connections between nodes.
- **IPCARF:** The prediction of lncRNA-disease associations is achieved by a synergistic combination of incremental principal component analysis and the RF algorithm.

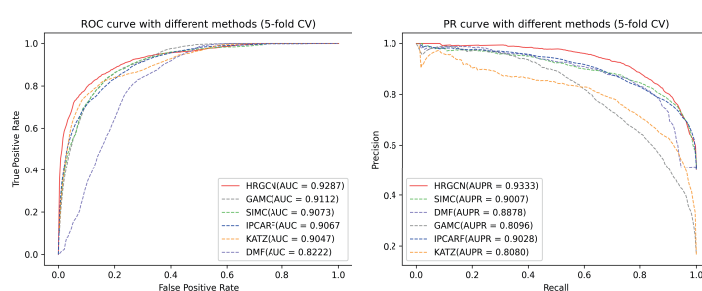


Figure 5. Performance comparison of HRGCNLDA and alternative methods for predicting lncRNA-disease associations under 5-CV on dataset1.

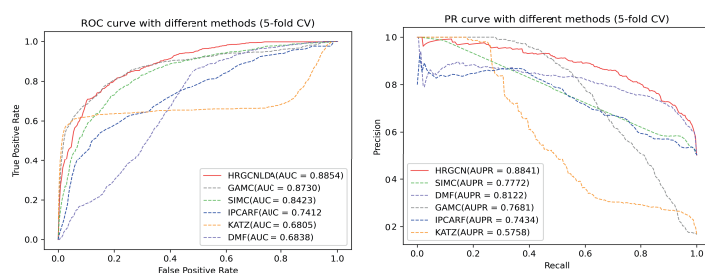


Figure 6. Performance comparison of HRGCNLDA and alternative methods for predicting lncRNA-disease associations under 5-CV on dataset2.

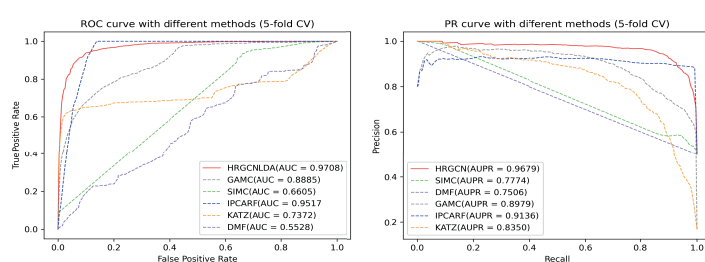


Figure 7. Performance comparison of HRGCNLDA and alternative methods for predicting lncRNA-disease associations under 5-CV on dataset3.

The experimental results on the 5-CV indicate that HRGCNLDA exhibits the best performance (see Figures 5–7).

3.3. Ablation experiment

To verify the efficiency of HRGCNLDA, we execute an ablation experiment by removing the hierarchical refinement component of HRGCNLDA and keeping other experimental settings unchanged. The final AUC obtained is 0.8691 and the AUPR is 0.8298. Compared with HRGCNLDA, the AUC and AUPR are reduced by 0.0596 and 0.1035, respectively (see Figure 8), and this experimental result further proves that the hierarchical refinement mechanism effectively improves the prediction accuracy by refining the layer embedding. In addition, we excluded the relevant miRNA information and conducted experiments (see Figure 9 and Table 5). The experimental results indicate that the enhancement of predictive performance by miRNA is of significant importance.

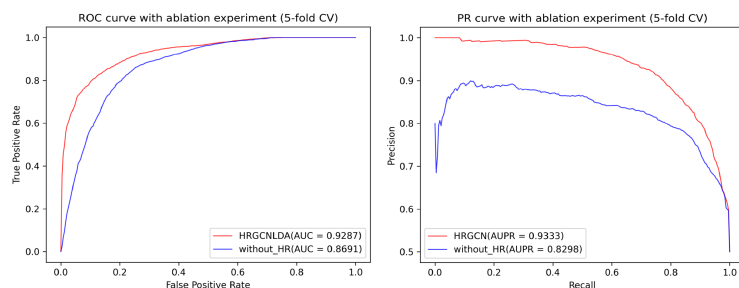


Figure 8. Performance comparison of ablation experiments and HRGCNLDA on dataset1.

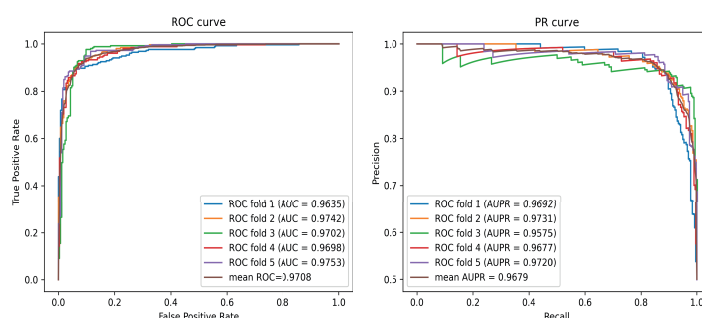


Figure 9. Performance of HRGCNLDA in lncRNA-disease associations prediction under 5-CV on dataset1 (without miRNA).

Table 5. Comparison of method performances assessed using 5-CV on dataset1.

models	AUC	Accuracy	Precision	Recall	F1-score	AUPR
HRGCNLDA	0.9287	0.8409	0.8239	0.8693	0.8453	0.9333
without miRNA	0.8929	0.8122	0.7852	0.8647	0.8223	0.9099

3.4. Effect of number of layers

We maintain consistent values for other parameters in a 5-CV setup and discuss how the count of layers influences the predictive result of the HRGCNLDA model. To be specific, we configure the layer count at 2, 3, 4 and 5, and the results are depicted in Figure 10. As illustrated in the figure, the model's AUC and AUPR values exhibit a rise in correspondence with the increment in the number of layers and reach the maximum at *layers* = 4, and then start to decrease. Therefore, we finally set the number of layers to 4.

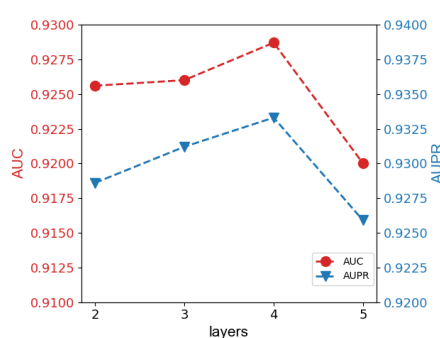


Figure 10. AUC and AUPR values for different layers on dataset1.

3.5. Case studies

To thoroughly assess the HRGCNLDA model's efficacy in predicting novel disease-associated lncRNAs, we conduct a case study on three prevalent and significant human diseases, namely, breast cancer, lung cancer and gastric cancer.

To be precise, we extract all positive samples (2687 known associations) from the association matrix of lncRNA-disease and randomly select a balanced set of nonpositive samples, then construct a training dataset for HRGCNLDA model training and finally predict the latent associations between lncRNAs and diseases to obtain the prediction score matrix. We screen the top 10 lncRNAs from the prediction score matrix for potential associations with specific diseases and identify the confirmed associations by retrieving them from published literature or publicly available databases (see Table 6).

Table 6. HRGCNLDA identified the top 10 novel lncRNAs related to breast cancer, lung cancer and gastric cancer.

Breast cancer		Lung cancer		Gastric cancer	
lncRNA	evidence	lncRNA	evidence	lncRNA	evidence
MIR17HG	PMID: 36943627	TUSC7	LncRNADisease v2.0	MALAT1	LncRNADisease v2.0
BANCR	Lnc2cancer v3.0	HOTTIP	LncRNADisease v2.0	XIST	LncRNADisease v2.0
SNHG1	Lnc2cancer v3.0	PCA3	PMID: 32388776	AFAP1-AS1	LncRNADisease v2.0
TUSC7	PMID: 35296964	SCHLAP1	Unconfirmed	NEAT1	LncRNADisease v2.0
PCA3	Unconfirmed	KCNQ1OT1	LncRNADisease v2.0	HOTTIP	LncRNADisease v2.0
HYMAI	Unconfirmed	TP53COR1	LncRNADisease v2.0	PCA3	Unconfirmed
CASC2	LncRNADisease v2.0	PRNCR1	Lnc2cancer v3.0	BCYRN1	LncRNADisease v2.0
PRNCR1	Lnc2cancer v3.0	ZEB1-AS1	Lnc2cancer v3.0	MIR17HG	Unconfirmed
TP53COR1	LncRNADisease v2.0	HYMAI	Unconfirmed	MIR124-2HG	Unconfirmed
PRINS	Unconfirmed	HCG9	PMID: 31576252	SCHLAP1	Unconfirmed

It begins with a case study on breast cancer, a prevalent cause of mortality among women, whose early diagnosis is crucial in preventing the disease, yet it remains challenging in the world [60]. Numerous experimental studies have demonstrated the abnormal expression of lncRNAs in breast cancer, signifying their strong association with the disease's progression [61]. By training and prediction of the HRGCNLDA model, we identify candidate lncRNAs linked to breast cancer and subsequently narrow down our selection to the top 10 lncRNAs. The analysis revealed that 7 out of the leading 10 candidate lncRNAs linked to breast cancer are validated by the publicly available datasets LncRNADisease v2.0, Lnc2cancer v3.0, or published literatures. Of these, among the leading 5 candidate lncRNAs, 4 of them have been confirmed to be closely related to breast cancer. For example, the lncRNA MIR17HG can potentially suppress the growth and movement of breast cancer cells by acting as a sponge for miR-454-3p through a ceRNA (competing endogenous RNA) mechanism, suggesting that targeting MIR17HG is a viable approach in the search for therapeutic candidates for screening breast cancer [62]. The lncRNA BANCR exhibited a significant increase in expression within breast cancer tissues when compared to their normal counterparts. After suppressing BANCR in MCF-7 breast cancer cells, there was a significant inhibition in cell proliferation and colony formation ability. Further studies demonstrated that this inhibition of BANCR promoted apoptosis in MCF-7 cells [63].

The second disease is lung cancer, a prominent contributor to cancer-related fatalities on a global scale, accounting for about 5.2% of all cancer deaths [64]. In the case study, we obtain the ranking of candidate lncRNAs related to lung cancer. Among the top 10, 8, among the top 5, 4 lncRNAs are proven

to be correlated with the development of lung cancer. For example, lncRNA TUSC7 exhibits reduced expression levels in cases of non-small cell lung carcinoma (NSCL) tissues and lung cancer cells than in normal cells, and upregulation of TUSC7 expression levels inhibited lung cancer cell proliferation in vitro [65]. The lncRNA HOTTIP exhibited a notable increase in expression within lung cancer cells. Furthermore, conclusive data suggests that HOTTIP is a pivotal mediator in fostering lung cancer progression, both in experimental models and clinical samples, and promotes cell cycle processes and inhibits apoptosis in lung tumor cells [66].

The final analysis pertains to gastric cancer, which held the fifth position among the most common malignancies in 2020 with about 1.1 million new infections, ranking as the world's fourth most common cause of cancer-related fatalities [67]. An expanding body of experiments has shown that lncRNA's significance in gastric cancer treatment is pivotal. By prediction of the HRGCNLDA model, the ranking of all candidate lncRNAs linked to gastric cancer is acquired. The analysis revealed that among the top 10 potential lncRNAs, 6 lncRNAs are verified to be relevant to gastric cancer, and among the top 5 candidate lncRNAs, all of them are confirmed to have some correlation with gastric cancer. For example, the lncRNA MALAT1 exhibited abnormally high expression levels in gastric cancer cell lines. In SGC-7901 cells, suppressing MALAT1 resulted in substantial cell cycle arrest in the G0 phase as well as a notably suppressed cell proliferation [68]. In gastric cancer tissues and cell lines, lncRNA XIST exhibited a pronounced upregulation. Specifically, XIST is implicated in cell cycle progression from the G1 phase to the S phase, as well as in protecting cells from apoptosis and promoting the growth of gastric cancer cells [69].

4. Conclusions

An increasing body of research has indicated that lncRNAs are instrumental in disease prevention, development, treatment, and prognosis, especially for cancer. Although numerous computational models developed by experts and scholars have been put forth to aid in discerning the lncRNA-disease associations, the GCN-based method for lncRNA-disease association prediction has some shortcomings. For instance, the learned node features are fewer and simpler and there is an over-smoothing problem when the number of network layers is stacked high. To address these issues, we introduce HRGCNLDA, a computational model for predicting lncRNA-disease associations using a hierarchical refinement GCN. The main merit of HRGCNLDA is that the similarity between each layer and ego layer is fully considered in the message propagation process, so that the weights of the layers that are more similar to the ego layer can be enhanced and the different can be reduced, thus preventing the model from being over-smoothing. What's more, a heterogeneous network graph of lncRNA-disease-miRNA was constructed to enrich the feature information of lncRNA and disease nodes [70].

In this paper, we propose a method for predicting lncRNA-disease associations using a GCN, which we abbreviate as HRGCNLDA. The method effectively tackles the model over-smoothing problem, and miRNA nodes information are incorporated into the heterogeneous network graph to enrich the feature information of nodes. Ultimately, our model attains an AUC value of 0.9287 and an AUPR value of 0.9333, respectively, on a 5-CV. Moreover, we also do ablation experiments by removing the hierarchical refinement component from HRGCNLDA and considering only the prediction performance of a normal GCN. The final AUC achieved is solely 0.8691, while the AUPR solely reaches 0.8298, and these results are specifically obtained through 5-CV, respectively, which further illustrates the accuracy

of our model. Finally, we also perform case studies on the whole dataset. Among the lncRNAs in the top 10 candidates potentially related to breast, lung, and gastric cancers, respectively, 7, 8, and 6 lncRNAs were confirmed by public datasets or published literature, and among the top 5 candidate lncRNAs, 4, 4, and 5 are confirmed, respectively, which sufficiently illustrates the validity of HRGCNLDA.

The HRGCNLDA model, despite its superior predictive performance regarding latent lncRNA-disease associations, still exhibits certain limitations. Our heterogeneous network graph comprises only three types of nodes, whereas the biological information exhibits extensive diversity. In future endeavors, incorporating additional bioinformatics sources may yield enhanced performance. Furthermore, acquiring the latest data from recently updated biological experiments can enhance the accuracy of model predictions.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (62172158), Natural Science Foundation of Hunan province (2023JJ30264), and Scientific Research Project of Hunan Education Department (22A0350, 22A0101).

We also extend our gratitude to the anonymous reviewers for their invaluable feedback and insightful suggestions, which have greatly contributed to the enhancement of this paper.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Y. J. Chi, D. Wang, J. P. Wang, W. D. Yu, J. C. Yang, Long non-coding rna in the pathogenesis of cancers, *Cells*, **8** (2019), 1015. <https://doi.org/10.3390/cells8091015>
2. S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, et al., Landscape of transcription in human cells, *Nature*, **489** (2012), 101–108. <https://doi.org/10.1038/nature11233>
3. A. T. Willingham, A. P. Orth, S. Batalov, E. C. Peters, B. G. Wen, P. Aza-Blanc, et al., A strategy for probing the function of noncoding rnas finds a repressor of nfat, *Science*, **309** (2005), 1570–1573. <https://doi.org/10.1126/science.1115901>
4. C. Xing, S. G. Sun, Z. Q. Yue, F. Bai, Role of lncrna lucat1 in cancer, *Biomed. Pharmacother.*, **134** (2021), 111158. <https://doi.org/10.1016/j.biopha.2020.111158>
5. L. Peng, M. Peng, B. Liao, G. H. Huang, W. B. Li, D. F. Xie, The advances and challenges of deep learning application in biological big data processing, *Curr. Bioinf.*, **13** (2018), 352–359. https://doi.org/10.1163/9789004392533_041

6. R. H. Wang, Y. Jiang, J. R. Jin, C. L. Yin, H. Q. Yu, F. S. Wang, et al., Deepbio: an automated and interpretable deep-learning platform for high-throughput biological sequence prediction, functional annotation and visualization analysis, *Nucleic Acids Res.*, **51** (2023), 3017–3029. <https://doi.org/10.1093/nar/gkad055>
7. L. H. Peng, J. W. Tan, W. Xiong, L. Zhang, Z. Wang, R. Y. Yuan, et al., Deciphering ligand-receptor-mediated intercellular communication based on ensemble deep learning and the joint scoring strategy from single-cell transcriptomic data, *Comput. Biol. Med.*, **163** (2023), 107137. <https://doi.org/10.1016/j.compbiomed.2023.107137>
8. W. Liu, Y. Yang, X. Lu, X. Z. Fu, R. Q. Sun, L. Yang, et al., Nsrgrn: a network structure refinement method for gene regulatory network inference, *Briefings Bioinf.*, **24** (2023), bbad129. <https://doi.org/10.1093/bib/bbad129>
9. J. C. Wang, Y. J. Chen, Q. Zou, Inferring gene regulatory network from single-cell transcriptomes with graph autoencoder model, *PLoS Genet.*, **19** (2023), e1010942. <https://doi.org/10.1371/journal.pgen.1010942>
10. L. Peng, C. Yang, L. Huang, X. Chen, X. Z. Fu, W. Liu, Rnmflp: predicting circrna-disease associations based on robust nonnegative matrix factorization and label propagation, *Briefings Bioinf.*, **23** (2022), bbac155. <https://doi.org/10.1093/bib/bbac155>
11. W. Liu, T. T. Tang, X. Lu, X. Z. Fu, Y. Yang, L. Peng, Mpclca: predicting circrna-disease associations by using automatically selected meta-path and contrastive learning, *Briefings Bioinf.*, **24** (2023), bbad227. <https://doi.org/10.1093/bib/bbad227>
12. L. L. Zhuo, S. Y. Pan, J. Li, X. Z. Fu, Predicting mirna-lncrna interactions on plant datasets based on bipartite network embedding method, *Methods*, **207** (2022), 97–102. <https://doi.org/10.1016/j.ymeth.2022.09.002>
13. Z. C. Zhou, Z. Y. Du, J. H. Wei, L. L. Zhuo, S. Y. Pan, X. Z. Fu, et al., Mham-npi: Predicting ncRNA-protein interactions based on multi-head attention mechanism, *Comput. Biol. Med.*, **163** (2023), 107143. <https://doi.org/10.1016/j.compbiomed.2023.107143>
14. X. Chen, L. Wang, J. Qu, N. N. Guan, J. Q. Li, Predicting mirna-disease association based on inductive matrix completion, *Bioinformatics*, **34** (2018), 4256–4265. <https://doi.org/10.1093/bioinformatics/bty503>
15. X. Chen, D. Xie, L. Wang, Q. Zhao, Z. H. You, H. Liu, Bnpmda: Bipartite network projection for mirna-disease association prediction, *Bioinformatics*, **34** (2018), 3178–3186. <https://doi.org/10.1093/bioinformatics/bty333>
16. L. Huang, L. Zhang, X. Chen, Updated review of advances in micrornas and complex diseases: towards systematic evaluation of computational models, *Briefings Bioinf.*, **23** (2022), bbac407. <https://doi.org/10.1093/bib/bbac407>
17. C. C. Wang, C. C. Zhu, X. Chen, Ensemble of kernel ridge regression-based small molecule-mirna association prediction in human disease, *Briefings Bioinf.*, **23** (2022), bbab431. <https://doi.org/10.1093/bib/bbab431>
18. Z. J. Li, Y. X. Zhang, Y. Bai, X. H. Xie, L. J. Zeng, Imc-mda: Prediction of mirna-disease association based on induction matrix completion, *Math. Biosci. Eng.*, **20** (2023), 10659–10674. <https://doi.org/10.3934/mbe.2023471>

19. Q. Qu, X. Chen, B. Ning, X. Zhang, H. Nie, L. Zeng, et al., Prediction of mirna-disease associations by neural network-based deep matrix factorization, *Methods*, **212** (2023), 1–9. <https://doi.org/10.1016/j.ymeth.2023.02.003>
20. L. Zhang, C. C. Wang, X. Chen, Predicting drug-target binding affinity through molecule representation block based on multi-head attention and skip connection, *Briefings Bioinf.*, **23** (2022), bbac468. <https://doi.org/10.1093/bib/bbac468>
21. L. Katusiime, Covid-19 and the effect of central bank intervention on exchange rate volatility in developing countries: The case of uganda, *National Accounting Rev.*, **5** (2023), 23–37. <https://doi.org/10.3934/NAR.2023002>
22. L. Grassini, Statistical features and economic impact of Covid-19, *National Accounting Rev.*, **5** (2023), 38–40. <https://doi.org/10.3934/NAR.2023003>
23. Z. Y. Bao, Z. Yang, Z. Huang, Y. R. Zhou, Q. H. Cui, D. Dong, Lncrnadisease 2.0: an updated database of long non-coding rna-associated diseases, *Nucleic Acids Res.*, **47** (2019), D1034–D1037. <https://doi.org/10.1093/nar/gky905>
24. S. W. Ning, J. Z. Zhang, P. Wang, H. Zhi, J. J. Wang, Y. Liu, et al., Lnc2cancer: a manually curated database of experimentally supported lncrnas associated with various human cancers, *Nucleic Acids Res.*, **44** (2016), D980–D985. <https://doi.org/10.1093/nar/gkv1094>
25. X. Chen, L. Huang, Computational model for disease research, *Briefings Bioinf.*, **24** (2023), bbac615. <https://doi.org/10.1093/bib/bbac615>
26. K. Albitar, K. Hussainey, Sustainability, environmental responsibility and innovation, *Green Finance*, **5** (2023), 85–88. <https://doi.org/10.3934/GF.2023004>
27. G. Desalegn, Insuring a greener future: How green insurance drives investment in sustainable projects in developing countries, *Green Finance*, **5** (2023), 195–210. <https://doi.org/10.3934/GF.2023008>
28. Y. Liang, Z. H. Zhang, N. N. Liu, Y. N. Wu, C. L. Gu, Y. L. Wang, Magcnse: predicting lncrna-disease associations using multi-view attention graph convolutional network and stacking ensemble model, *BMC Bioinf.*, **23** (2022). <https://doi.org/10.1186/s12859-022-04715-w>
29. Y. Kim, M. Lee, Deep learning approaches for lncrna-mediated mechanisms: A comprehensive review of recent developments, *Int. J. Mol. Sci.*, **24** (2023), 10299. <https://doi.org/10.3390/ijms241210299>
30. Z. Q. Zhang, J. L. Xu, Y. N. Wu, N. N. Liu, Y. L. Wang, Y. Liang, Capsnet-lda: predicting lncrna-disease associations using attention mechanism and capsule network based on multi-view data, *Briefings Bioinf.*, **24** (2022), bbac531. <https://doi.org/10.1093/bib/bbac531>
31. N. Dwarika, The risk-return relationship and volatility feedback in south africa: a comparative analysis of the parametric and nonparametric bayesian approach, *Quant. Finance Econ.*, **7** (2023), 119–146. <https://doi.org/10.3934/QFE.2023007>
32. N. Dwarika, Asset pricing models in south africa: A comparative of regression analysis and the bayesian approach, *Data Sci. Finance Econ.*, **3** (2023), 55–75. <https://doi.org/10.3934/DSFE.2023004>

33. Y. Q. Lin, X. J. Chen, H. Y. Lan, Analysis and prediction of american economy under different government policy based on stepwise regression and support vector machine modelling, *Data Sci. Finance Econ.*, **3** (2023), 1–13. <https://doi.org/10.3934/DSFE.2023001>
34. N. Sheng, L. Huang, Y. T. Lu, H. Wang, L. L. Yang, L. Gao, et al., Data resources and computational methods for lncrna-disease association prediction, *Comput. Biol. Med.*, **153** (2023), 106527. <https://doi.org/10.1016/j.compbiomed.2022.106527>
35. J. H. Wei, L. L. Zhuo, S. Y. Pan, X. Z. Lian, X. J. Yao, X. Z. Fu, Headtailtransfer: An efficient sampling method to improve the performance of graph neural network method in predicting sparse ncRNA-protein interactions, *Comput. Biol. Med.*, **157** (2023), 106783. <https://doi.org/10.1016/j.compbiomed.2023.106783>
36. P. Xuan, S. X. Pan, T. G. Zhang, Y. Liu, H. Sun, Graph convolutional network and convolutional neural network based method for predicting lncrna-disease associations, *Cells*, **8** (2019), 1012. <https://doi.org/10.3390/cells8091012>
37. M. F. Leung, A. Jawaid, S. W. Ip, C. H. Kwok, S. Yan, A portfolio recommendation system based on machine learning and big data analytics, *Data Sci. Finance Econ.*, **3** (2023), 152–165. <https://doi.org/10.3934/DSFE.2023009>
38. Q. W. Wu, J. F. Xia, J. C. Ni, C. H. Zheng, Gaerf: predicting lncrna-disease associations by graph auto-encoder and random forest, *Briefings Bioinf.*, **22** (2021), bbaa391. <https://doi.org/10.1093/bib/bbaa391>
39. N. Sheng, L. Huang, Y. Wang, J. Zhao, P. Xuan, L. Gao, et al., Multi-channel graph attention autoencoders for disease-related lncrnas prediction, *Briefings Bioinf.*, **23** (2022), bbab604. <https://doi.org/10.1093/bib/bbab604>
40. L. Peng, C. Yang, Y. F. Chen, W. Liu, Predicting circrna-disease associations via feature convolution learning with heterogeneous graph attention network, *IEEE J. Biomed. Health. Inf.*, **27** (2023), 3072–3082. <https://doi.org/10.1109/JBHI.2023.3260863>.
41. X. Liu, C. Z. Song, F. Huang, H. T. Fu, W. J. Xiao, W. Zhang, Graphcdr: a graph neural network method with contrastive learning for cancer drug response prediction, *Briefings Bioinf.*, **23** (2022), bbab457. <https://doi.org/10.1093/bib/bbab457>
42. G. Y. Fu, J. Wang, C. Domeniconi, G. X. Yu, Matrix factorization-based data fusion for the prediction of lncrna–disease associations, *Bioinformatics*, **34** (2018), 1529–1537. <https://doi.org/10.1093/bioinformatics/btx794>
43. Z. Y. Lu, K. B. Cohen, L. Hunter, Generif quality assurance as summary revision, in *Biocomputing 2007*, World Scientific, (2007), 269–280. https://doi.org/10.1142/9789812772435_0026
44. J. H. Li, S. Liu, H. Zhou, L. H. Qu, J. H. Yang, starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data, *Nucleic Acids Res.*, **42** (2014), D92–D97. <https://doi.org/10.1093/nar/gkt1248>
45. W. Lan, Y. Dong, Q. F. Chen, R. Q. Zheng, J. Liu, Y. Pan, et al., Kganeda: predicting circrna-disease associations based on knowledge graph attention network, *Briefings Bioinf.*, **23** (2022), bbab494. <https://doi.org/10.1093/bib/bbab494>

46. Z. H. Guo, Z. H. You, D. S. Huang, H. C. Yi, Z. H. Chen, Y. B. Wang, A learning based framework for diverse biomolecule relationship prediction in molecular association network, *Commun. Biol.*, **3** (2020). <https://doi.org/10.1038/s42003-020-0858-8>
47. D. Wang, J. Wang, M. Lu, F. Song, Q. H. Cui, Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases, *Bioinformatics*, **26** (2010), 1644–1650. <https://doi.org/10.1093/bioinformatics/btq241>
48. X. Chen, Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA, *Sci. Rep.*, **5** (2015), 13186. <https://doi.org/10.1038/srep13186>
49. X. Chen, G. Y. Yan, Novel human lncrna-disease association inference based on lncrna expression profiles, *Bioinformatics*, **29** (2013), 2617–2624. <https://doi.org/10.1093/bioinformatics/btt426>
50. D. Anderson, U. Ulrych, Accelerated american option pricing with deep neural networks, *Quant. Finance Econ.*, **7** (2023), 207–228. <https://doi.org/10.3934/QFE.2023011>
51. T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, preprint, arXiv:1609.02907. <https://doi.org/10.48550/arXiv.1609.02907>
52. L. Peng, Y. Tu, L. Huang, Y. Li, X. Z. Fu, X. Chen, Daestb: inferring associations of small molecule–mirna via a scalable tree boosting model based on deep autoencoder, *Briefings Bioinf.*, **23** (2022), bbac478. <https://doi.org/10.1093/bib/bbac478>
53. Z. Y. Chu, S. C. Liu, W. Zhang, Hierarchical graph representation learning for the prediction of drug-target binding affinity, *Inf. Sci.*, **613** (2022), 507–523. <https://doi.org/10.1016/j.ins.2022.09.043>
54. M. Chen, Z. W. Wei, Z. F. Huang, B. L. Ding, Y. L. Li, Simple and deep graph convolutional networks, in *Proceedings of the 37th International Conference on Machine Learning*, PMLR, (2020), 1725–1735.
55. X. Chen, Katzlda: Katz measure for the lncrna-disease association prediction, *Sci. Rep.*, **5** (2015), 16840. <https://doi.org/10.1038/srep16840>
56. C. Q. Lu, M. Y. Yang, F. Luo, F. X. Wu, M. Li, Y. Pan, et al., Prediction of lncrna–disease associations based on inductive matrix completion, *Bioinformatics*, **34** (2018), 3357–3364. <https://doi.org/10.1093/bioinformatics/bty327>
57. X. M. Wu, W. Lan, Q. F. Chen, Y. Dong, J. Liu, W. Peng, Inferring lncRNA-disease associations based on graph autoencoder matrix completion, *Comput. Biol. Chem.*, **87** (2020), 107282. <https://doi.org/10.1016/j.compbiolchem.2020.107282>
58. M. Zeng, C. Q. Lu, Z. H. Fei, F. X. Wu, Y. H. Li, J. X. Wang, et al., Dmflda: a deep learning framework for predicting lncrna–disease associations, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **18** (2020), 2353–2363. <https://doi.org/10.1109/TCBB.2020.2983958>
59. R. Zhu, Y. Wang, J. X. Liu, L. Y. Dai, Ipcarf: improving lncrna-disease association prediction using incremental principal component analysis feature selection and a random forest classifier, *BMC Bioinf.*, **22** (2021). <https://doi.org/10.1186/s12859-021-04104-9>
60. Y. S. Sun, Z. Zhao, Z. N. Yang, F. Xu, H. J. Lu, Z. Y. Zhu, et al., Risk factors and preventions of breast cancer, *Int. J. Biol. Sci.*, **13** (2017), 1387–1397. <https://doi.org/10.7150/ijbs.21635>

61. H. Jin, W. Du, W. T. Huang, J. J. Yan, Q. Tang, Y. B. Chen, et al., lncRNA and breast cancer: Progress from identifying mechanisms to challenges and opportunities of clinical treatment, *Mol. Ther.–Nucleic Acids*, **25** (2021), 613–637. <https://doi.org/10.1016/j.omtn.2021.08.005>
62. J. J. Xu, M. S. Hu, Y. Gao, Y. S. Wang, X. N. Yuan, Y. Yang, et al., Lncrna mir17hg suppresses breast cancer proliferation and migration as cerna to target fam135a by sponging mir-454-3p, *Mol. Biotechnol.*, **65** (2023), 2071–2085. <https://doi.org/10.1007/s12033-023-00706-1>
63. K. X. Lou, Z. H. Li, P. Wang, Z. Liu, Y. Chen, X. L. Wang, et al., Long non-coding rna bancr indicates poor prognosis for breast cancer and promotes cell proliferation and invasion, *Eur. Rev. Med. Pharmacol. Sci.*, **22** (2018), 1358–1365.
64. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, A. Jemal, Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: Cancer J. Clinicians*, **68** (2018), 394–424. <https://doi.org/10.3322/caac.21492>
65. Z. W. Wang, Y. Y. Jin, H. T. Ren, X. L. Ma, B. F. Wang, Y. L. Wang, Downregulation of the long non-coding RNA TUSC7 promotes NSCLC cell proliferation and correlates with poor prognosis, *Am. J. Transl. Res.*, **8** (2016), 680–687.
66. H. P. Deng, L. Chen, T. Fan, B. Zhang, Y. Xu, Q. Geng, Long non-coding rna hottip promotes tumor growth and inhibits cell apoptosis in lung cancer, *Cell. Mol. Biol.*, **61** (2015), 34–40.
67. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, et al., Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J. Clinicians*, **71** (2021), 209–249. <https://doi.org/10.3322/caac.21660>
68. J. Q. Wang, L. P. Su, X. H. Chen, P. Li, Q. Cai, B. Q. Yu, et al., MALAT1 promotes cell proliferation in gastric cancer by recruiting SF2/ASF, *Biomed. Pharmacother.*, **68** (2014), 557–564. <https://doi.org/10.1016/j.biopha.2014.04.007>
69. L. Ma, Y. J. Zhou, X. J. Luo, H. Gao, X. B. Deng, Y. J. Jiang, Long non-coding RNA XIST promotes cell growth and invasion through regulating miR-497/MACCC1 axis in gastric cancer, *Oncotarget*, **8** (2017), 4125–4135. <https://doi.org/10.18632/oncotarget.13670>
70. H. T. Fu, F. Huang, X. Liu, Y. Qiu, W. Zhang, Mvgn: data integration through multi-view graph convolutional network for predicting links in biomedical bipartite networks, *Bioinformatics*, **38** (2022), 426–434. <https://doi.org/10.1093/bioinformatics/btab651>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)