International Journal of Electronics Communication and Computer Engineering
Volume 15, Issue 5, ISSN (Online): 2249–071X

Predicting lncRNA-Disease Associations through Attention Graph Convolutional Autoencode

Xiaoqian Li[1], Shanwen Zhang[2*] and Shenao Yuan[3]
[1,2,3]Xijing University, Xi'an, China.
*Corresponding author email id: wjdw716@163.com

Date of publication (dd/mm/yyyy): 14/10/2024

***Abstract* – This thesis takes lncRNA, disease as the research object, and starts from the aspects of heterogeneous network construction, feature information extraction, and association prediction model establishment, respectively, and proposes an improved graph self-encoder prediction model (AGCELDA for short) for predicting disease-related lncRNAs, in response to the problem of incomplete modelling of association relationships in the existing lncRNA-disease association prediction model. The model Firstly, a heterogeneous graph is constructed using similar information of lncRNAs, diseases and known lncRNA-disease associations; subsequently, a graph encoder embedded with an attention mechanism is used to model the lncRNA-disease association relationship to obtain high-quality, low-dimensional representation vectors; and finally, the graph decoder is used to reconstruct the association relationship between lncRNAs and diseases for potential association prediction. Using this method can better capture the association relationship between nodes in the heterogeneous graph, thus effectively improving the performance of lncRNA-disease association prediction. In the experiments, the model was subjected to five-fold cross-validation and compared with other methods, and the experiments proved that the AUC accuracy of the AGCELDA method was relatively high. Finally, a case study was also conducted to demonstrate the capability of AGCELDA in identifying candidate lncRNAs that may be associated with diseases.**

***Keywords* – lncRNA, Disease, Association Prediction, Deep Autocoder.**

I. INTRODUCTION

With the development of biotechnology and the accumulation of related theories, a large amount of evidence suggests that mutations and dysregulation of lncRNAs are closely related to the development of various complex human diseases, including cancer [1]. Studying the potential associative relationship between lncRNAs and diseases can enhance human understanding of the molecular mechanisms of diseases, and also facilitate the prevention, diagnosis, prognosis and treatment of clinical diseases.

With the improvement of computer computing power and the accumulation of biological databases, researchers are committed to developing computational models that are more convenient and effective than traditional biological experiments to predict associations between lncRNAs and diseases. In the past decade, researchers have proposed a large number of computational methods for predicting potential LDA [2]. Sun et al [3] proposed a random-wandering LDA prediction model (RWRlncD) based on the functional similarity network of lncRNAs to infer potential human LDA, but the method has some limitations and is not applicable to the absence of any known disease-associated LncRNAs. Chen et al [4] developed the prediction model KATZLDA using Katz metric for predicting potential lncRNA-disease associations on heterogeneous networks. With the application of machine learning and deep learning in biology [5], a number of LDA prediction methods based on different machine learning have been proposed, such as Bayesian classifier-based prediction [6], random forest-based prediction [7], and prediction based on orthonormal Laplacian regularised least squares [8]. Chen et al [9] proposed a Laplacian regularised least squares semi-supervised learning method LRLSLDA for identifying potential associations between lncRNAs and diseases, which is the first computational model for

predicting LDA.LRLSLDA calculates lncRNA similarity and disease similarity, constructs two classifiers based on Laplace regularized least squares in the disease space and lncRNA space respectively and combine these two classifiers into one classifier to get the final association probability between disease and lncRNA. Deep learning has been applied to various prediction problems in biology [10]. Xuan et al [11-13] proposed different deep learning based lncRNA disease prediction models such as GCNLDA, CNNDLP, and LDAPred. GCNLDA uses graphical convolutional neural network. CNNDLP employs convolutional neural network and convolutional auto encoder. LDAPred uses convolutional neural network and information flow propagation. Currently, matrix decomposition [14] has been applied to identify potential LDAs. Therefore, some researchers have combined matrix decomposition and deep learning [15] to improve the performance of predicting LDA. Although researchers have explored the field of predicting potential LDA in depth and made great contributions, there are still some limitations of the existing methods. Therefore, bridging the gap of existing methods and developing more efficient and stable prediction methods remains a challenging research endeavour.

II. Proposed Method

A. *Overview*

An improved graph self-encoder model based on multi-source information, referred to as AGCELDA, is proposed to predict potential associations between lncRNA and disease. The model first constructs a heterogeneous graph using information from lncRNA, diseases, and known lncRNA-disease associations. Next, a graph encoder with an attention mechanism models the lncRNA-disease association relationships to obtain high-quality low-dimensional representation vectors. Finally, the graph decoder, also equipped with an attention mechanism, reconstructs the lncRNA-disease associations for potential prediction. The known associations are structured in the heterogeneous graph, allowing nodes to aggregate features through information propagation to

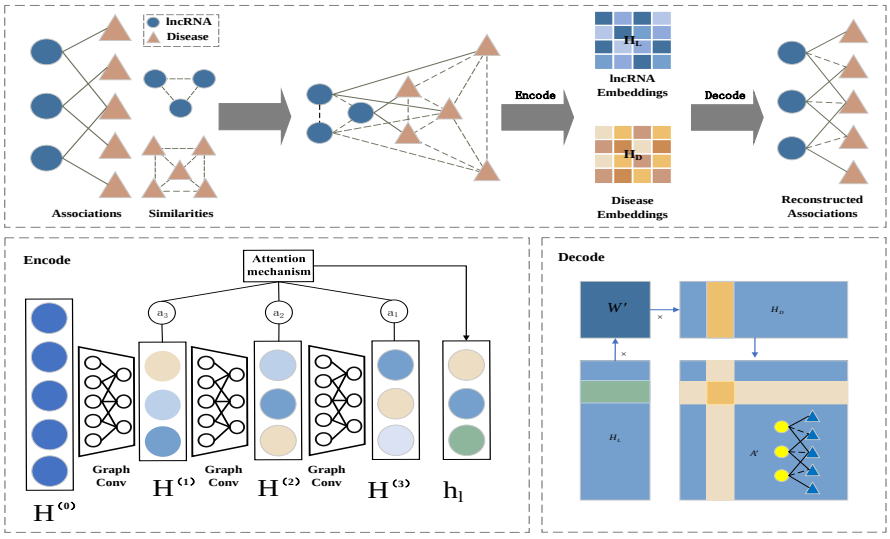

Fig. 1. The workflow of AGCELDA.

model their relationships. By combining graph attention and graph convolution, the model effectively aggregates features between nodes. The graph self-encoder captures the associations between nodes, addressing the issue of incomplete modeling and thereby improving the performance of lncRNA-disease association prediction. The overall model is illustrated in Figure 1.

B. *Disease Semantic Similarity*

According to the approach of Wang [16] and others, a disease is usually modelled as a Directed Acyclic Graph [17] (DAG), where each node denotes a disease or symptom, and each edge denotes a relationship. DAG for disease d_i is a set of nodes N_{d_i} that includes the disease d_i itself and all its ancestors and a set of all direct edges E_{d_i} from parent to child nodes, i.e., $DAG(N_{d_i}, E_{d_i})$. The semantic contribution of disease to disease in DAG_{d_i} is given in Equation (1):

$$\begin{cases} C_{d_i}(d) = 1, & if\ d = d_i \\ C_{d_i}(d) = \max\left\{\varepsilon \times C_{d_i}(d') \middle| d' \in children\ of\ d\right\} & if\ d \neq d_i \end{cases} \quad (1)$$

where ε denotes the semantic contribution factor, in DAG_{d_i}. Here, it is assumed that more distant ancestor nodes contribute less to a specific semantic value of node d_i. Therefore, $\varepsilon \in (0,1)$ is usually set to 0.5. For the computation of the semantic value $D_v(d_i)$ for disease d_i, see Equation (2):

$$D_v(d_i) = \sum_{d \in N_{d_i}} C_{d_i}(d) \quad (2)$$

By considering the relative positions of the two diseases in DAG_{d_i}, it is assumed that diseases sharing most of DAG tend to have higher semantic similarity. Therefore, the semantic similarity $SD(d_i, d_j)$ of the two diseases d_i and d_j is calculated based on their positions in the DAG graph with respect to their ancestor disease nodes, see equation (3):

$$SD(d_i, d_j) = \frac{\sum_{d \in N_{d_i} \cap d \in N_{d_j}} (D_{d_i}(d) + D_{d_j}(d))}{D_v(d_i) + D_v(d_j)} \quad (3)$$

where $D_{d_i}(d)$ denotes the semantic value of Disease d in relation to Disease d_i, and similarly, $D_{d_j}(d)$ denotes the semantic value of Disease d in relation to Disease d_j.

C. *lncRNA Sequence Similarity*

Sequence similarity between lncRNAs and lncRNAs is calculated by Levenshtein distance [18] (also known as edit distance). Specifically, for two different lncRNA sequence information, it needs to be calculated by dynamic programming method. In this case, the editing cost of inserting and deleting lncRNA sequence characters is set to 1, while the editing cost of replacing sequence characters is set to 2. Then, the sequence similarity $S(l_i, l_j)$ between two different lncRNAs can be computed as shown in Eq. (4):

$$SL(l_i, l_j) = 1 - \frac{dis}{s(l_i) + s(l_j)} \quad (4)$$

Where $s(l_i)$ denotes the sequence information of lncRNA l_i, $s(l_j)$ denotes the sequence information of lncRNA l_j, and dis denotes the minimum editing cost that needs to be spent to convert from lncRNA l_i

sequence to lncRNA l_j, and $SL(l_i, l_j)$ denotes the sequence similarity between lncRNA l_i and lncRNA l_j. It can be seen that the smaller the editing cost between two lncRNAs, the greater the sequence similarity between these two lncRNAs.

D. *Constructing a Heterogeneous Map of Multiple Sources of Information*

The heterogeneous network was constructed based on lncRNA-disease association, disease-disease similarity and lncRNA-lncRNA similarity. We denote lncRNA-disease associations by a binary matrix $A \in \{0,1\}_{N \times M}$, where M, N denote the number of diseases and lncRNAs, respectively. A_{ij} is equal to 1 if drug l_i is associated with disease d_j; otherwise A_{ij} = 0. The pairwise similarity between N lncRNAs is denoted as the similarity matrix SL, where $SL(l_i, l_j)$ serves as its (i, j) entry, and the pairwise similarity between M diseases is denoted as the similarity matrix SD, where $SD(d_i, d_j)$ serves as its (i, j) entry. We normalise the similarity matrix by $SL = SD_l^{-\frac{1}{2}} SLSD_l^{-\frac{1}{2}}$ and $SD = SD_d^{-\frac{1}{2}} SDSD_d^{-\frac{1}{2}}$, where $SD_l = diag(\sum_j SL_{ij})$ and $SD_d = diag(\sum_j SD_{ij})$. Finally, we construct the heterogeneous network defined by the adjacency matrix as shown in equation (5):

$$A_H = \begin{bmatrix} SL & A \\ A^{\mathrm{T}} & SD \end{bmatrix} \in \mathbb{R}^{(N+M)\times(N+M)} \tag{5}$$

III. EXPERIMENT

A. *Dataset*

The known LDAs used in the dataset were obtained by download from the LncRNADiseasev2.0, Lnc2Cancerv2.0, and GeneRIF databases, with a total of 240 lncRNAs, 412 diseases, and 2,697 associations.

B. *Graphical Convolutional Neural Network*

GCN [19] is a multi-layer connected neural network architecture for learning low-dimensional representations of nodes from graph-structured data. Each layer of GCN reconstructs the embeddings by aggregating information about its neighbours through the direct links of the graph as input to the next layer. Specifically, given a network with a corresponding neighbourhood matrix G, the layerwise propagation rule of a GCN is formulated as:

$$H^{(l+1)} = f(H^{(l)}, G) = \partial(D^{-\frac{1}{2}} G D^{-\frac{1}{2}} H^{(l)} W^{(l)}) \tag{6}$$

where $H^{(l)}$ is the embedding of nodes in layer l, $D = diag(\sum_j G_{ij})$ is the degree matrix of G, $W^{(l)}$ is the layer-specific trainable weight matrix, and $\partial(\cdot)$ is the nonlinear activation function. To construct GCN-based encoders to learn low-dimensional representations of drugs and diseases, we consider combining node similarity and directly linked association information by deploying GCNs on the constructed heterogeneous graph A_H. We use GCNs as a tool to control node similarity and direct linkage. First, we introduce a penalty factor η to control the contribution of similarity in the GCN propagation process. Specifically, we set the input graph G to be:

$$G = \begin{bmatrix} \eta SL & A \\ A^{\mathrm{T}} & \eta SD \end{bmatrix} \tag{7}$$

We then initialise the embedding to:

$$H^{(0)} = \begin{bmatrix} 0 & A \\ A^{\mathrm{T}} & 0 \end{bmatrix} \tag{8}$$

With the above setup, the first layer of our GCN encoder is formulated as:

$$H^{(1)} = \partial(D^{-\frac{1}{2}} G D^{-\frac{1}{2}} H^{(0)} W^{(0)}) \tag{9}$$

where $W^{(0)} \in \mathbb{R}^{(N+M)\times k}$ is the input-hidden weight matrix, $H^{(1)} \in \mathbb{R}^{(N+M)\times k}$ is the first layer embedding of the nodes (lncRNA and disease) of the heterogeneous network A_H, k is the dimension of the embedding, and G is defined in Equation (7). The subsequent layers of our GCN encoder are for $l = 1, 2, ..., l$, where $W^{(l)} \in \mathbb{R}^{k\times k}$. After l iterations, we can obtain the k dimension embedding of l from different graph convolution layers. Exponential linear units [20] are used as nonlinear activation functions in all graph convolution layers, which not only accelerates the learning process but also significantly enhances the generalisation performance. Embeddings at different layers capture different structural information about the heterogeneous network. For example, the first layer captures direct link information and higher layers capture multi-hop neighbour information (higher order proximity) by iteratively updating the embedding [21]. Considering that the contributions of different embeddings in different layers are inconsistent, we introduce an attention mechanism to combine these embeddings to obtain the final embedding of drugs and diseases as $\begin{bmatrix} H_L \\ H_D \end{bmatrix} = \sum a_l H^{(l)}$, where $H_L \in \mathbb{R}^{N\times k}$ is the final embedding of lncRNAs, $H_D \in \mathbb{R}^{N\times k}$ is the final embedding of diseases, and a_l is automatically learnt by the neural network and initialised as $1/(l+1), l = 1, 2, ..., l.$

To reconstruct the adjacency matrix of the lncRNA-disease association, the bilinear decoder A, created by Huang [22], is used as shown in equation (10):

$$A' = sigmoid(H_L W' H_D^{\mathrm{T}}) \tag{10}$$

where $W' \in \mathbb{R}^{k\times k}$ is the trainable matrix. The predictive score for the association between lncRNA l_i and disease d_j is given by the corresponding (i, j) entry of A', denoted a'_{ij}.

C. *Optimization*

All trainable weight matrices (W' and $W^{(l)}$) are initialised by the Xaiver initialisation method [23]. We then minimise the loss function using the Adam optimizer [24]. The Adam optimizer can iteratively update the weights of the neural network based on the training data. To prevent overfitting, we introduce node deletion [25] and regular deletion [26] into the graph convolutional layer. This node loss can be thought of as training different models on a variety of small sub-networks to predict unknown lncRNA-disease pairs by integrating these small models [27]. In addition, cyclic learning rates [28] are used in the optimisation process. The simple cyclic learning rate makes the learning rate vary between the maximum and minimum learning rate, helping us to balance the training speed and ACC. (1)

D. *Experimental Setting*

In our experiments, we used five-fold cross-validation (5-CV) to evaluate the performance of the prediction methods. All known drug-disease associations were randomly divided into five equally sized subsets. The cross-validation process was repeated five times, with each subset in turn serving as the test set and the remaining four subsets serving as the training set. In each fold, predictive models were constructed on the known associations in the training set and used to predict the associations in the test set. We use AUPR and AUC as the primary metrics as they can measure the performance of the method without any specific threshold. There are several hyper-parameters in AGCELDA, such as embedding dimensionality k, the number of layers l, the initial learning rate of the optimiser lr, the total number of training sessions in AGCELDA α, two dropout rates (node dropout and regular dropout) β, γ, and the penalty factor η in heterogeneous networks, Consider different combinations of these parameters, $\alpha \in \{500, 1000, 2000, 4000\}$, β, $\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$ and $\eta \in \{2, 4, 6, 8, 10\}$. By tuning the parameters empirically, we set parameters $k = 64$, $l = 3$, $lr = 0.008$, $\alpha = 4000$, $\beta = 0.6$, $\gamma = 0.4$, $\eta = 6$ for AGCELDA in the following experiments.

E. *Effect of Layer Attention Mechanism*

Layer attention is an integral part of the AGCELDA network architecture and is responsible for managing and quantifying the interdependence of different convolutional layers. Here, we discuss the role of layer attention mechanisms. We build AGCELDA models, abbreviated as AGCELDA-L1, AGCELDAL2, and AGCELDA-L3, using only the embeddings of the 1st layer of AGCELDA of $l = 1, 2, 3$. The results of all the models evaluated on the main dataset by 5-CV are shown in Table 3. AGCELDA-L1 and AGCELDAL2 produce better results than AGCELDA-L3 produce better results, indicating that the first and second level embeddings contain more information than the third level embedding. The results may be caused by excessive smoothing of the GCN. AGCELDA with integrated embedding at the third layer produces better results than AGCELDA-L1, AGCELDA-L2 and AGCELDA-L3. It is believed that the lth convolutional layer of the GCN captures the lth order approximation and that the attention weights indicate the contribution of the embeddings of the different convolutional layers to the final embedding. We performed 10 5-CVs on AGCELDA and visualised the attention weights of the three convolutional layers in Fig. 2. The attentional weights of the three layers are different, with layer 1 > layer 2 > layer 3, which is in line with our expectation of a higher contribution from lower-order proximity and a lower contribution from higher-order proximity. These results also help explain the performance of AGCELDA-L1, AGCELDA-L2 and AGCELDA-L3 in Table 3. Therefore, it is necessary to pay different attention to the convolutional layers when building high-precision prediction models.

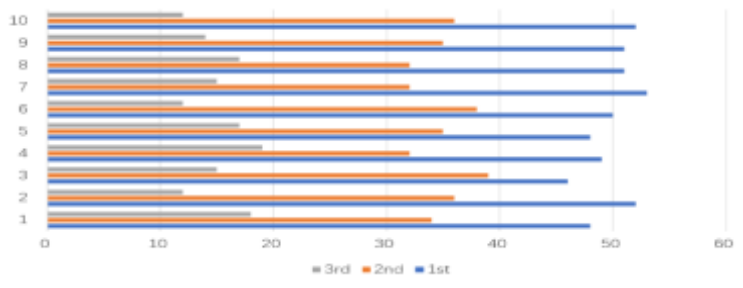

Fig. 2. Attention weights for three convolution layers in AGCELDA.

In addition, we consider alternative ways of combining embeddings at different layers. AGCELDA-AVE assigns uniform weights to different embeddings; AGCELDA-CON directly concatenates different embeddings. As shown in Table 1, AGCELDA produces better results than AGCELDA-AVE and AGCELDA-CON, demonstrating the effectiveness of the attention mechanism in AGCELDA. To better represent the differences, we use a visual bar chart to show them, see Figure 3.

Table 1. Performance of AGCELDA based on different embeddings.

Models	AUPR	AUC
AGCELDA	0.8037	0.8928
AGCELDA-AVE	0.7125	0.8174
AGCELDA-CON	0.7436	0.8237
AGCELDA-L1	0.7137	0.8564
AGCELDA-L2	0.7112	0.8430
AGCELDA-L3	0.6928	0.7318

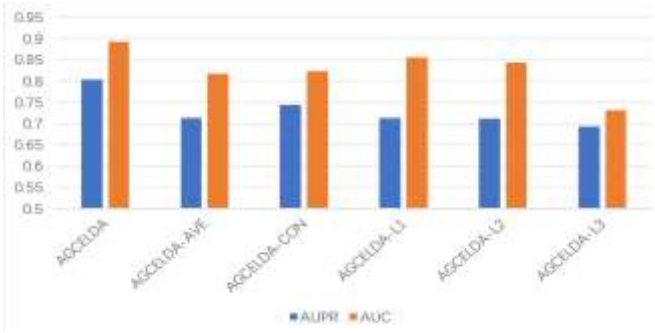

Fig. 3. Performance of AGCELDA based on different embeddings.

F. *Comparison with Other Methods*

In this section, in order to demonstrate the prediction performance of AGCELDA, four state-of-the-art prediction methods are selected for performance analysis on the dataset, these methods include LRLSLDA [29], MFLDA [30], TPGLDA [31], LDAP [32], The performance comparison between AGCELDA and other state-of-the-art methods is shown in Table 2. The final experimental results represent the average of 10 5-CVs. The results show that the average values of 5-CV for LRLSLDA, MFLDA, TPGLDA, LDAP and AGCELDA are 0.8286, 0.8503, 0.8423, 0.8856, and 0.9028, respectively. The results show that AGCELDA outperforms the other state-of-the-art methods. In particular, it improves by 7.42% compared to the earlier LRLSLDA model.

Table 2. Performance comparison.

	Acc.	Pre.	Re.	F1.	AUC	AUPR
LRLSLDA	0.7466	0.8133	0.7977	0.8054	0.8286	0.7203
MFLDA	0.7671	0.7921	0.7623	0.7769	0.8503	0.8721
TPGLDA	0.7522	0.7864	0.8116	0.7988	0.8423	0.7468
LDAP	0.8087	0.7521	0.7387	0.7453	0.8856	0.7711
AGCELDA	0.7982	0.8038	0.8102	0.8070	0.9028	0.8637

In order to give a more visual representation of the performance of our model compared to the performance of other models, radar plots are given, as in Fig. 4:

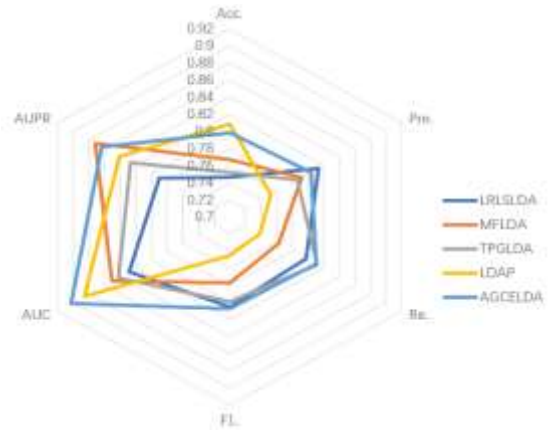

Fig. 4. Performance comparison.

G. *Impact of Data Set Size on Model Performance*

In order to further assess the effect of dataset size on the performance of the AGCELDA model, two-fold cross validation and ten-fold cross validation were used on the dataset, and their performance is shown in Table 3.

Table 3. Performance of 2-fold and 10-fold cross validation.

	Acc.	Pre.	Re.	F1.	AUC	AUPR
2-fold average	0.8267 ±0.0175	0.6267 ±0.0253	0.8877 ±0.0164	0.7347 ±0.0136	0.8919 ±0.0118	0.8877 ±0.0116
10-fold average	0.8405±0.0184	0.6357 ±0.0197	0.8955 ±0.0155	0.7436 ±0.0178	0.9045 ±0.0114	0.9031 ±0.0111

From the comparative results, it can be seen that increasing the number of training samples improves the prediction performance. The 10-fold average has a higher AUC and smaller standard deviation than the 2-fold average, which proves that the accuracy and stability of the model improves with the increase of the dataset, and it is expected that the predictive ability of the proposed model will be further improved with the increase of the data from the future studies.

IV. CASE STUDY

To further demonstrate the correctness and rationality of AGCELDA in identifying potential lncRNA-diseases, this subsection conducts a case study of three common human diseases, including gastric cancer and cervical cancer. The specific procedure is to rank the candidate lncRNAs for each disease based on the final prediction score. The top ranked lncRNAs were selected for analysis in the databases Lnc2cancer and LncRNA Disease.

Table 4. Predicted gastric cancer-related lncRNAs.

Ranking	**lncRNA**	**Prove**
1	HOTAIR	Lnc2cancer
2	MEG3	Lnc2cancer

Ranking	lncRNA	Prove
3	UCAI	Lnc2cancer
4	TUSC7	LncRNADisease
5	EPB41L4A-AS1	Unconfirmed
6	CAT1	LncRNADisease
7	LINC00261	LncRNADisease
8	FOXCUT	Lnc2cancer
9	LSINCTS	LncRNADisease
10	AFAPI-AS1	Lnc2cancer

Gastric cancer is the most serious malignant tumour in gastric diseases, of which males are the majority. AGCELDA was used to predict the 10 potential lncRNAs that have the highest correlation scores with gastric cancer to analyse whether they are associated or not. The specific information is shown in Table 4. Of the top 10 predicted novel lncRNAs, 9 were successfully confirmed to be associated with gastric cancer by the database. For example, researchers found that TUSC7 was down-regulated in gastric cancer samples, was an independent prognostic indicator for disease-free survival (DFS) and disease-specific survival (DSS) in gastric cancer patients, and that TUSC7 could be determined to inhibit the growth of tumour cells in vitro and in vivo [33]. Li et al. [34] found that CAT1 promotes gastric cancer by negatively regulating miR-219-1 for both tumourigenesis and progression. It should be noted that although EPB41L4A-AS1 has not been found to be associated with gastric cancer in these two databases, it has been demonstrated that the expression of EPB41L4A-AS1 in tumours is lower than that in normal tissues by searching for relevant literature, and thus it can be inferred that it may be an oncogene in gastric cancer [35].

Table 5. Predicted cervical cancer-related lncRNAs.

Ranking	lncRNA	Prove
1	MEG3	Lnc2cancer
2	GAS5	Lnc2cancer
3	UCAI	Lnc2cancer
4	PVTI	LncRNADisease
5	XIST	Lnc2cancer
6	LSINCT5	Unconfirmed
7	BANCR	Unconfirmed
8	CCAT1	Lnc2cancer
9	LSINCTS	LncRNADisease
10	HULC	Lnc2cancer

Cervical cancer is one of the most common malignant tumours in gynaecology, which seriously endangers women's physical and mental health. Therefore, early diagnosis and treatment of patients is particularly important. After the completion of the case study, the top 10 lncRNAs in terms of association score with

cervical cancer were also used for further validation. As can be seen in Table 5, 8 of the 10 novel lncRNAs predicted by AGCELDA have been confirmed by Lnc2cancer or LncRNADiseasev2.0. This proves that these 8 predicted lncRNAs are indeed associated with cervical cancer. For example, researchers found that XIST is extremely highly expressed in cervical cancer tissues and cell lines and acts as a ceRNA in cervical cancer progression by regulating miR-200a [36]. Meanwhile, a study proved to determine the role of CCAT1 in cervical cancer cell proliferation and invasion, and the expression of CCAT1 in cervical cancer tissues was higher than that in neighbouring normal tissues. Overexpression of CCAT1 promoted cervical cancer cell proliferation, colony formation and invasion in vitro [37]. Although 2 lncRNAs have not been confirmed in the database, experimental results suggest that these 2 lncRNAs are also closely associated with cervical cancer.

V. CONCLUSION

This paper proposes an improved graph autoencoder model (AGCELDA) for predicting potential associations between lncRNAs and diseases. The model first constructs a heterogeneous graph that integrates information from lncRNAs, diseases, and known associations. Next, the graph encoder models the lncRNA-disease associations and generates low-dimensional representation vectors. Finally, the graph decoder reconstructs these relationships for potential association predictions. Through information propagation, nodes in the heterogeneous graph aggregate features, effectively combining graph attention and graph convolution. The graph autoencoder models the aggregated features to capture the relationships between nodes, addressing the issue of incomplete modeling of associations and thus improving prediction performance. In five-fold cross-validation experiments, AGCELDA achieved AUC and AUPR values of 0.9028 and 0.8637, respectively, outperforming other models. The experimental results indicate that AGCELDA is capable of better predicting potential associations between lncRNAs and diseases.

ACKNOWLEDGMENTS

This study was supported in part by the National Natural Science Foundation of China (No. 62172338). ‘Research on Multi-source Adverse Drug Reaction Extraction Method Based on Quantum Deep Learning and Knowledge Graph’, supported by the National Natural Science Foundation of China (Grant No. 62172338), from January 2022 to December 2025.

REFERENCES

[1] Chen X, Yan C.C., Zhang X, et al. Long non-coding RNAs and complex diseases: from experimental results to computational models [J]. Briefings in bioinformatics, 2017, 18(4): 558576.
[2] Fan Y, Chen M, Pan X. GCRFLDA: scoring lncRNA-disease associations using graph convolution matrix completion with conditional random field [J]. Briefings in Bioinformatics, 2022, 23(1): bbab361.
[3] Sun J, Shi H, Wang Z, et al. Inferring novel lncRNA–disease associations based on a random walk model of a lncRNA functional similarity network [J]. Molecular BioSystems, 2014, 10(8): 2074-2081.
[4] Chen X. KATZLDA: KATZ measure for the lncRNA-disease association prediction [J]. Scientific reports, 2015, 5(1): 1-11.
[5] Wainberg M, Merico D, Delong A, et al. Deep learning in biomedicine [J]. Nature biotechnology, 2018, 36(9): 829-838.
[6] Yu J, Xuan Z, Feng X, et al. A novel collaborative filtering model for LncRNA-disease association prediction based on the Naive Bayesian classifier [J]. BMC bioinformatics, 2019, 20(1): 1-13.
[7] Yao D, Zhan X, Zhan X, et al. A random forest based computational model for predicting novel lncRNA-disease associations [J]. BMC bioinformatics, 2020, 21(1): 1-18.
[8] Deng L, Li W, Zhang J. LDAH2V: Exploring meta-paths across multiple networks for lncRNA-disease association prediction [J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2019, 18(4): 1572-1581.
[9] Chen X, Yan G-Y. Novel human lncRNA–disease association inference based on lncRNA expression profiles [J]. Bioinformatics, 2013, 29(20): 2617-2624.
[10] Zeng M, Zhang F, Wu F-X, et al. Protein–protein interaction site prediction through combining local and global features with deep neural networks [J]. Bioinformatics, 2020, 36(4): 1114-1120.
[11] Xuan P, Pan S, Zhang T, et al. Graph convolutional network and convolutional neural network based method for predicting lncRNA-disease associations [J]. Cells, 2019, 8(9): 1012.
[12] Xuan P, Sheng N, Zhang T, et al. CNNDLP: a method based on convolutional autoencoder and convolutional neural network with adj-

-acent edge attention for predicting lncRNA–disease associations [J]. International journal of molecular sciences, 2019, 20(17): 42 60.
[13] Xuan P, Jia L, Zhang T, et al. LDAPred: a method based on information flow propagation and a convolutional neural network for the prediction of disease-associated lncRNAs [J]. International journal of molecular sciences, 2019, 20(18): 4458.
[14] Wang Y, Yu G, Wang J, et al. Weighted matrix factorization on multi-relational data for LncRNA-disease association prediction [J]. Methods, 2020, 173: 32-43. [50] Fu G, Wang J, Domeniconi C, et al. Matrix factorization-based data fusion for the prediction of lncRNA–disease associations [J]. Bioinformatics, 2018, 34(9): 1529-1537.
[15] Zeng M, Lu C, Zhang F, et al. SDLDA: lncRNA-disease association prediction based on singular value decomposition and deep learning [J]. Methods, 2020, 179: 73-80.
[16] Wang D, Wang J, Lu M, et al. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases [J]. Bioinformatics, 2010, 26(13): 1644-50.
[17] Digitale J C, Martin J N, Glymour M M. Tutorial on directed acyclic graphs [J]. Journal of Clinical Epidemiology, 2022, 142: 264-7.
[18] Yujian L, Bo L. A normalized Levenshtein distance metric [J]. IEEE transactions on pattern analysis and machine intelligence, 2007, 29(6): 1091-5.
[19] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR), 2017. https://iclr.cc/archive/ www/2017.html.
[20] Clevert D-A, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). In: International Conference on Learning Representations (ICLR), 2016.
[21] Wang X, He X, Wang M et al. Neural graph collaborative filtering. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY: Assoc Computing Machinery, 2019, pp. 165-74.
[22] Y-A H, Hu P, KCC C, et al. Graph convolution for predicting associations between miRNA and drug resistance. Bioinformatics 2019;36:851–8.
[23] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010, pp. 249–56. http://proceedings.mlr.press/v9/ glorot10a/glorot 10a.pdf.
[24] Kingma DP, Ba J. Adam: a method for stochastic optimization. In: International Conference on Learning Representations (ICLR), 2015. https://iclr.cc/archive/www/2015.html.
[25] van den Berg R, Kipf TN, Welling M. Graph convolutional matrix completion. In: KDD, 2018. https://www.kdd.org/ kdd2018/deep-learning-day.
[26] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15:1929–58.
[27] Zhu L, Hong Z, Zheng H. Predicting gene-disease associations via graph embedding and graph convolutional networks. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019, pp. 382–9.
[28] Smith LN. Cyclical learning rates for training neural networks. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). New York, NY: IEEE, 2017, pp. 464–72.
[29] Chen X, Yan G-Y. Novel human lncRNA–disease association inference based on lncRNA expression profiles[J]. Bioinformatics, 2013, 29(20): 2617-2624.
[30] Fu G, Wang J, Domeniconi C, et al. Matrix factorization-based data fusion for the prediction of lncRNA–disease associations[J]. Bioinformatics, 2018, 34(9): 1529-1537.
[31] Ding L, Wang M, Sun D, et al. TPGLDA: Novel prediction of associations between lncRNAs and diseases via lncRNA-disease-gene tripartite graph [J]. Scientific reports, 2018, 8(1): 1-11.
[32] Lan W, Li M, Zhao K, et al. LDAP: a web server for lncRNA-disease association prediction [J]. Bioinformatics, 2017, 33(3): 458-460.
[33] Qi P, Xu M D, Shen X H, et al. Reciprocal repression between TUSC 7 and mi R-23b in gastric cancer [J]. International journal of cancer, 2015, 137(6): 1269-1278.
[34] Yanfeng, Li, Guanyu, et al. lncRNA CCAT1 contributes to the growth and invasion of gastric cancer via targeting miR-219-1 [J]. Journal of Cellular Biochemistry, 2019, 120(12): 1945719468.
[35] Delshad E, Shamsabadi F T, Bahramian S, et al. In silico identification of novel lncRNAs with a potential role in diagnosis of gastric cancer [J]. Journal of Biomolecular Structure & Dynamics, 2020, 38(7): 1954-1962.
[36] Zhu H, Zheng T, Yu J, et al. LncRNA XIST accelerates cervical cancer progression via upregulating Fus through competitively binding with miR-200a[J]. Biomedicine & Pharmacotherapy, 2018, 105: 789-797.

AUTHOR'S PROFILE

First Author
Xiaoqian Li, was born in Shandong, China, in 1998. She received the B.S. degree in information and computing science from the Xi'an University of Art and Science, Xi'an, Shaanxi, China, in 2021. She is pursuing the M.S. degree in electronic information with Xijing University, Xi'an, Shaanxi, China. Her current research includes data mining, bioinformatics, and computer vision. email id: 347204787@qq.com

Second Author
Shanwen Zhang, was born in Shaanxi, China. He received the B.S. degree in mathematics from Northwest University, Xi'an, China, in 1988, the M.S. degree in applied mathematics from Northwest Polytechnic University, Xi'an, in 1995, and the Ph.D. degree in electromagnetic field and microwave from Air Force Engineering University, Xi'an, in 2001.He is currently a Professor with Xijing University, Xi'an, and a Visiting Scholar with the Department of Computer Science, Virginia Tech, Blacksburg, VA, USA. His research interests include machine learning and its application in data mining, including machine learning, image processing, data reduction, data mining, feature selection, and wavelet transforms.

Third Author
Shenao Yuan, was born in Henan, China, in 2001. He received the B.S. degree in Electronic Information Engineering from Hubei Institute of Technology in 2023. He is currently pursuing the M.S. degree in new generation electronic information technology with Xijing University, Xi'an, China. His research interests include pattern recognition, underwater image processing. email id: y15294841834@163.com