



## OPEN ACCESS

## EDITED BY

Qi Zhao,  
University of Science and Technology Liaoning,  
China

## REVIEWED BY

Wen Zhang,  
Huazhong Agricultural University, China  
Jian Huang,  
University of Electronic Science and  
Technology of China, China

## \*CORRESPONDENCE

Xuequn Shang  
✉ shang@nwpu.edu.cn

RECEIVED 26 January 2023

ACCEPTED 04 April 2023

PUBLISHED 26 April 2023

## CITATION

Gao M and Shang X (2023) Identification of associations between lncRNA and drug resistance based on deep learning and attention mechanism. *Front. Microbiol.* 14:1147778. doi: 10.3389/fmicb.2023.1147778

## COPYRIGHT

© 2023 Gao and Shang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Identification of associations between lncRNA and drug resistance based on deep learning and attention mechanism

Meihong Gao and Xuequn Shang\*

School of Computer Science, Northwestern Polytechnical University, Xi'an, China

**Introduction:** Abnormal lncRNA expression can lead to the resistance of tumor cells to anticancer drugs, which is a crucial factor leading to high cancer mortality. Studying the relationship between lncRNA and drug resistance becomes necessary. Recently, deep learning has achieved promising results in predicting biomolecular associations. However, to our knowledge, deep learning-based lncRNA-drug resistance associations prediction has yet to be studied.

**Methods:** Here, we proposed a new computational model, DeepLDA, which used deep neural networks and graph attention mechanisms to learn lncRNA and drug embeddings for predicting potential relationships between lncRNAs and drug resistance. DeepLDA first constructed similarity networks for lncRNAs and drugs using known association information. Subsequently, deep graph neural networks were utilized to automatically extract features from multiple attributes of lncRNAs and drugs. These features were fed into graph attention networks to learn lncRNA and drug embeddings. Finally, the embeddings were used to predict potential associations between lncRNAs and drug resistance.

**Results:** Experimental results on the given datasets show that DeepLDA outperforms other machine learning-related prediction methods, and the deep neural network and attention mechanism can improve model performance.

**Discussion:** In summary, this study proposes a powerful deep-learning model that can effectively predict lncRNA-drug resistance associations and facilitate the development of lncRNA-targeted drugs. DeepLDA is available at <https://github.com/meihonggao/DeepLDA>.

## KEYWORDS

lncRNA-drug resistance associations, deep neural networks, graph attention mechanisms, similarity networks, embeddings

## 1. Introduction

Long non-coding RNA (lncRNA) is a transcript longer than 200 nucleotides, which is transcribed from the genome, cannot be translated into functional proteins, and has heterogeneity in organisms (Mattick and Rinn, 2015; Koch, 2017; Bridges et al., 2021; Zhao et al., 2021). More and more lncRNAs have been proven to affect various biological processes to cause cancer through transcription initiation, transcriptional and post-transcriptional regulation, and are no longer the so-called “transcriptional noise” (Long et al., 2017; Peng et al., 2017; Jiang et al., 2019; Gao et al., 2021, 2022b). In addition, cancer is the leading cause of death worldwide (Bray et al., 2020; Ferlay et al., 2021; Gao and Shang, 2022b; Xia et al., 2022). The current primary therapeutic approach for cancer is chemotherapy, which uses chemical drugs to treat patients. However, tumor cells can become resistant to chemotherapy

during treatment, leading to treatment failure (Rebucci and Michiels, 2013; Hu et al., 2019; Rehman et al., 2021). Thus, drug resistance is still a significant challenge in cancer therapy, and its underlying mechanisms have not been fully elucidated.

lncRNAs have recently been identified as a novel mechanism of drug resistance and have received extensive attention in cancer research (Bester et al., 2018; Sun et al., 2019; Barth et al., 2020; Jiang et al., 2020; Singh et al., 2022; Zhou et al., 2022). The abnormal expression of lncRNA can lead to the resistance of tumor cells to anticancer drugs, which is a crucial factor leading to high cancer mortality. For example, lncRNA HOTAIR is found to be upregulated in tumors, such as breast cancer, gastric cancer, esophageal cancer, and leukemia (Xue et al., 2016; Zhu et al., 2022). It not only participates in the formation of multidrug resistance of tumor cells but also is closely related to the degree of malignancy and poor prognosis of tumors. In addition, the overexpression of lncRNA UCA1 in urothelial carcinoma is associated with chemotherapy resistance, and silencing lncRNA UCA1 can inhibit the migration and invasion of non-small cell lung cancer cells and reverse the drug resistance of cancer cells (Wang et al., 2017; Liu et al., 2019). Furthermore, lncRNA DILA1 is associated with the drug resistance of breast cancer cells, which makes cancer cells resistant to tamoxifen by inhibiting the degradation of cyclin D1 (Shi et al., 2020). Overall, there is a close relationship between lncRNA and drug resistance, and the study of lncRNA-drug resistance association becomes crucial.

Some databases have provided experimentally validated lncRNA-drug resistance association data (Dai et al., 2017; Li et al., 2020). However, existing information is small compared with the unknown one. Although biological experiments can identify new lncRNA-drug resistance associations, they are challenging due to high time and financial costs. Computational methods can predict potential associations between lncRNAs and drug resistance, but to our knowledge, only two related works have been proposed. One is LRGCPND (Li et al., 2021), which infers the relationship between noncoding RNAs and drug resistance based on linear residual graph convolution. The other is GSLRDA (Zheng J. et al., 2022), which uses light graph convolutional networks (GCNs), data augmentation, and self-supervision to identify associations between ncRNAs and drug resistance. There is still much for exploration in lncRNA-drug resistance association prediction. In recent years, machine learning has achieved remarkable results in predicting biomolecular association, such as lncRNA-gene association (Zhang et al., 2020; Zhao et al., 2020; Gao and Shang, 2022a; Gao et al., 2022a), lncRNA-miRNA association (Liu et al., 2020; Zhang et al., 2021a,b), miRNA-drug resistance association (Huang et al., 2020; Niu et al., 2022; Zheng K. et al., 2022), and ncRNA-drug resistance association (Li et al., 2021; Zheng J. et al., 2022). Inspired by this, machine learning methods can be used to predict potential lncRNA-drug resistance associations to explore the impact of lncRNAs on the drug resistance of cancer cells.

In this study, we proposed a deep learning-based computational model, DeepLDA, which used deep neural network and graph attention mechanism to learn embeddings of lncRNAs and drugs for predicting potential lncRNA-drug resistance associations. We first used the known relationship between lncRNAs and drug resistance to construct similarity networks for lncRNAs and

drugs. Subsequently, deep GCN were used to automatically extract features from multiple attributes of the raw data of nodes. These features were then used as input to graph attention network (GAT) module for embedding learning. Finally, lncRNA and drug embeddings were used to predict potential associations between lncRNAs and drug resistance. Experimental results show that the prediction performance of DeepLDA is better than other ncRNA-drug resistance association prediction methods, and the deep neural network and attention mechanism are proved to improve the model performance. In summary, this study proposes a new computational model that can effectively complete the task of lncRNA-drug resistance association prediction, help to understand the lncRNA-related drug resistance mechanism, accelerate drug development, and promote the development of targeted therapy.

## 2. Materials and methods

We designed a new computational model, DeepLDA, to predict candidate lncRNA-drug associations based on deep learning and graph attention mechanism (Figure 1). Firstly, lncRNA-drug resistance association data were collected and preprocessed (Figure 1A). Then, we proposed a deep learning module based on graph neural network and graph attention mechanism to learn lncRNA and drug embeddings (Figure 1B). Finally, these learned embeddings were used to identify potential associations between lncRNAs and drug resistance (Figure 1C).

### 2.1. Data collection and experimental setup

lncRNAs-drug resistance associations were collected from NoncoRNA (Li et al., 2020) and ncDR (Dai et al., 2017) datasets. Here, experimentally validated association terms were extracted for subsequent prediction analysis. After preprocessing, these items are transformed into association networks (Table 1). As we can see in Table 1, the number of lncRNAs, drugs, and lncRNA-drug resistance associations in the NoncoRNA-related association network are 3,601, 71, and 3,802, respectively. In the ncDR-related association network, the numbers of lncRNAs, drugs and lncRNA-drug resistance associations are 162, 31, and 184, respectively.

The model was evaluated on balanced and unbalanced datasets. On the unbalanced dataset, we used all lncRNA-drug association items for a more practical simulation. Here, all negative samples were selected for training and testing. Due to the limited understanding of the regulatory relationship between lncRNAs and drug resistances, the number of positive samples was much lower than that of negative samples, which resulted in an imbalance of samples. On balanced datasets, the same amount of negative samples was sampled as the positive sample before training and testing. During training and testing, we performed 10-fold cross-validation, as shown in our previous study (Gao et al., 2022a). Finally, four metrics, the AUC, AUPR, F1-score, and MCC, were calculated to evaluate model performance.

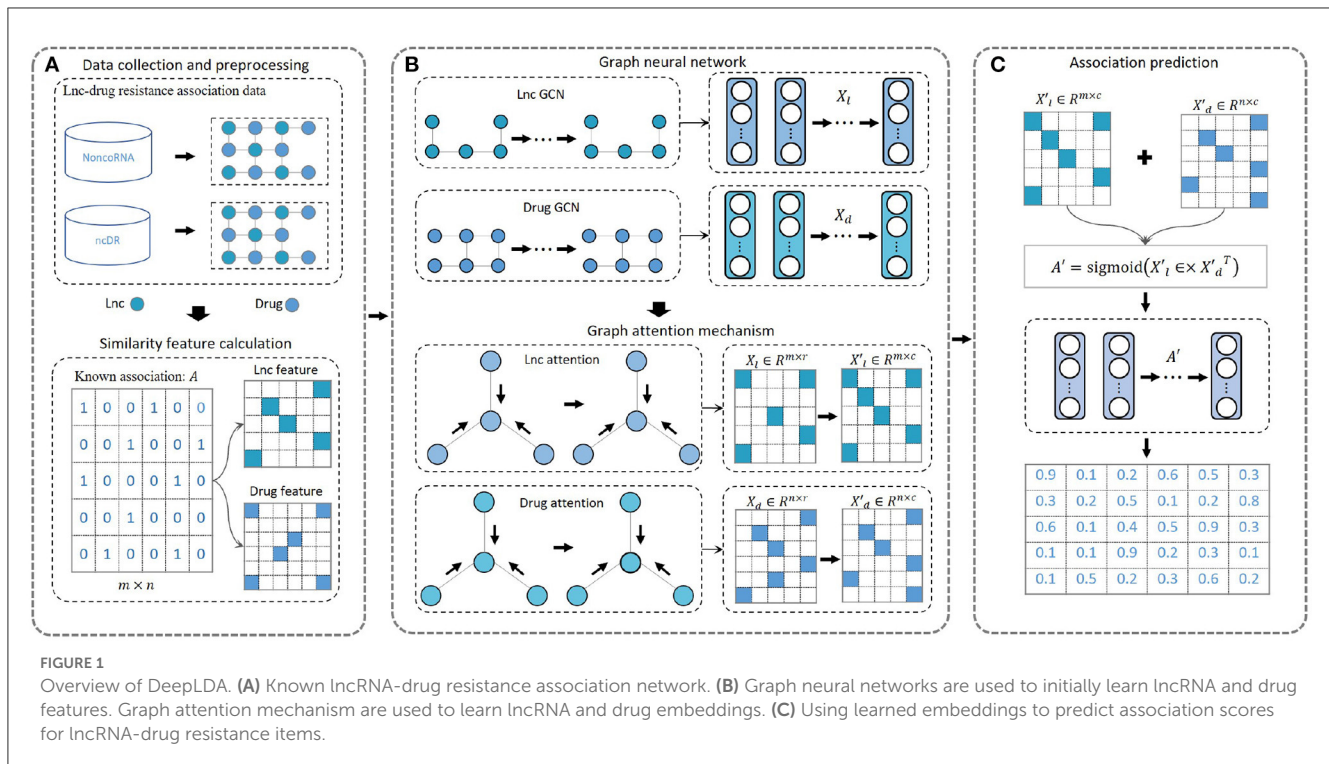


TABLE 1 Details of lncRNA-drug resistance associations.

Dataset	LncRNA	Drug	LncRNA-drug resistance
NoncoRNA	3,599	71	3,802
ncDR	162	31	184

## 2.2. Similarity calculation

We obtained lncRNA and drug features through lncRNA-drug resistance association information. Specifically, we assumed that lncRNAs with similar functions have similar drug resistance patterns. The similarity between lncRNAs was calculated using the Gaussian kernel function as follows:

$$G_l(i, j) = \exp(-\alpha_l \|A(i, :) - A(j, :)\|^2) \quad (1)$$

$$\alpha_l = \frac{1}{m} \sum_{k=1}^m \|A(k, :)\|^2 \quad (2)$$

where  $A \in R^{m \times n}$  represents the known lncRNA-drug resistance associations,  $\|X\|$  represents the Euclidean distance from  $X$  to the origin, and  $m$  and  $n$  represent the number of lncRNAs and drugs related to the association network, respectively. Similarly, we calculated the similarity between drugs as follows:

$$G_d(i, j) = \exp(-\alpha_d \|A(:, i) - A(:, j)\|^2) \quad (3)$$

$$\alpha_d = \frac{1}{n} \sum_{k=1}^n \|A(:, k)\|^2 \quad (4)$$

Finally, we obtained similarity features  $G_l \in R^{m \times m}$  for lncRNAs and  $G_d \in R^{n \times n}$  for drugs.

## 2.3. Embedding learning

We designed a deep learning module based on graph neural network and graph attention mechanism to learn lncRNA and drug embeddings (Figure 1B). GCN was used to initially extract lncRNA features, and its layer propagation formula is as follows:

$$X_l = \text{softmax}(\hat{A}_l \text{ReLU}(\hat{A}_l X W_l^{(0)})) W_l^{(1)} \quad (5)$$

where  $\hat{A}_l = \tilde{D}_l^{-\frac{1}{2}} \tilde{A}_l \tilde{D}_l^{-\frac{1}{2}}$ ,  $\tilde{A}_l = G_l + I$  and  $I$  represents the identity matrix,  $\tilde{D}_l$  represents the degree matrix of matrix  $\tilde{A}_l$ , and  $W_l^{(0)}$  and  $W_l^{(1)}$  represents layer-specific weight matrices. Further, we learned lncRNA embeddings via the graph attention mechanism. Here, the input to the GAT is  $X_l \in R^{m \times r} = \{\tilde{x}_{l_1}, \tilde{x}_{l_2}, \dots, \tilde{x}_{l_m}\}$ , and the attention coefficient between lncRNA  $l_i$  and lncRNA  $l_j$  is defined as follows:

$$e_{l_i l_j} = a(W_l \tilde{x}_{l_i}, W_l \tilde{x}_{l_j}) \quad (6)$$

where  $W_l \in R^{r \times c}$  is a parameter matrix, and  $a$  is a projection:  $R^{c \times c} \rightarrow R$ . Furthermore, we normalized  $e_{l_i l_j}$  across all choices of  $l_j$  as follows:

$$\alpha_{l_i l_j} = \text{softmax}_{l_j}(e_{l_i l_j}) = \frac{\exp(e_{l_i l_j})}{\sum_{l_k \in N_{l_i}} \exp(e_{l_i l_k})} \quad (7)$$

where  $N_{l_i}$  is the neighbor of node  $l_i$  and  $\alpha_{l_i l_j}$  can be fully expanded as:

$$\alpha_{l_i l_j} = \frac{\exp(\text{LeakyReLU}(\tilde{a}^T [W_l \tilde{x}_{l_i} \| W_l \tilde{x}_{l_j}]))}{\sum_{l_k \in N_{l_i}} \exp(\text{LeakyReLU}(\tilde{a}^T [W_l \tilde{x}_{l_i} \| W_l \tilde{x}_{l_k}]))} \quad (8)$$

where  $\vec{a} \in R^{2c}$  is a weight vector and  $\parallel$  is the concatenation operation. Based on this, we obtained the output feature of  $l_i$  as:

$$\vec{x}'_{l_i} = \sigma \left( \sum_{l_j \in N(l_i)} \alpha_{l_i l_j} W_l \vec{x}_{l_j} \right) \quad (9)$$

where  $\vec{x}'_{l_i}$  was further calculated by the multi-head attention mechanism as:

$$\vec{x}'_{l_i} = \parallel_{k=1}^K \sigma \left( \sum_{l_j \in N(l_i)} \alpha_{l_i l_j}^k W_l^k \vec{x}_{l_j} \right) \quad (10)$$

where  $K$  is the number of heads,  $\alpha_{l_i l_j}^k$  is computed by the  $k$ -th head, and  $W_l^k$  is the corresponding weight matrix. Specially, the multi-head attention mechanism in the last layer is:

$$\vec{x}'_{l_i} = \sigma \left( \frac{1}{K} \sum_{k=1}^K \sum_{l_j \in N(l_i)} \alpha_{l_i l_j}^k W_l^k \vec{x}_{l_j} \right) \quad (11)$$

After the above operation, lncRNA embeddings are expressed as  $X'_l \in R^{m \times c} = \{\vec{x}'_{l_1}, \vec{x}'_{l_2}, \dots, \vec{x}'_{l_m}\}$ .

Similar to the lncRNA embeddings learning process, drug features were initially extracted by GCN, which layer propagation formula is:

$$X_d = \text{softmax} \left( \hat{A}_d \text{ReLU} \left( \hat{A}_d X' W_d^{(0)} \right) W_d^{(1)} \right) \quad (12)$$

where  $\hat{A}_d = \tilde{D}_d^{-\frac{1}{2}} \tilde{A}_d \tilde{D}_d^{-\frac{1}{2}}$ ,  $\tilde{A}_d = G_d + I$ ,  $\tilde{D}_d$  represents the degree matrix of matrix  $\tilde{A}_d$ , and  $W_d^{(0)}$  and  $W_d^{(1)}$  represents layer-specific weight matrices. Through the above operations, we obtained drug features  $X_d \in R^{n \times r}$ . Further, we learned drug embeddings through graph attention mechanism which input is  $X_d = \{\vec{x}_{d_1}, \vec{x}_{d_2}, \dots, \vec{x}_{d_m}\}$ . Then the attention coefficient between drug  $d_i$  and drug  $d_j$  is:

$$e_{d_i d_j} = a(W_d \vec{x}_{d_i}, W_d \vec{x}_{d_j}) \quad (13)$$

where  $W_d \in R^{r \times c}$  is a parameter matrix. Furthermore, we normalized  $e_{d_i d_j}$  across all choices of  $d_j$  as follows:

$$\alpha_{d_i d_j} = \text{softmax}_{d_j} (e_{d_i d_j}) = \frac{\exp(e_{d_i d_j})}{\sum_{d_k \in N_{d_i}} \exp(e_{d_i d_k})} \quad (14)$$

where  $N_{d_i}$  is the neighbor of node  $d_i$ . The above formula is fully expanded as follows:

$$\alpha_{d_i d_j} = \frac{\exp \left( \text{LeakyReLU} \left( \vec{a}_d^T [W_d \vec{x}_{d_i} \parallel W_d \vec{x}_{d_j}] \right) \right)}{\sum_{d_k \in N_{d_i}} \exp \left( \text{LeakyReLU} \left( \vec{a}_d^T [W_d \vec{x}_{d_i} \parallel W_d \vec{x}_{d_k}] \right) \right)} \quad (15)$$

where  $\vec{a}_d \in R^{2c}$  is a weight vector. Based on this, we obtained the output feature of  $d_i$  as follows:

$$\vec{x}'_{d_i} = \sigma \left( \sum_{d_j \in N(d_i)} \alpha_{d_i d_j} W_d \vec{x}_{d_j} \right) \quad (16)$$

where  $\vec{x}'_{d_i}$  was calculated by the multi-head attention mechanism as follows:

$$\vec{x}'_{d_i} = \parallel_{k=1}^K \sigma \left( \sum_{d_j \in N(d_i)} \alpha_{d_i d_j}^k W_d^k \vec{x}_{d_j} \right) \quad (17)$$

where  $\alpha_{d_i d_j}^k$  is computed by the  $k$ -th head and  $W_d^k$  is the corresponding weight matrix. Specially, the multi-head attention mechanism in the last layer is as follows:

$$\vec{x}'_{d_i} = \sigma \left( \frac{1}{K} \sum_{k=1}^K \sum_{d_j \in N(d_i)} \alpha_{d_i d_j}^k W_d^k \vec{x}_{d_j} \right) \quad (18)$$

Finally, we obtained lncRNA embeddings as  $X'_d \in R^{n \times c} = \{\vec{x}'_{d_1}, \vec{x}'_{d_2}, \dots, \vec{x}'_{d_n}\}$ .

## 2.4. Association prediction

After obtaining lncRNA and drug embeddings, we used them to predict potential lncRNA-drug resistance associations. The association score is equal to  $A' = \sigma(X'_l \times X'_d^T)$  and  $\sigma$  is a sigmoid function. To make the prediction result as close as possible to the real relationship between lncRNA and drug resistance, the reconstruction loss is defined as follows:

$$\text{Loss} = \sum_{i=1}^m \sum_{j=1}^n (A'_{ij} - A_{ij}) \quad (19)$$

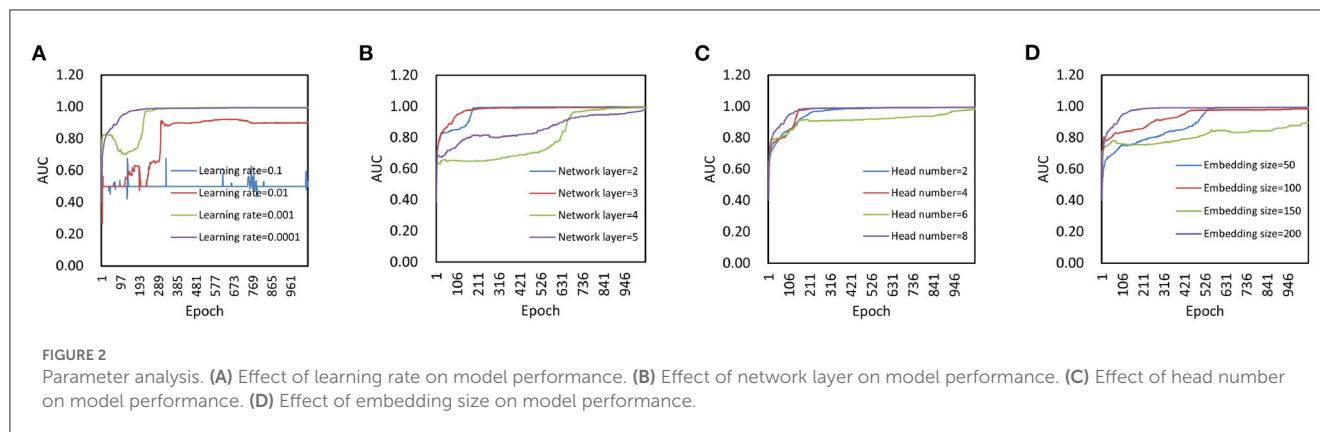
## 3. Results

### 3.1. Parameter analysis

We set the parameters learning rate, network layer, head number, and embedding size in the model as follows. We first changed the learning rate in  $\{0.1, 0.01, 0.001, 0.0001\}$  to determine its effect on model performance (Figure 2A). As we can found, when the learning rate is equal to 0.1, the model is difficult to converge. When the learning rate is in  $\{0.01, 0.001\}$ , the model does not achieve optimal performance. Therefore, 0.0001 is used as the learning rate value. Then, we changed the network layer in  $\{2, 3, 4, 5\}$  to determine its effect on model performance (Figure 2B). It can be found that a small number of layers can speed up the convergence of the model, and a large number of layers will make the model prone to overfitting. In this experiment, we choose 3 as the network layer.

After that, we compared the model performance with different head number in graph attention mechanism module to determine its impact on model performance. Specifically, we changed the head number in  $\{2, 4, 6, 8\}$  for analysis (Figure 2C). As we can see, there is not much difference in final performance between models with different head number. A large number of attention head can speed up the convergence of the model, and a small number of attention head will make the model converge slowly. In this experiment, we make the head number of the attention mechanism equal to 8. Finally, we set the embedding size in  $\{10, 50, 100, 200\}$  to verify its impact on model prediction performance (Figure 2D). When the





embedding size is set to 200, the final AUC is slightly larger than that of the other groups. Thus, we choose 200 as the embedding size.

### 3.2. Performance evaluation

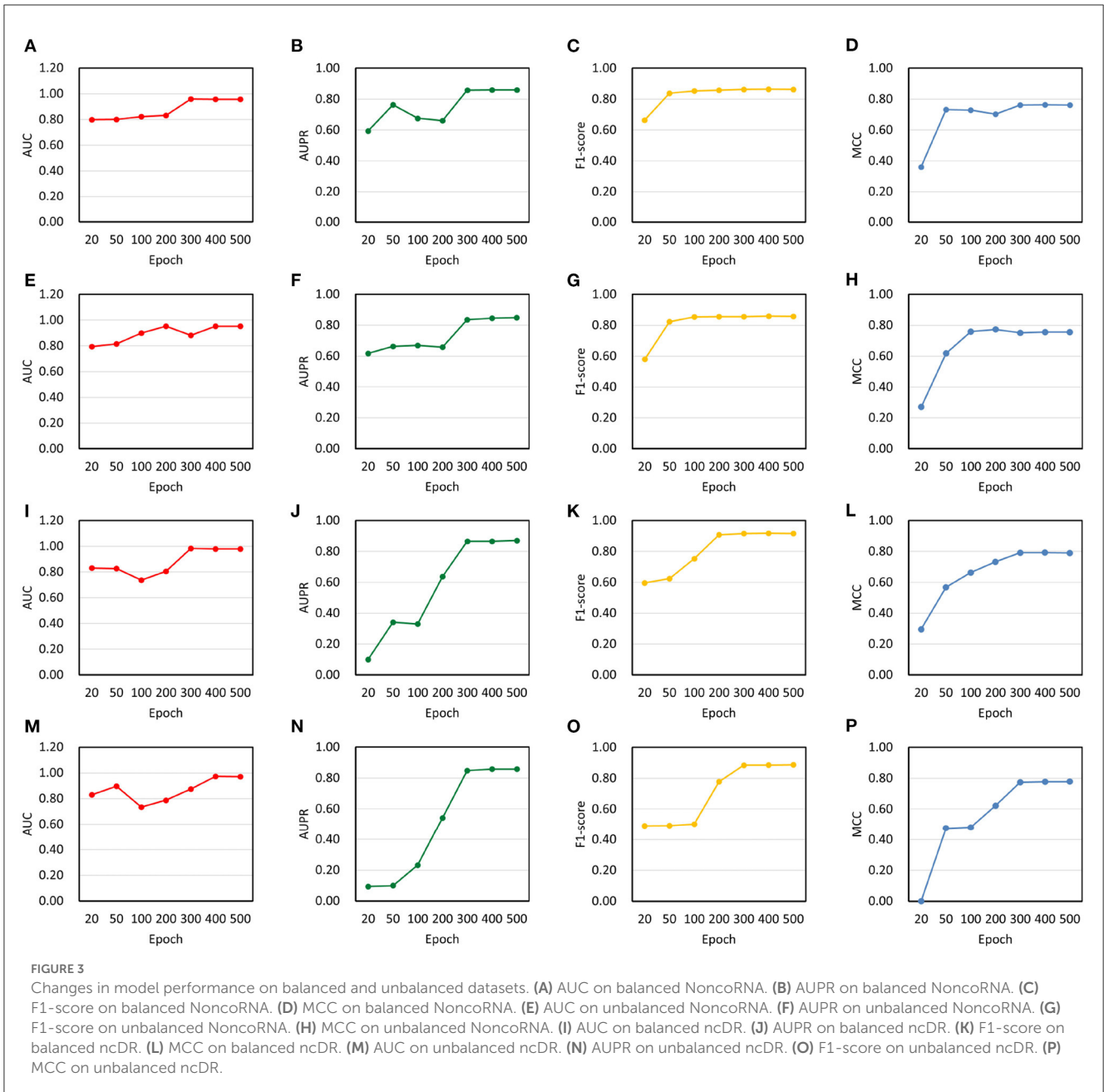
In order to evaluate model performance, we analyzed changes in AUC, AUPR, F1-score and MCC on NoncoRNA and ncDR. As a result, four experimental groups were obtained, including balanced NoncoRNA, balanced ncDR, unbalanced NoncoRNA, and unbalanced ncDR (Figure 3). It can be found that the AUC, AUPR, F1-score and MCC of DeepLDA are stable at 0.96, 0.86, 0.86, and 0.76, respectively, on balanced NoncoRNA (Figures 3A–D). On unbalanced NoncoRNA, the AUC, AUPR, F1 score and MCC of DeepLDA are stable at 0.95, 0.85, 0.86, and 0.76, respectively (Figures 3E–H). In addition, the AUC, AUPR, F1-score and MCC of DeepLDA are stable at 0.98, 0.87, 0.92, and 0.79, respectively, on balanced ncDR (Figures 3I–L). On unbalanced ncDR, the AUC, AUPR, F1 score and MCC of DeepLDA are stable at 0.97, 0.86, 0.89, and 0.78, respectively (Figures 3M–P). The results show that the model performance on the balanced dataset is slightly better than that on the unbalanced dataset. The phenomenon is caused by the proportion of samples in the dataset. Since the number of positive and negative samples in the balanced datasets is the same, the bias caused by unbalanced samples can be eliminated to a certain extent in the prediction task. Thus, our method performs better on balanced datasets than on unbalanced datasets. For unbalanced datasets, although the prediction performance is slightly inferior to the balanced datasets, it is still at a high level, which proves the robustness of our model.

We further calculated average performance metrics, which is equal to the average value of AUC, AUPR, F1-score and MCC (Table 2). It can be found that when the epoch is around 300, the average metrics are close to convergence on the balanced dataset, and when the epoch is around 400, the average metrics are close to convergence on the unbalanced dataset. Specifically, the average metrics on NoncoRNA- and ncDR-related balanced datasets are stable at 0.86 and 0.89, respectively, and the average metrics on NoncoRNA- and ncDR-related unbalanced datasets are stable at 0.85 and 0.87, respectively. The above experimental results demonstrate that our model has satisfactory performance in predicting lncRNA-drug resistance associations and converges

faster on balanced datasets than on unbalanced datasets since the data size of balanced datasets is smaller than that of unbalanced datasets. Moreover, the results further verify that the model performance better on the balanced dataset than on the unbalanced dataset, and that the model can also achieve ideal performance on unbalanced datasets.

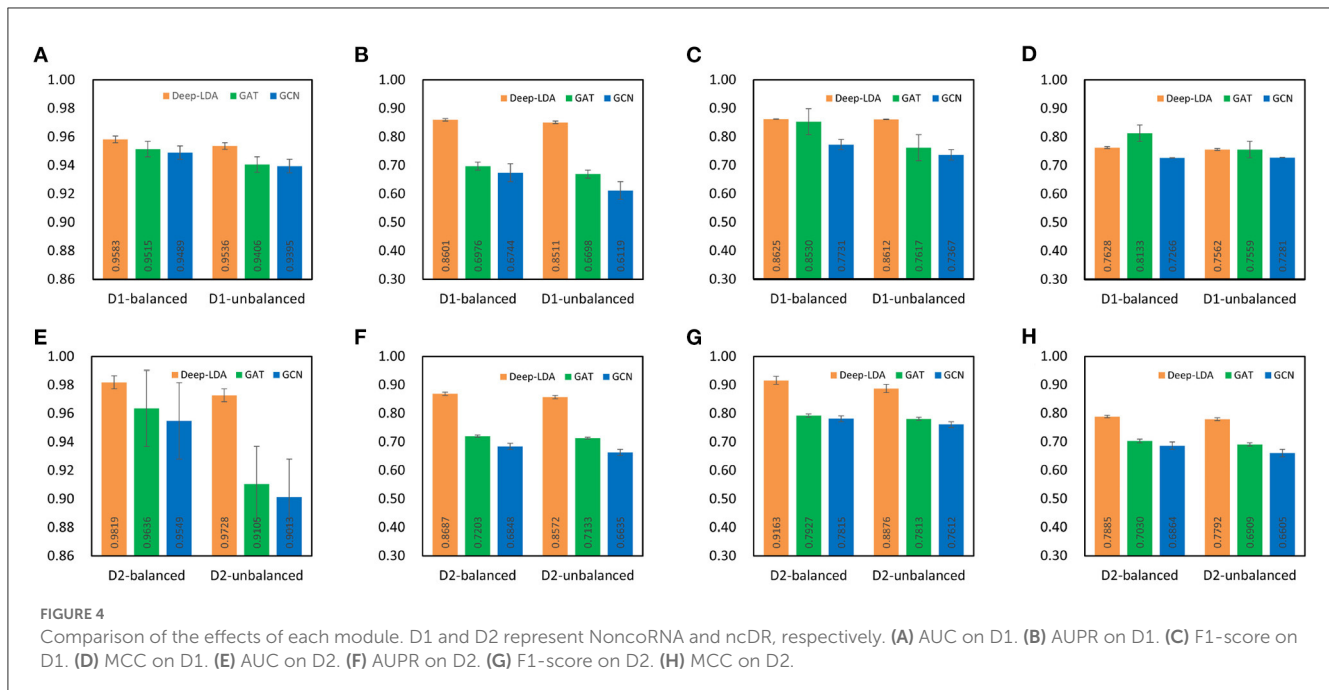
### 3.3. Effect of each module

To demonstrate the effectiveness of GCN and GAT on the lncRNA-drug resistance association prediction task, we compared the performance of GCN, GAT, and DeepLDA under the same experimental setting. As a result, we find that DeepLDA outperforms GAT and GCN on the given datasets (Figure 4). The AUC, AUPR, F1-score, and MCC of DeepLDA are 0.9583, 0.8601, 0.8625, and 0.7628, respectively, on balanced NoncoRNA (Figures 4A–D), are 0.9536, 0.8511, 0.8612, and 0.7562, respectively, on unbalanced NoncoRNA (Figures 4A–D), are 0.9819, 0.8687, 0.9163, and 0.7885, respectively, on balanced ncDR (Figures 4E–H), and are 0.9728, 0.8572, 0.8876, and 0.7792, respectively, on unbalanced ncDR (Figures 4E–H). Through comparative analysis, it can be found that the AUCs of DeepLDA on balanced NoncoRNA, unbalanced NoncoRNA, balanced ncDR, and unbalanced ncDR are 0.0071, 0.0138, 0.0190, and 0.0684 higher than the best AUC in GAT and GCN, respectively. The AUPRs of DeepLDA are 0.2329, 0.2707, 0.2060, and 0.2017 higher than the best AUPR in GAT and GCN on balanced NoncoRNA, unbalanced NoncoRNA, balanced ncDR, and unbalanced ncDR, respectively. The F1-scores of DeepLDA are 0.0111, 0.1306, 0.1559, 0.1361 higher than the best F1-score in GAT and GCN on balanced NoncoRNA, unbalanced NoncoRNA, balanced ncDR, and unbalanced ncDR, respectively. The MCCs of DeepLDA is 0.0004, 0.1216, and 0.1278, higher than the best MCC in GAT and GCN on unbalanced NoncoRNA, balanced ncDR, and unbalanced ncDR, respectively. At the same time, the MCC of DeepLDA is higher than GCN and slightly lower than GAT on balanced NoncoRNA. As for the average performance, which are 0.8609, 0.8555, 0.8889, and 0.8742 on balanced NoncoRNA, unbalanced NoncoRNA, balanced ncDR, and unbalanced ncDR, respectively (Table 3). Compared with the best average performance in GAT



**TABLE 2** Changes in average model performance on balanced and unbalanced datasets.

Epoch	NoncoRNA-balanced	NoncoRNA-unbalanced	ncDR-balanced	ncDR-unbalanced
20	0.6042	0.5660	0.4553	0.3535
50	0.7835	0.7301	0.5897	0.4908
100	0.7696	0.7959	0.6204	0.4872
200	0.7634	0.8099	0.7707	0.6825
300	0.8608	0.8315	0.8889	0.8454
400	0.8617	0.8534	0.8890	0.8736
500	0.8605	0.8538	0.8890	0.8734



**TABLE 3** Comparison of average performance metrics of each module.

Method	Balanced noncoRNA	Unbalanced noncoRNA	Balanced ncDR	Unbalanced ncDR
Deep-LDA	0.8609	0.8555	0.8889	0.8742
GAT	0.8289	0.7820	0.7949	0.7740
GCN	0.7808	0.7541	0.7769	0.7466

and GCN, the improvements are 0.0387, 0.0940, 0.1182, and 0.1295, respectively.

Overall, DeepLDA has an advantage over GAT and GCN in predicting lncRNA-drug resistance associations because DeepLDA combines the feature learning capabilities of GCN and GAT to capture the local and global features of nodes effectively. In addition, GAT has a strong learning ability and generalization ability, and can handle complex and variable lncRNA-drug resistance association data only by learning association-related nodes and their neighbor information, which significantly improves the model prediction performance.

### 3.4. Comparison with other methods

We compared the performance of DeepLDA with other association prediction methods, GSLRDA (Zheng J. et al., 2022), LRGPND (Li et al., 2021), and GCMR (Huang et al., 2020), to verify its effectiveness. Among them, GSLRDA and LRGPND were designed to predict ncRNA-drug resistance associations, and GCMR was designed to predict miRNA-drug resistance associations. As a result, these methods perform better in predicting lncRNA-drug resistance associations on balanced datasets than on unbalanced datasets, as we expected (Figure 5). For dataset NoncoRNA and ncDR, DeepLDA performs better on ncDR

than NoncoRNA (Figure 6). As we can see in Figures 5, 6, DeepLDA outperforms other prediction methods in AUC, AUPR, F1-score, and MCC. The AUC of DeepLDA is 0.0321, 0.0538, 0.0434, and 0.0470, better than the second-best method on balanced NoncoRNA, balanced ncDR, unbalanced NoncoRNA and unbalanced ncDR, respectively. The AUPR of DeepLDA is 0.0976, 0.0389, 0.1773, and 0.0388, better than the second-best method on balanced NoncoRNA, balanced ncDR, unbalanced NoncoRNA, and unbalanced ncDR, respectively. The F1-score of DeepLDA is 0.0887, 0.1092, 0.1389, and 0.1036, better than the second-best method on balanced NoncoRNA, balanced ncDR, unbalanced NoncoRNA and unbalanced ncDR, respectively. The MCC of DeepLDA is 0.0304 and 0.0648 better than the second-best method on balanced NoncoRNA and unbalanced NoncoRNA, respectively. At the same time, the MCC of DeepLDA is slightly inferior to the second-best method on balanced ncDR and unbalanced ncDR, respectively.

To further demonstrate the effectiveness of DeepLDA, we compared the average performance of different prediction methods (Table 4). As we can see in Table 4, the average performance of GCMR on balanced NoncoRNA, unbalanced NoncoRNA, balanced ncDR, and unbalanced ncDR are 0.8094, 0.7710, 0.8149, and 0.7885 respectively. The average performance of GSLRDA on balanced NoncoRNA, unbalanced NoncoRNA, balanced ncDR, and unbalanced ncDR are 0.7800, 0.7591, 0.8468, and 0.8303, respectively. As for LRGPND, its average

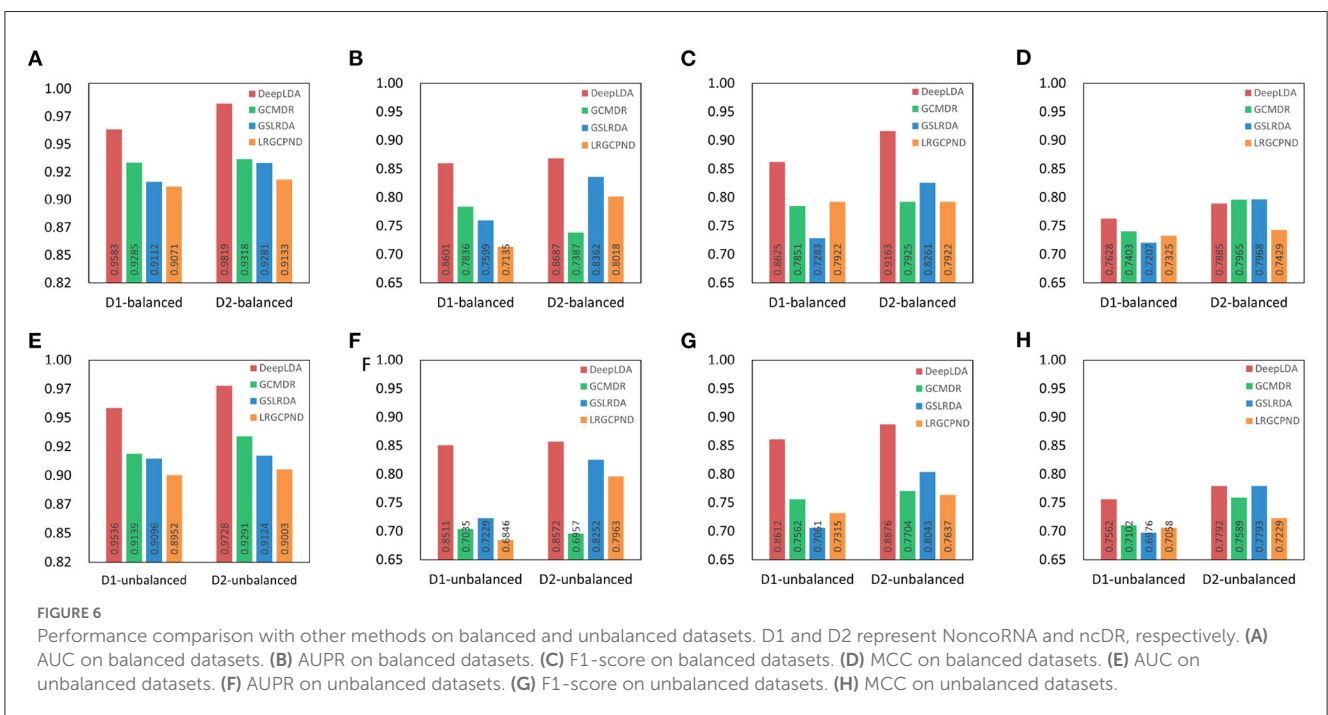
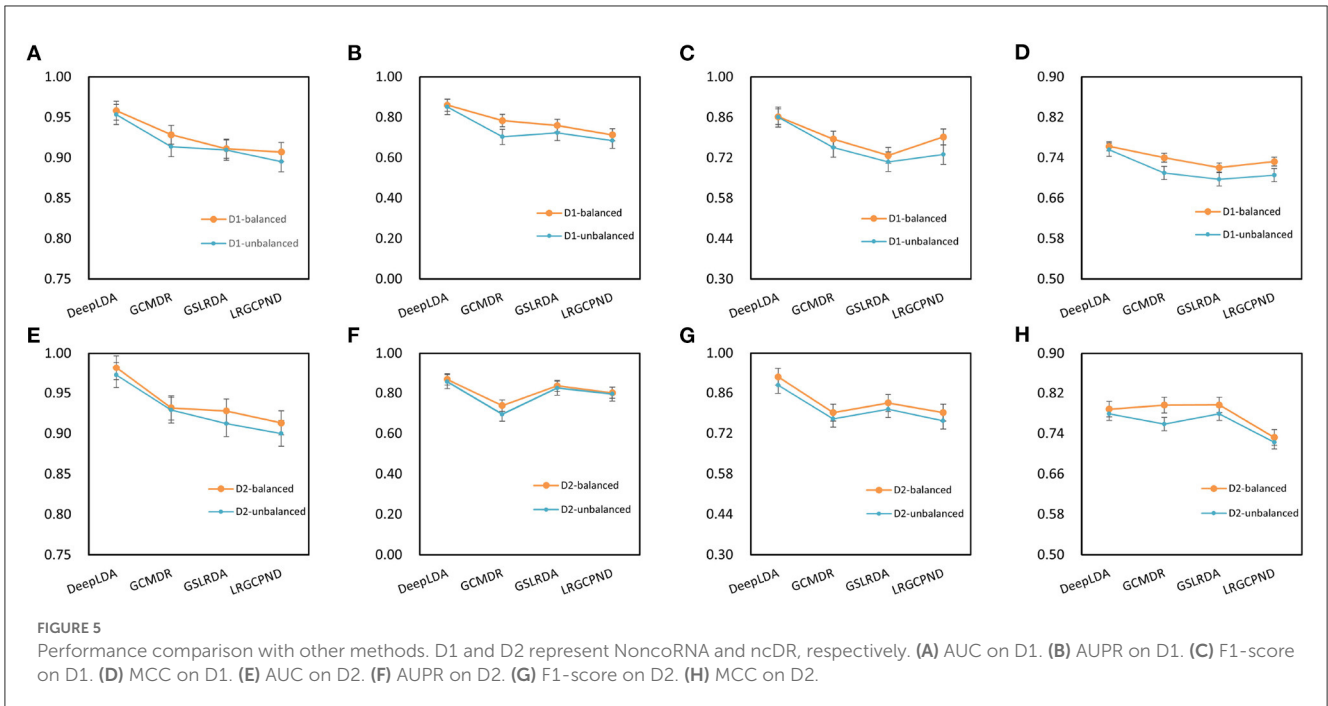


TABLE 4 Comparison of average performance metrics of different methods.

Method	Balanced noncoRNA	Unbalanced noncoRNA	Balanced ncDR	Unbalanced ncDR
Deep-LDA	0.8609	0.8555	0.8889	0.8742
GCMR	0.8094	0.7710	0.8149	0.7885
GSLRDA	0.7800	0.7591	0.8468	0.8303
LRGCPND	0.7863	0.7543	0.8126	0.7958



performance on balanced NoncoRNA, unbalanced NoncoRNA, balanced ncDR, and unbalanced ncDR are 0.7863, 0.7543, 0.8126, and 0.7958, respectively. Comparative analysis shows that the average performance of DeepLDA on balanced NoncoRNA, unbalanced NoncoRNA, balanced ncDR, and unbalanced ncDR are 0.0637, 0.1097, 0.0497, and 0.0529 better than the suboptimal method, respectively.

Overall, the performance of DeepLDA is significantly superior to GSLRDA, LRGCPND, and GCMDR. The performance advantage of DeepLDA is attributed to the following two points. First, GCN performs end-to-end learning of feature information and structural information of lncRNAs and drugs, which can comprehensively capture global information and represent node features well. Second, the graph attention mechanism aggregates the neighbor information of nodes according to the attention coefficient to obtain its embedding, which can efficiently represent local neighbor information.

## 4. Discussion

lncRNA plays an important role in carcinogenesis and can lead to the resistance of tumor cells to chemotherapeutic drugs (Wang et al., 2017; Ashrafizaveh et al., 2021), which is an essential factor leading to high cancer mortality (Vasan et al., 2019). Therefore, identifying lncRNA-drug resistance associations becomes crucial for revealing the impact of lncRNAs on the drug resistance of tumor cells. Recently, machine learning has achieved promising results in predicting biomolecular association. However, to the best of our knowledge, little work has been done on machine learning to predict lncRNA-drug resistance associations. In this study, we proposed DeepLDA, a powerful deep learning model based on graph neural network and graph attention mechanism for revealing the potential relationship between lncRNAs and drug resistance. DeepLDA first used known association items to construct similarity networks of lncRNAs and drugs. Subsequently, deep graph neural networks were used for preliminary learning of lncRNA and drug features. Finally, these learned features were input to GAT to learn lncRNA and drug embeddings to predict potential association pairs. Experimental results show that DeepLDA outperforms other machine learning methods in predicting lncRNA-drug resistance pairs on the given datasets. In summary, our proposed a computational model, DeepLDA, that can effectively complete the prediction task of the association between lncRNAs and drug resistance. DeepLDA can provide valuable insights for drug design and open new avenues for lncRNA-related research.

On the one hand, the association between miRNA/circRNA and drug resistance in cancer cells has been confirmed (Leonetti et al., 2019; Xu et al., 2020; Pan et al., 2021; Wang et al., 2022). DeepLDA provides a reference for the study of the relationship between miRNA/circRNA and drug resistance, and the prediction process is as follows. Firstly, GCN learns the features of miRNAs/circRNAs and drugs, which can comprehensively capture global information and represent node features well. Subsequently, the graph attention mechanism aggregates the neighbor features of nodes according to the attention coefficient to obtain their embeddings. Finally, the learned embeddings can be used to effectively predict potential

associations between miRNAs/circRNAs and drug resistance. On the other hand, DeepLDA can facilitate the development of targeted therapies. Despite recent discoveries in cancer treatment, the study of resistance to chemotherapy, radiation therapy, targeted therapy, and immunotherapy remains a major challenge. lncRNAs are widely recognized as universal regulators of multiple cancers, such as proliferation, apoptosis, invasion, metastasis, and genome instability (He et al., 2021; Nandwani et al., 2021). Based on this, lncRNAs can be used as therapeutic adjuvants and components of tumor-agnostic therapeutic strategies to improve anticancer responses to existing treatment modalities. Therefore, lncRNA-based targeted therapy can be developed on the basis of DeepLDA to intervene in lncRNA drug resistance.

Despite DeepLDA has significant advantages in predicting potential lncRNA-drug resistance associations, its limitations should be informed. The known association matrix remains sparse. In future work, we will collect more lncRNA-drug resistance associations and employ other feature learning methods to better explore the feature information of nodes to improve the model performance. In addition, we will combine other biological characteristics to carry out lncRNA-drug resistance association prediction analysis to effectively mine the regulatory mechanism of lncRNA.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

MG designed and implemented the method. MG and XS wrote this manuscript. Both authors contributed to the article and approved the submitted version.

## Funding

Thanks for the support of the National Natural Science Foundation of China (grant numbers 61772426 and U1811262).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Ashrafzavah, S., Ashrafzadeh, M., Zarrabi, A., Husmandi, K., Zabolian, A., Shahinozaman, M., et al. (2021). Long non-coding RNAs in the doxorubicin resistance of cancer cells. *Cancer Lett.* 508, 104–114. doi: 10.1016/j.canlet.2021.03.018
- Barth, D. A., Juracek, J., Slaby, O., Pichler, M., and Calin, G. A. (2020). lncRNA and mechanisms of drug resistance in cancers of the genitourinary system. *Cancers* 12, 2148. doi: 10.3390/cancers12082148
- Bester, A. C., Lee, J. D., Chavez, A., Lee, Y.-R., Nachmani, D., Vora, S., et al. (2018). An integrated genome-wide CRISPR approach to functionalize lncRNAs in drug resistance. *Cell* 173, 649–664. doi: 10.1016/j.cell.2018.03.052
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R., Torre, L., Jemal, A., et al. (2020). Erratum: Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 70, 313. doi: 10.3322/caac.21609
- Bridges, M. C., Daulagala, A. C., and Kourtidis, A. (2021). Lncation: lncRNA localization and function. *J. Cell Biol.* 220, e202009045. doi: 10.1083/jcb.202009045
- Dai, E., Yang, F., Wang, J., Zhou, X., Song, Q., An, W., et al. (2017). ncdR: a comprehensive resource of non-coding RNAs involved in drug resistance. *Bioinformatics* 33, 4010–4011. doi: 10.1093/bioinformatics/btx523
- Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., et al. (2021). Cancer statistics for the year 2020: an overview. *Int. J. Cancer* 149, 778–789. doi: 10.1002/ijc.33588
- Gao, M., Guo, Y., Xiao, Y., and Shang, X. (2021). Comprehensive analyses of correlation and survival reveal informative lncRNA prognostic signatures in colon cancer. *World J. Surg. Oncol.* 19, 1–15. doi: 10.1186/s12957-021-02196-4
- Gao, M., Liu, S., Qi, Y., Guo, X., and Shang, X. (2022a). Gae-lga: integration of multi-omics data with graph autoencoders to identify lncRNA-pg associations. *Brief. Bioinform.* 23, bbac452. doi: 10.1093/bib/bbac452
- Gao, M., Liu, S., Qi, Y., Guo, X., and Shang, X. (2022b). ImreLnc: Identifying immune-related lncRNA characteristics in human cancers based on heuristic correlation optimization. *Front. Genet.* 12, 2768. doi: 10.3389/fgene.2021.792541
- Gao, M., and Shang, X. (2022a). “Identification of lncRNA-related protein-coding genes using multi-omics data based on deep learning and matrix completion,” in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Las Vegas, NV: IEEE Computer Society), 3307–3314.
- Gao, M.-H., and Shang, X.-Q. (2022b). Artificial intelligence-based prediction for cancer susceptibility, recurrence and survival. *Progr. Biochem. Biophys.* 49, 1687–1702. doi: 10.16476/j.pibb.2021.0334
- He, J., Zhu, S., Liang, X., Zhang, Q., Luo, X., Liu, C., et al. (2021). lncRNA as a multifunctional regulator in cancer multi-drug resistance. *Mol. Biol. Rep.* 48, 1–15. doi: 10.1007/s11033-021-06603-7
- Hu, Y.-B., Yan, C., Mu, L., Mi, Y.-L., Zhao, H., Hu, H., et al. (2019). Exosomal wnt-induced dedifferentiation of colorectal cancer cells contributes to chemotherapy resistance. *Oncogene* 38, 1951–1965. doi: 10.1038/s41388-018-0557-91
- Huang, Y.-A., Hu, P., Chan, K. C., and You, Z.-H. (2020). Graph convolution for predicting associations between miRNA and drug resistance. *Bioinformatics* 36, 851–858. doi: 10.1093/bioinformatics/btz621
- Jiang, M.-C., Ni, J.-J., Cui, W.-Y., Wang, B.-Y., and Zhuo, W. (2019). Emerging roles of lncRNA in cancer and therapeutic opportunities. *Am. J. Cancer Res.* 9, 1354.
- Jiang, W., Xia, J., Xie, S., Zou, R., Pan, S., Wang, Z.-w., et al. (2020). Long non-coding RNAs as a determinant of cancer drug resistance: Towards the overcoming of chemoresistance via modulation of lncRNAs. *Drug Resist. Updates* 50, 100683. doi: 10.1016/j.drug.2020.100683
- Koch, L. (2017). Screening for lncRNA function. *Nat. Rev. Genet.* 18, 70–70. doi: 10.1038/nrg.2016.168
- Leonetti, A., Assaraf, Y. G., Veltsista, P. D., El Hassouni, B., Tiseo, M., and Giovannetti, E. (2019). MicroRNAs as a drug resistance mechanism to targeted therapies in EGFR-mutated NSCLC: Current implications and future directions. *Drug Resist. Updates* 42, 1–11. doi: 10.1016/j.drug.2018.11.002
- Li, L., Wu, P., Wang, Z., Meng, X., Zha, C., Li, Z., et al. (2020). NonCorna: a database of experimentally supported non-coding RNAs and drug targets in cancer. *J. Hematol. Oncol.* 13, 1–4. doi: 10.1186/s13045-020-00849-7
- Li, Y., Wang, R., Zhang, S., Xu, H., and Deng, L. (2021). LrgcPnd: predicting associations between ncRNA and drug resistance via linear residual graph convolution. *Int. J. Mol. Sci.* 22, 10508. doi: 10.3390/ijms221910508
- Liu, H., Ren, G., Chen, H., Liu, Q., Yang, Y., and Zhao, Q. (2020). Predicting lncRNA-miRNA interactions based on logistic matrix factorization with neighborhood regularized. *Knowl. Based Syst.* 191, 105261. doi: 10.1016/j.knsys.2019.105261
- Liu, X., Huang, Z., Qian, W., Zhang, Q., and Sun, J. (2019). Silence of lncRNA uca1 rescues drug resistance of cisplatin to non-small-cell lung cancer cells. *J. Cell. Biochem.* 120, 9243–9249. doi: 10.1002/jcb.28200
- Long, Y., Wang, X., Youmans, D. T., and Cech, T. R. (2017). How do lncRNAs regulate transcription? *Sci. Adv.* 3, ea02110. doi: 10.1126/sciadv.a02110
- Mattick, J. S., and Rinn, J. L. (2015). Discovery and annotation of long noncoding RNAs. *Nat. Struct. Mol. Biol.* 22, 5–7. doi: 10.1038/nsmb.2942
- Nandwani, A., Rathore, S., and Datta, M. (2021). lncRNAs in cancer: regulatory and therapeutic implications. *Cancer Lett.* 501, 162–171. doi: 10.1016/j.canlet.2020.11.048
- Niu, Y., Song, C., Gong, Y., and Zhang, W. (2022). MiRNA-drug resistance association prediction through the attentive multimodal graph convolutional network. *Front. Pharmacol.* 12, 3997. doi: 10.3389/fphar.2021.799108
- Pan, G., Liu, Y., Shang, L., Zhou, F., and Yang, S. (2021). EMT-associated miRNAs and their roles in cancer stemness and drug resistance. *Cancer Commun.* 41, 199–217. doi: 10.1002/cac2.12138
- Peng, W.-X., Koirala, P., and Mo, Y.-Y. (2017). lncRNA-mediated regulation of cell signaling in cancer. *Oncogene* 36, 5661–5667. doi: 10.1038/ncr.2017.184
- Rebucci, M., and Michiels, C. (2013). Molecular aspects of cancer cell resistance to chemotherapy. *Biochem. Pharmacol.* 85, 1219–1226. doi: 10.1016/j.bcp.2013.02.017
- Rehman, S. K., Haynes, J., Collignon, E., Brown, K. R., Wang, Y., Nixon, A. M., et al. (2021). Colorectal cancer cells enter a diapause-like dtp state to survive chemotherapy. *Cell* 184, 226–242. doi: 10.1016/j.cell.2020.11.018
- Shi, Q., Li, Y., Li, S., Jin, L., Lai, H., Wu, Y., et al. (2020). lncRNA dila1 inhibits cyclin D1 degradation and contributes to tamoxifen resistance in breast cancer. *Nat. Commun.* 11, 5513. doi: 10.1038/s41467-020-19349-w
- Singh, D., Assaraf, Y. G., and Gacche, R. N. (2022). Long non-coding RNA mediated drug resistance in breast cancer. *Drug Resist. Updates* 63, 100851. doi: 10.1016/j.drug.2022.100851
- Sun, R., Wang, R., Chang, S., Li, K., Sun, R., Wang, M., et al. (2019). Long non-coding RNA in drug resistance of non-small cell lung cancer: a mini review. *Front. Pharmacol.* 10, 1457. doi: 10.3389/fphar.2019.01457
- Vasan, N., Baselga, J., and Hyman, D. M. (2019). A view on drug resistance in cancer. *Nature* 575, 299–309. doi: 10.1038/s41586-019-1730-1
- Wang, H., Guan, Z., He, K., Qian, J., Cao, J., and Teng, L. (2017). lncRNA ucal in anti-cancer drug resistance. *Oncotarget* 8, 64638. doi: 10.18632/oncotarget.18344
- Wang, X., Zhang, J., Cao, G., Hua, J., Shan, G., and Lin, W. (2022). Emerging roles of circular RNAs in gastric cancer metastasis and drug resistance. *J. Exp. Clin. Cancer Res.* 41, 1–13. doi: 10.1186/s13046-022-02432-z
- Xia, C., Dong, X., Li, H., Cao, M., Sun, D., He, S., et al. (2022). Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin. Med. J.* 135, 584–590. doi: 10.1097/CM9.00000000000002108
- Xu, T., Wang, M., Jiang, L., Ma, L., Wan, L., Chen, Q., et al. (2020). CircRNAs in anticancer drug resistance: recent advances and future potential. *Mol. Cancer* 19, 1–20. doi: 10.1186/s12943-020-01240-3
- Xue, X., Yang, Y. A., Zhang, A., Fong, K., Kim, J., Song, B., et al. (2016). lncRNA hotair enhances ER signaling and confers tamoxifen resistance in breast cancer. *Oncogene* 35, 2746–2755. doi: 10.1038/ncr.2015.340
- Zhang, L., Liu, T., Chen, H., Zhao, Q., and Liu, H. (2021a). Predicting lncRNA-miRNA interactions based on interactome network and graphlet interaction. *Genomics* 113, 874–880. doi: 10.1016/j.ygeno.2021.02.002
- Zhang, L., Yang, P., Feng, H., Zhao, Q., and Liu, H. (2021b). Using network distance analysis to predict lncRNA-miRNA interactions. *Interdiscipl. Sci. Comp. Life Sci.* 13, 535–545. doi: 10.1007/s12539-021-00458-z
- Zhang, Y., Yi, T., Ji, H., Zhao, G., Xi, Y., Dong, C., et al. (2020). Designing a general method for predicting the regulatory relationships between long noncoding RNAs and protein-coding genes based on multi-omics characteristics. *Bioinformatics* 36, 2025–2032. doi: 10.1093/bioinformatics/btz886
- Zhao, T., Hu, Y., Peng, J., and Cheng, L. (2020). DeepLpP: a novel deep learning method for prioritizing lncRNA target genes. *Bioinformatics* 36, 4466–4472. doi: 10.1093/bioinformatics/btaa428
- Zhao, Z., Guo, Y., Liu, Y., Sun, L., Chen, B., Wang, C., et al. (2021). Individualized lncRNA differential expression profile reveals heterogeneity of breast cancer. *Oncogene* 40, 4604–4614. doi: 10.1038/s41388-021-01883-6
- Zheng, J., Qian, Y., He, J., Kang, Z., and Deng, L. (2022). Graph neural network with self-supervised learning for noncoding RNA-drug resistance association prediction. *J. Chem. Inf. Model.* 62, 3676–3684. doi: 10.1021/acs.jcim.2c00367
- Zheng, K., Zhao, H., Zhao, Q., Wang, B., Gao, X., and Wang, J. (2022). NasMDR: a framework for miRNA-drug resistance prediction using efficient neural architecture search and graph isomorphism networks. *Brief. Bioinform.* 23, bbac338. doi: 10.1093/bib/bbac338
- Zhou, M., Liu, L., Wang, J., and Liu, W. (2022). The role of long noncoding RNAs in therapeutic resistance in cervical cancer. *Front. Cell Dev. Biol.* 10, 1060909. doi: 10.3389/fcell.2022.1060909
- Zhu, C., Wang, X., Wang, Y., and Wang, K. (2022). Functions and underlying mechanisms of lncRNA hotair in cancer chemotherapy resistance. *Cell Death Discov.* 8, 1–10. doi: 10.1038/s41420-022-01174-3