




# DeepWalk-aware graph attention networks with CNN for circRNA–drug sensitivity association identification

Guanghui Li , Youjun Li, Cheng Liang  and Jiawei Luo 

Corresponding authors: Guanghui Li, School of Information Engineering, East China Jiaotong University, Nanchang 330013, China. Tel.: +86-0791-87046245; E-mail: ghli16@hnu.edu.cn, Jiawei Luo, College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China. Tel.: +86-0731-88821971; E-mail: luojiawei@hnu.edu.cn

## Abstract

Circular RNAs (circRNAs) are a class of noncoding RNA molecules that are widely found in cells. Recent studies have revealed the significant role played by circRNAs in human health and disease treatment. Several restrictions are encountered because forecasting prospective circRNAs and medication sensitivity connections through biological research is not only time-consuming and expensive but also incredibly ineffective. Consequently, the development of a novel computational method that enhances both the efficiency and accuracy of predicting the associations between circRNAs and drug sensitivities is urgently needed. Here, we present DGATCCDA, a computational method based on deep learning, for circRNA–drug sensitivity association identification. In DGATCCDA, we first construct multimodal networks from the original feature information of circRNAs and drugs. After that, we adopt DeepWalk-aware graph attention networks to sufficiently extract feature information from the multimodal networks to obtain the embedding representation of nodes. Specifically, we combine DeepWalk and graph attention network to form DeepWalk-aware graph attention networks, which can effectively capture the global and local information of graph structures. The features extracted from the multimodal networks are fused by layer attention, and eventually, the inner product approach is used to construct the association matrix of circRNAs and drugs for prediction. The ultimate experimental results obtained under 5-fold cross-validation settings show that the average area under the receiver operating characteristic curve value of DGATCCDA reaches 91.18%, which is better than those of the five current state-of-the-art calculation methods. We further guide a case study, and the excellent obtained results also show that DGATCCDA is an effective computational method for exploring latent circRNA–drug sensitivity associations.

**Keywords:** circRNA–drug associations; DeepWalk; attention mechanism; global information

## INTRODUCTION

Most frequently seen in eukaryotic cells, circular RNAs (circRNAs) are a family of single-stranded closed-loop noncoding RNA molecules with certain tissue specializations and temporal sequences that are created by the reverse splicing of mRNA precursors (pre-mRNA) during genome transcription [1, 2]. CircRNA lacks a poly(A) tail at the 3' ends and a cap structure at the 5' ends in comparison with ordinary linear RNA [3]. For continuous and steady expression in cells, its distinctive structure renders it highly stable and resistant to nucleic acid exonuclease destruction [4]. It has been discovered that a large number of circRNAs that exhibit abnormal expression in tissues and cells have the capacity to modify the cell cycle or control cell division and apoptosis. For instance, Xie *et al.* [5] discovered that overexpression of hsa\_circ\_0006470 effectively triggered cell cycle arrest and reduced the proliferation, migration and viability of gastric cancer cells, preventing the development of stomach cancer. Furthermore, Wang [6] hypothesized that the miR-520 h/CDC42 axis may be a mechanism by which circ\_SKA3 controls cell cycle progression, invasion, colony formation, apoptosis and migration. Additionally, circRNAs may influence biological processes by controlling transcription, acting as miRNA sponges, interacting with proteins or in certain situations

autotranslating [7, 8]. According to recent studies, circRNAs are said to have a considerable impact on how sensitive cells are to drugs. For instance, circSMARCA5 enhances the chemosensitivity of human breast cancer cells to bleomycin [9] and cisplatin. CircCDYL, an autophagy-associated protein, plays a significant role in promoting the malignant progression of breast cancer cells. Moreover, it has been observed to diminish the clinical response to chemotherapy in breast cancer patients [10], and accumulating evidence strongly supports the association between circRNAs and tumor chemoresistance, suggesting their potential significance in this context [11]. As shown by the high expression of circAKT3 in cisplatin-resistant gastric cancer cells and the promotion of paclitaxel resistance by circ-PVT1 in gastric cancer cells [12], identifying the sensitivity associations between circRNAs and drugs is crucial for circRNA-based drug discovery and disease treatment.

A computational method for circRNA–drug sensitivity association identification is urgently needed because there are so few studies in this field, and traditional biological experiments used to investigate these relationships are time- and money-consuming. Deng *et al.* [13] advanced a computational approach for the circRNA–drug sensitivity prediction. The GATECDA model proposed by Deng *et al.* calculates the similarity of circRNAs using

Guanghui Li is an associate professor at East China Jiaotong University. His research interests include computational biology and deep learning.

Youjun Li is a graduate student of East China Jiaotong University. His research interests include machine learning and bioinformatics.

Cheng Liang is an associate professor at Shandong Normal University. Her research interests include bioinformatics and deep learning.

Jiawei Luo is a professor at Hunan University. Her research interests include complex network and machine learning.

Received: July 21, 2023. Revised: September 26, 2023. Accepted: November 20, 2023

© The Author(s) 2023. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

the Gaussian interaction profile kernel (GIP) kernel similarity and the sequence similarity of the host genes. A drug similarity matrix is calculated by the structural similarity of drugs and GIP kernel similarity, after which the GATE graph attention-based self-encoder is utilized to extract features; finally, a multilayer perceptron is employed to explore the relationships among circRNA and drug sensitivities using the feature representations of circRNAs and drugs. The results of several experiments have demonstrated the effectiveness of GATECDA in identifying potential associations. To our best knowledge, GATECDA represents the pioneering computational approach for predicting circRNA–drug sensitivity associations. Following this, Yang et al. [14] also proposed a computational-based approach, MNGACDA, to predict the sensitivity associations of circRNAs and drugs. MNGACDA first uses a multimodal strategy to extract features from circRNA similarity networks, drug similarity networks and association networks of circRNAs and drugs, respectively, and then it uses a multilayer multihead graph attention network (GAT) [15] with a residual module to extract information from the original features. According to the final trial results, MNGACDA is more successful in terms of circRNA–drug sensitivity prediction and has demonstrated its efficacy. However, only a few highly advanced computational methods focus on this field, and the information that is currently available is insufficient. As a result, there is an urgent need to develop computational methods that are both more accurate and more efficient. Recently, with the development of graph neural networks, there are more and more computational methods successfully applied in various association prediction fields. The successful application of deep learning in these fields, such as circRNA–disease association prediction [16–19] and miRNA–disease association prediction [20–22], has significant inspirations to our current research. To do this, we advance an approach based on DeepWalk-aware graph attention networks. We first build the similarity network of circRNAs based on GIP kernel similarity and the sequence data of circRNAs, build the similarity network of drugs based on the structural similarity and GIP kernel similarity, and build the circRNA–drug network based on known circRNA–drug associations. We separately perform feature extraction via DeepWalk-aware graph attention networks for each network. We extract the topological features of each network using the DeepWalk method [23], update the adjacency matrix of each network according to the DeepWalk features, replace the original adjacency matrix with the updated adjacency matrix, and then apply the multilayer GAT to extract both similarity features and DeepWalk features. We utilize an attention convolution neural network (CNN) to fuse the GAT views layers to obtain fused embedding representations. Finally, layer attention is used to fuse the features of the association network and the features of the similarity network to obtain the final embedding representations of circRNAs and drugs, and the final feature representations are used by the inner product to construct the predicted adjacency matrix of circRNAs and drugs. The ablation experiments conducted in this study demonstrate the individual impact of each key module on the resulting prediction performance, and finally, the validity of our model and its superiority over the current computational methods are demonstrated by a 5-fold cross-validation and the comparative analysis with other existing models.

## MATERIALS AND METHODS

### Datasets

Our dataset is downloaded from Ref. [13]. The utilized drug sensitivity data are sourced from the database GDSC [24]. The circRNA–drug sensitivity associations, on the other hand, are obtained from

the circRic [25] database. This comprehensive dataset encompasses a total of 404 circRNAs and 80 076 associations involving 250 drugs. The Wilcoxon test is used for each individual circRNA to find a drug sensitivity that is strongly correlated with its circRNA expression. While a connection is considered significant if its false discovery rate is less than 0.05. As a result, 4134 connections involving 218 drugs and 271 circRNAs qualify as these statistically significant relationships that are retrieved as the benchmark dataset. We create an association matrix  $A \in \mathbb{R}^{271 \times 218}$  based on these linkages. For each element in matrix  $A$ ,  $A_{ij} = 1$  indicates a direct association between circRNA  $i$  and the drug  $j$ , and  $A_{ij} = 0$  indicates that the association is unknown but that a potential association may exist. In addition to the connections between circRNAs and drug sensitivities, we curate the sequences of circRNA host genes from the database NCBI Gene [26], and the drug structure information is obtained from the PubChem [27].

### Similarity in sequence of circRNAs host genes

Following the methodology described in Ref. [9], we denote similarity in sequence of circRNAs host genes as similarity between circRNAs. Through the ratio function of the Levenshtein package in Python, the similarity is determined using the Levenshtein distance metric applied to their respective nucleotide sequences. The circRNA sequence similarity is represented as a CSS  $\in \mathbb{R}^{M \times M}$  matrix, where  $M$  represents the total number of circRNAs present in the dataset.

### Structural similarity of drugs

Because the structure of a drug has a great influence on its function, we obtain the similarity of a drug according to its structure. Following the acquisition of structure data from the PubChem [27], we first employ RDKit [28] to determine the topological fingerprint of each medication before using the Tanimoto approach to compute the structural similarity between pharmaceuticals. Finally, a structural similarity matrix for drugs is constructed and represented by the matrix  $DSS \in \mathbb{R}^{N \times N}$ , where  $N$  is the total number of drugs.

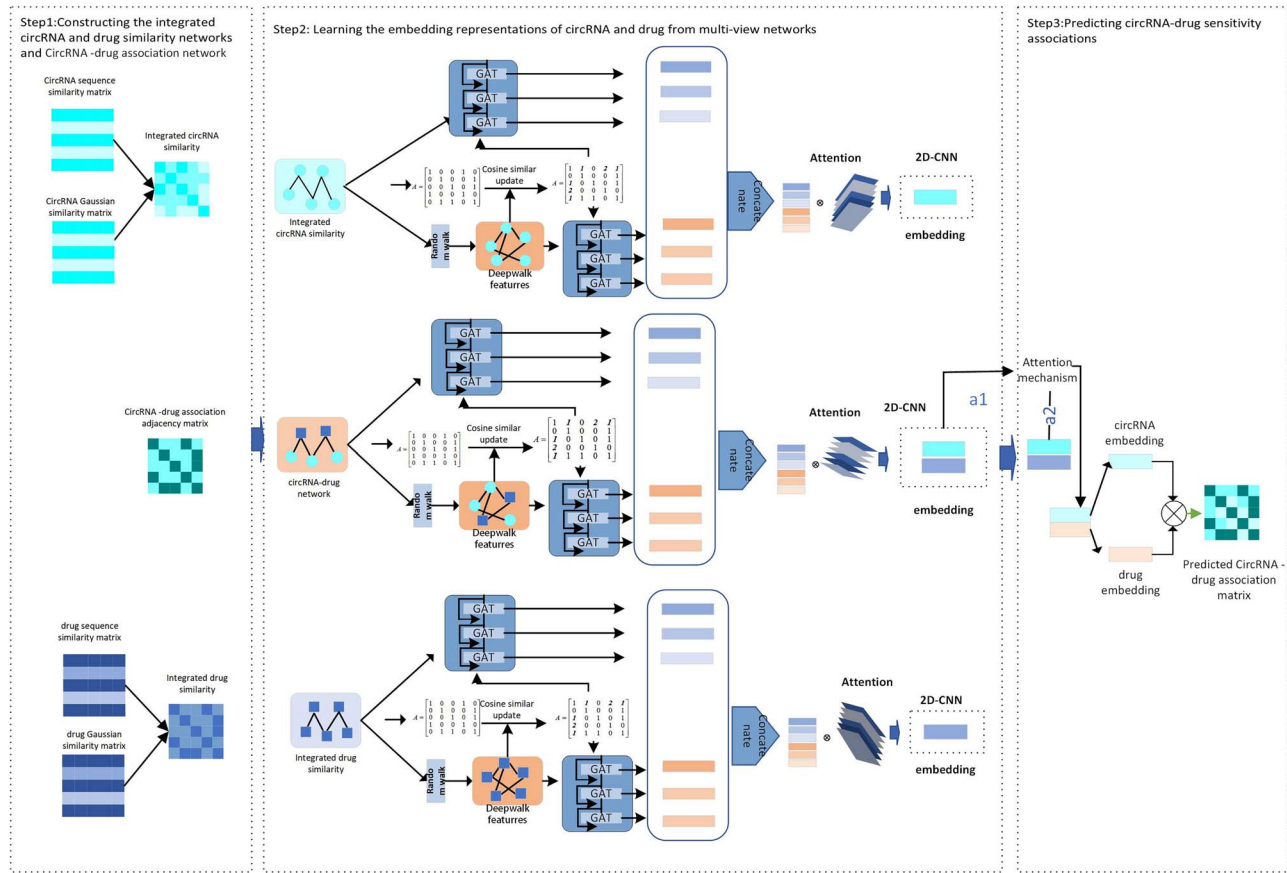
### Integrated circRNA similarity matrix

Although we obtained the sequence similarity of circRNAs, considering that not all pairs of circRNAs have sequence similarity, the constructed circRNA sequence similarity exhibits sparsity and may lack adequate informational content in certain instances. Therefore, we add Gaussian kernel similarity and use GIP kernel (GIPK) similarity to further enhance the sequence similarity information of circRNAs, and we fuse the sequence similarity of circRNAs with Gaussian kernel similarity to obtain the final circRNA similarity matrix.

GIPK similarity is extensively employed in biological entity similarities computations. GIPK similarity is calculated according to the reference [29], where we calculate the GIPK similarity by relying on the association matrix  $A$ , assuming that higher similarity between circRNAs represents a higher likelihood that they are associated with the same drugs. Therefore, the GIPK similarity matrix  $CGS$  for the circRNAs can be obtained.

After obtaining the GIPK similarity matrix  $CGS$ , we fuse the sequence similarity matrix  $CSS$  of the circRNAs with  $CGS$  to obtain the integrated circRNA similarity matrix  $CS \in \mathbb{R}^{M \times M}$ , where  $M$  indicates the number of circRNAs:

$$CS_{ij} = \begin{cases} \frac{(CSS_{ij} + CGS_{ij})}{2}, & \text{if } CSS_{ij} \neq 0 \\ CGS_{ij}, & \text{otherwise} \end{cases} \quad (1)$$



**Figure 1:** The workflow chart of our DGATCCDA method.

### Integrated drug similarity matrix

This part is similar to the circRNA similarity calculation method. We calculate the GIPK similarity of drug DGS and fuse the obtained drug structure similarity matrix DSS and GIPK similarity of drug DGS to get the integrated drug similarity matrix  $DS \in R^{N \times N}$ , where  $N$  indicates the number of drugs:

$$DS_{ij} = \begin{cases} \frac{(DSS_{ij} + DGS_{ij})}{2}, & \text{if } DSS_{ij} \neq 0 \\ DGS_{ij}, & \text{otherwise} \end{cases} \quad (2)$$

### DGATCCDA framework

Our DGATCCDA model applies a new feature extraction algorithm called the DeepWalk-aware graph attention network. GAT has been widely used in bioinformatics and has achieved significant results, e.g. MKGAT [30] and MNGACDA [14], but most of the existing computational methods in bioinformatics only take into account the local neighborhood information on the graph, while ignoring the global structural information. Inspired by this, we add DeepWalk [23] method to learn the topological information of the graph. DeepWalk is a random walk-based embedding algorithm, the idea of which is to represent nodes as fixed-length vectors, so that the similarity between nodes can be measured by the distance between vectors. We combine GAT and DeepWalk to learn more efficient node feature representations through DeepWalk-aware graph attention networks. The whole process of DGATCCDA is depicted in Figure 1 and consists of the following three main steps:

(i) Building multimodal networks: DGATCCDA learns circRNA and drug features from multisource information networks, which

include the circRNA–drug association network, integrated drug similarity network and integrated circRNA similarity network respectively.

(ii) Extracting features: Feature extraction is performed on the multimodal networks constructed in the first step. We first use the DeepWalk-aware GAT to learn the local neighborhood information of each node and the global structure information of the network, and then use the attention-based CNN for further feature fusion to obtain the final node representation.

(iii) Predicting association scores: Finally, we utilize an inner product decoder to decode the extracted circRNA and drug embedding features, thereby obtaining the predicted association matrix.

### DeepWalk-aware graph attention networks

Extracting features with GAT alone can only take into account the local information of the graph, not the global structural information, so we use DeepWalk-aware GAT to solve this problem. We combine DeepWalk and GAT to capture the local features and global topological features of the graph. The graph's adjacency matrix is updated by computing the cosine similarity between nodes using the acquired DeepWalk features. Finally, the original similarity features and DeepWalk features are conducted with the multilayer GAT for information transfer and fusion, respectively.

### GAT module

The circRNA–drug association network, integrated drug similarity network and integrated circRNA similarity network are denoted as  $A_N$ ,  $D_N$  and  $C_N$ , respectively.  $AS$ ,  $DS$  and  $CS$  are the integrated feature matrices corresponding to  $A_N$ ,  $D_N$  and  $C_N$ , respectively.

Since the processes of learning node embedding representations in the above three networks as similar, we use  $C_N$  as an example to provide a detailed explanation of the process performed by the DeepWalk-aware graph attention networks.

The feature matrix CS of  $C_N$  is denoted by  $H^{(0)}$ , and  $H^{(0)} \in \mathbb{R}^{M \times M}$  is used as the initial input of the GAT, where  $M$  denotes the number of circRNAs.  $G \in \mathbb{R}^{M \times M}$  is the adjacency matrix of  $C_N$ , and in the circRNA (or drug) similarity networks, we only link up each circRNA's (or drug's) 25 closest neighbors. For the input  $H^{l-1} = \{h_1^{l-1}, h_2^{l-1}, h_3^{l-1}, \dots, h_M^{l-1}\}$ ,  $h_i^{l-1} \in \mathbb{R}^{F^{l-1}}$  at layer  $l$ ,  $F^{l-1}$  is the embedding dimensionality of each node at layer  $l$ . The similarity between node  $h_i^{l-1}$  and its neighbor node  $h_j^{l-1}$  is first calculated to obtain the attention weight coefficient  $e_{ij}^{l-1}$ :

$$e_{ij}^{l-1}(h_i^{l-1}, h_j^{l-1}) = \text{Leaky Re Lu}(a^T [Wh_i^{l-1} \| Wh_j^{l-1}]) \quad (3)$$

where  $a \in \mathbb{R}^{2F}$  is the trained vector of attention parameters. *LeakyReLU* is the nonlinear activation function (with a negative slope of 0.2), and  $W$  is the weight matrix that serves to project the node features into the dimensional space of the next layer.

In the next step, we adopt the softmax function to regularize the attention coefficients and obtain the final attention score:

$$\theta_{ij}^{l-1} = \text{soft max}(e_{ij}^{l-1}) = \frac{\exp(e_{ij}^{l-1})}{\sum_{t \in N_i} \exp(e_{it}^{l-1})} \quad (4)$$

where  $N_i$  represents the set of neighboring nodes of node  $i$  and  $\theta_{ij}^{l-1}$  is the attention score between node  $i$  and node  $j$  in layer  $l$ . Next, the computed attention score is used for information transfer to extract information about node  $i$  and its neighbor nodes to obtain the updated embedding representation of node  $i$ :

$$h_i^l = \sigma \left( \sum_{j \in N_i} \theta_{ij}^{l-1} h_j^{l-1} \right) \quad (5)$$

where  $\sigma$  represents a nonlinear activation function.

Finally, the node feature matrix  $H^l = (h_1^l, h_2^l, \dots, h_M^l)$ ,  $H^l \in \mathbb{R}^{M \times F^l}$  is obtained after performing information transfer through the  $l$ th GAT layer,  $F^l$  denotes the embedding dimensionality of the  $l$ th layer, and  $M$  denotes the number of nodes.

## Residual module

The deeper layers of the stacked GAT tend to lead to an over-smoothing problem, i.e. the learned node representations become highly indistinguishable, so a residual algorithm is added to solve the above problem, and the following is the process of updating the feature matrix of the GAT with the addition of the residual module:

$$H^{l+1} = H_{\text{resd}}^l + H_{\text{aggr}}^l = \text{Linear}(H^l) + \text{Aggregate}(H^l, N_h) \quad (6)$$

where  $H_{\text{resd}}^l$  represents a linear projection layer,  $H_{\text{aggr}}^l$  represents the information aggregation process in GAT and  $N_h$  represents the set of neighboring nodes for each node.

## DeepWalk

The original GAT algorithm only aggregates and conveys information about local neighboring nodes without considering the topological information of the input graph. To compensate for the lack of features extracted by the GAT, we are inspired by the

reference [31] and introduce DeepWalk to extract the topological features of the graph. We utilize a deep walking method to extract the topological information of the graph and learn to encode a potential representation of the relationship between nodes. DeepWalk is used as an unsupervised feature extraction method that has been applied in the field of bioinformatics with many applications and has been proven to be effective [32–34]. The DeepWalk method consists of random walks and SkipGram [35] to learn the embedding representations of nodes. First, for each node, a random walk of a certain length is performed to obtain a sequence as a representative of that node. Specifically, each time a neighbor node is randomly selected as the next node and recorded, and the walk length is set to  $t$  in the experiment. Then, the SkipGram is used to obtain the final DeepWalk embedding representation. The SkipGram maximizes the cooccurrence probability of the surrounding nodes to obtain the node vector representation, and its objective function is

$$\min_{\Phi} - \log \Pr(\{h_{i-w}, \dots, h_{i+w}\} \setminus h_i | \Phi(h_i)) \quad (7)$$

where  $h_i$  denotes the  $i$ th node,  $\Phi(h_i) \in \mathbb{R}^d$  denotes the vector representation after projection on  $h_i$  and  $d$  is the embedding dimension of DeepWalk. Further optimization is performed using the SkipGram module:

$$\Pr(\{h_{i-w}, \dots, h_{i+w}\} \setminus h_i | \Phi(h_i)) = \prod_{j=i-w, j \neq i}^{i+w} \Pr(h_j | \Phi(h_i)) \quad (8)$$

DeepWalk is described in further depth in Ref. [23]. Finally, we get the DeepWalk feature representation  $\bar{H} = [\bar{h}_1, \dots, \bar{h}_i, \dots, \bar{h}_M] \in \mathbb{R}^{M \times d}$  for each node and  $\bar{h}_i \in \mathbb{R}^d$  is DeepWalk feature vector of the  $i$ th node.

## Updating the adjacency matrix

Taking into consideration that the defined adjacency matrix  $G$  does not incorporate the global structural information of the graph, we adjust the matrix  $G$  with the obtained DeepWalk embedding features. We can determine if a stronger connection between the nodes exists by utilizing the DeepWalk technique to obtain an embedding that can indicate how similar their node structures are. By employing cosine similarity, the degrees of similarity between nodes can be ascertained. The following is the procedure for calculating the cosine similarity and updating the adjacency matrix:

$$\cos(\theta[i, j]) = \frac{\bar{h}_i \cdot \bar{h}_j}{\|\bar{h}_i\| \|\bar{h}_j\|}, \quad (9)$$

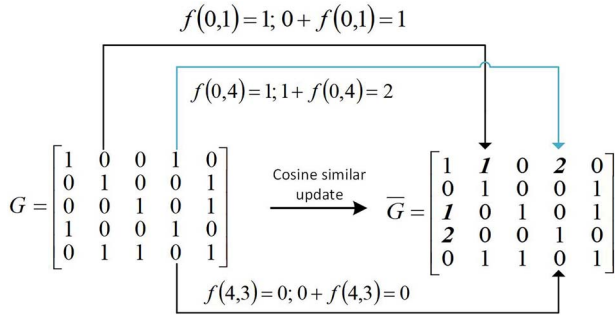
$$f(i, j) = \begin{cases} 1, & \text{if } \cos(\theta[i, j]) \geq \eta, \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

$$\bar{G}_{ij} = G_{ij} + f(i, j), \quad i, j = 1, 2, \dots, m, \quad (11)$$

where  $\bar{h}_i \in \mathbb{R}^d$  and  $\bar{h}_j \in \mathbb{R}^d$  are the DeepWalk features of node  $i$  and node  $j$ , respectively.  $\eta$  is the threshold value for updating the adjacency matrix.  $G_{ij}$  denotes the association between node  $i$  and node  $j$ , which is equal to 1 for an association and 0 for no association. An example of the process for updating the adjacency matrix can be seen in Figure 2.

In general, we use a threshold to update the adjacency matrix  $G$  instead of directly updating the matrix with the original similarity. The greater the cosine similarity is between node  $i$  and node  $j$ , the





**Figure 2:** Example of the adjacency matrix update process.

greater the possibility of an association between the two nodes, and we increase the cosine similarity to 1 by setting a threshold, thus allowing the GAT to extract graph and node information more effectively. The updated adjacency matrix  $\bar{G}$  is utilized for the edges of each subsequent layer of the GAT input, and we will also use the feature matrix  $\bar{H}$  obtained by DeepWalk as the GAT input for feature information transfer, this is similar to the calculation process of  $H^l$  obtained for the original input features after utilizing the GAT.

### Attention CNN

After employing DeepWalk and GAT, we obtain the feature information extracted from different layers and stack them to obtain  $H_{C_N} = (H_{C_N}^1, \dots, H_{C_N}^i, \dots, H_{C_N}^M, \bar{H}_{C_N}^1, \dots, \bar{H}_{C_N}^L) \in R^{V \times M \times F}$ .  $H_{C_N}^i$  denotes the initial similarity matrix obtained after the  $i$ th GAT layer,  $\bar{H}_{C_N}^L$  denotes the DeepWalk feature matrix obtained after the  $i$ th GAT layer,  $V$  denotes the numbers of views in different GAT layers,  $M$  denotes the amount of nodes, and  $F$  is the dimensionality of the node features. We assign different importance scores to the feature matrices acquired from different GAT output layers for the similarity features and DeepWalk features via an attention mechanism. First, we calculate the importance scores of different GAT layer views  $\theta_{C_N}$ :

$$\theta_{C_N} = \text{FNN}_{C_N}(\text{GAP}_{C_N}(H_{C_N})) \quad (12)$$

where  $\text{FNN}_{C_N}$  denotes a fully connected layer with two sublayers and  $\text{GAP}_{C_N}$  is a global mean pooling layer. Eventually, the different views are fused by calculating the importance scores  $\theta_{C_N}$  of the different GAT views obtained:

$$H_{C_N}^{\text{Att}} = \text{CNN}_{C_N}(\sigma(\theta_{C_N} \cdot H_{C_N})) \quad (13)$$

where  $\text{CNN}_{C_N}$  is the 2D convolutional neural network (CNN) used to fuse different GAT layer views and  $\sigma$  denotes the nonlinear activation function.  $H_{C_N}^{\text{Att}} \in R^{M \times F}$  denotes the final feature matrix after fusing different feature views.

### Prediction

Finally, we get the embedding representation  $H_{C_N}^{\text{Att}}$  of  $C_N$  from the DeepWalk-aware GAT and attention CNN. Similarly, we can also obtain the embedding representation  $H_{D_N}^{\text{Att}} \in R^{N \times F}$  of  $D_N$  and the  $H_{A_N}^{\text{Att}} \in R^{(M+N) \times F}$  of  $A_N$ .  $H_{C_N}^{\text{Att}}$  and  $H_{D_N}^{\text{Att}}$  are stitched together to obtain

$H_{B_N}^{\text{Att}} = \begin{bmatrix} H_{C_N}^{\text{Att}} \\ H_{D_N}^{\text{Att}} \end{bmatrix}$ . Then, the attention mechanism is used to fuse  $H_{A_N}^{\text{Att}}$  and  $H_{B_N}^{\text{Att}}$ .

$$H^{\text{Att}} = a \cdot H_{A_N}^{\text{Att}} + b \cdot H_{B_N}^{\text{Att}} \quad (14)$$

where  $a$  and  $b$  are the corresponding attention parameters, which are automatically learned by the neural network,  $H^{\text{Att}} \in R^{(M+N) \times F}$  denotes the fused embedding representation, where  $H^{\text{Att}} = \begin{bmatrix} H_C \\ H_D \end{bmatrix}$ ,  $H_C$  denotes the final circRNA feature matrix and  $H_D$  denotes the final drug feature matrix. After that, to construct the adjacency matrix, we employ the inner product operation. Predicting the associations between circRNAs and drugs is a binary classification task, so the activation function is the sigmoid function:

$$U' = \text{sigmoid}(H_C H_D^T) \quad (15)$$

where  $U'$  denotes the adjacency matrix predicted by the DGATCCDA model,  $U'_{ij}$  is the prediction score between the  $i$ th circRNA and the  $j$ th drug, and a higher value indicates a higher probability of an edge between the circRNA and the drug. We used the binary cross-entropy loss to train the DGATCCDA model and optimize the model parameters. The positive samples in our study consist of circRNA and drug pairs with known associations, while the negative samples comprise with unknown associations, and the sets of negative and positive samples used for training are denoted by  $y^+$  and  $y^-$ , respectively. Because in the training data, the number of negative samples surpasses the number of positive samples by a considerable margin, the network may be overly sensitive to the negative samples and fail to correctly identify the positive samples that do not appear in the training data. This leads to poor model generalization in practice. Therefore, in this experiment, to ensure equal representations of the positive and negative samples during the training phase, we randomly choose an equal number of negative samples to the number of positive samples. Ultimately, the following is used as the formula for the loss function:

$$L = - \sum_{(i,j) \in y^+ \cup y^-} [U_{ij} \ln U'_{ij} + (1 - U_{ij}) \ln (1 - U'_{ij})] \quad (16)$$

where  $(i,j)$  denotes the pair containing circRNA  $i$  and drug  $j$ , and  $U'_{ij}$  and  $U_{ij}$  denote the predicted association scores and true association scores between circRNA  $i$  and drug  $j$ , respectively.

## RESULT

### Evaluation criteria

We employ 5-fold cross validation in this experiment to unbiasedly test the model's effectiveness in predicting the correlations between circRNAs and drugs. We will randomly divide the samples into five roughly equal subsets, commonly referred to as folds, and the training set is constructed by sequentially selecting four of the folds, while the remaining fold is designated as the test set.

In addition, we mainly select the following values to assess the performance of our model: the area under the ROC curve (AUC), the area under the P-R curve (AUPR), accuracy, precision, recall, the F1-measure and specificity. The formulas for these metrics are described in Eqs. (17–22). The AUC measures the performance of the model by calculating the relationship between the false-positive rate and the true-positive rate at different probability or label thresholds. The AUPR plots the P-R curve by calculating the precision and recall values of the classifier under different precision and recall thresholds and then calculates the area under the curve to obtain the AUPR value:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (17)$$

$$Acc = \frac{TP + TN}{TN + TP + FP + FN} \quad (18)$$

$$Pre = \frac{TP}{FP + TP} \quad (19)$$

$$Rec = \frac{TP}{FN + TP} \quad (20)$$

$$F1 = \frac{2TP}{2TP + FN + FP} \quad (21)$$

$$Spec = \frac{TN}{FP + TN} \quad (22)$$

where TN and TP are the numbers of correctly predicted unassociated circRNA–drug pairs and associated pairs, respectively. FP refers to incorrectly determining negative class samples as positive class samples, i.e. incorrectly predicting unassociated circRNA–drug pairs, and FN refers to incorrectly determining positive class samples as negative classes samples, i.e. incorrectly predicting associated circRNA–drug pairs.

### Comparison with other models

In this section, since there are very few computational methods are available for exploring latent circRNA–drug sensitivity associations, currently only GATECDA [13] and MNGACDA [14], we compare DGATCCDA with computational methods in other association prediction fields to better assess the performance of our model. Finally, we conduct a comparison between our model and five state-of-the-art methods, including MNGACDA [14], GraphCDA [36], GATECDA [13], MINIMDA [37] and MKGAT [30]. Among them, GraphCDA is a model proposed to excavate the latent circRNA–disease associations, and MINIMDA and MKGAT are the models designed to identify the associations between miRNA and diseases. All methods are implemented under the same experimental conditions and all of them use the optimal parameters derived from their respective papers.

MNGACDA [14]: a method is employed to explore latent circRNA and drug sensitivity associations, it uses graph self-encoders and attention mechanisms to learn multimodal networks.

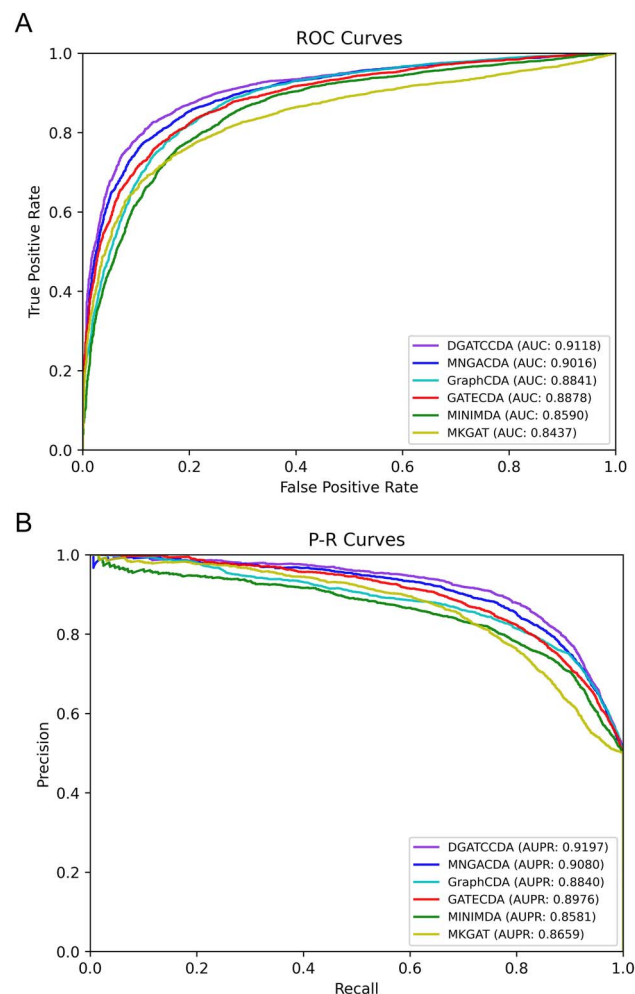
GraphCDA [36]: utilizing GAT and GCN for learning graph representation, and finally predicting circRNA and disease associations using random forests.

GATECDA [13]: a computational method that uses graph self-encoders to extract features from circRNA and drug networks separately and eventually utilizes a fully connected layer for circRNA–drug sensitivity association identification.

MINIMDA [37]: a computational method for miRNA and disease association identification by learning multimodal networks, using a high-order GCN to extract features, and finally entering a multilayer perceptron.

MKGAT [30]: A computational model for predicting miRNA–disease associations by employing a multilayer GAT for feature extraction and a dual Laplacian regularized least squares for decoding.

A comparison among the AUC values obtained by several models under 5-fold cross-validation can be seen in Figure 3(A), and DGATCCDA has the highest AUC value of 0.9118 among the six models. The comparison among the AUPR values can be seen in Figure 3(B), and DGATCCDA is also the highest among them. In addition, other performance evaluation metrics including the accuracy, precision, recall, F1 score and specificity, are compared in Table 1, and the results of DGATCCDA are 0.8459, 0.8444, 0.8492, 0.8461 and 0.8429, respectively. It can be observed that DGATCCDA is almost superior to the other five methods in terms of most evaluation metrics, except for its low Recall value. In order to



**Figure 3:** (A) Comparison among the ROC and PR curves produced by DGATCCDA and the other existing methods in 5-CV experiments. (B) Comparison among the ROC and PR curves produced by DGATCCDA and the other existing methods in 5-CV experiments.

further characterize the differences between our model and other models, we use t-test to determine whether DGATCCDA is significantly different from other models. As you can see in Table 2, there are very significant differences between DGATCCDA and other four models under the significance level of 0.05. The difference between DGATCCDA and MNGACDA is not significant, but DGATCCDA is superior to MNGACDA in terms of most evaluation metrics. Therefore, the results suggest that DGATCCDA is a very effective computational method for excavating circRNA and drug sensitivity association.

### Parameter setting

In this experiment, we adopt 5-fold cross validation to determine the effect of different hyperparameters on the DGATCCDA model and compare the resulting evaluation metric values. The main parameters in the experiment are the amount of GAT layers  $L$ , the dimensionality of the GAT layer embedding  $F$  and the embedding dimensionality of DeepWalk  $d$ . In addition, the random walk length of DeepWalk is set to 30, the number of walks is set to 10, the window size is set to 5 for the SkipGram model and the threshold  $\eta$  of the adjacency matrix is updated to 0.92.

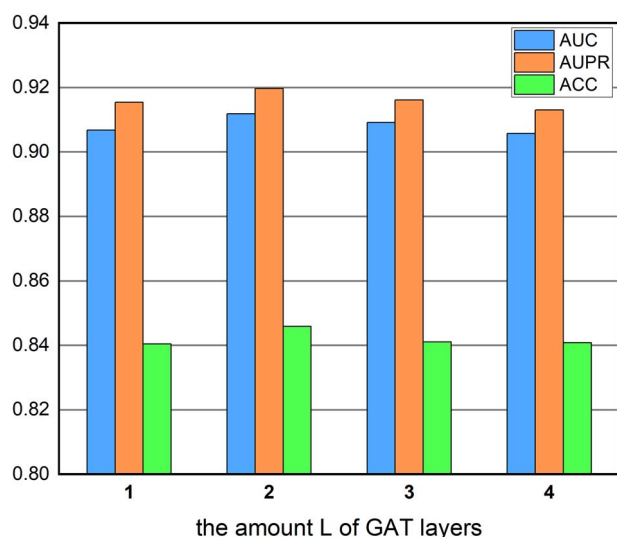
**Table 1.** Comparison with other methods

Method	Accuracy	Precision	Recall	F1-score	Specificity
DGATCCDA	<b>0.8459</b>	<b>0.8444</b>	0.8492	<b>0.8461</b>	<b>0.8429</b>
MNGACDA	0.8291	0.8100	0.8597	0.8341	0.7981
GraphCDA	0.8095	0.7765	<b>0.8748</b>	0.8210	0.7439
GATECDA	0.8105	0.7853	0.8556	0.8184	0.7644
MINIMDA	0.7841	0.7464	0.8608	0.7994	0.7077
MKGAT	0.7821	0.7808	0.7857	0.7827	0.7784

Bold values indicate the best results.

**Table 2.** The differences between DGATCCDA and other methods in terms of 5-fold CV

DGATCCDA versus	MNGACDA	GraphCDA	GATECDA	MINIMDA	MKGAT
P-value	0.079	1.1e-03	1.4e-03	5.38e-05	3.67e-06

**Figure 4:** Parameter analysis regarding the number of GAT layers  $L$ .

### The number of GAT layers $L$

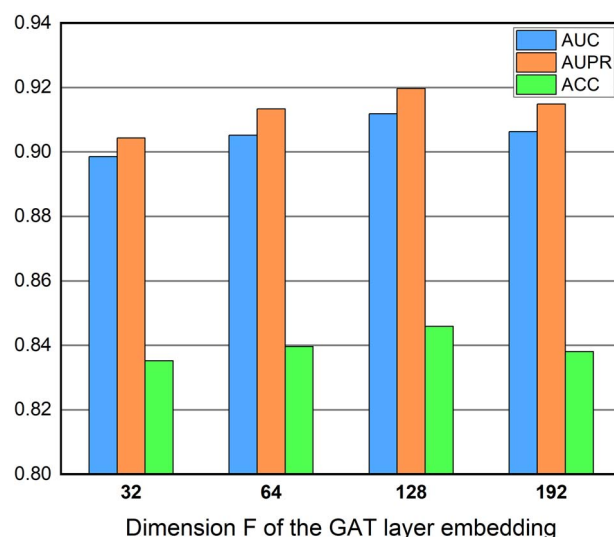
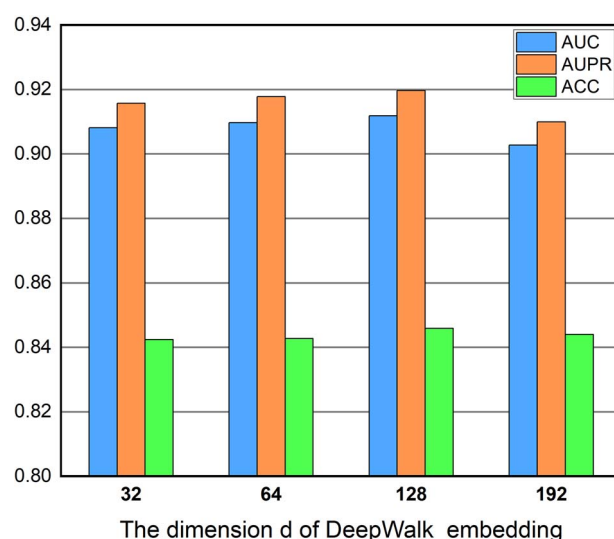
Generally, additional GAT layers may gather more node data. However, an excessive increase in the amount of layers can lead to oversmoothing and overfitting problems. We analyze the  $L$ , and the experimental results can be seen in Figure 4, where the best prediction performance is obtained when we utilize two layers.

### The dimension $F$ of the GAT layer embedding

The dimension of the learnable parameter matrix for each feature depends on its embedding size. The impact of  $F$  can be observed in Figure 5, where clear changes in the AUC and AUPR values are induced at various feature dimensions. Finally, we adjusted the hidden layer dimension to 128 to guarantee prediction performance of the model.

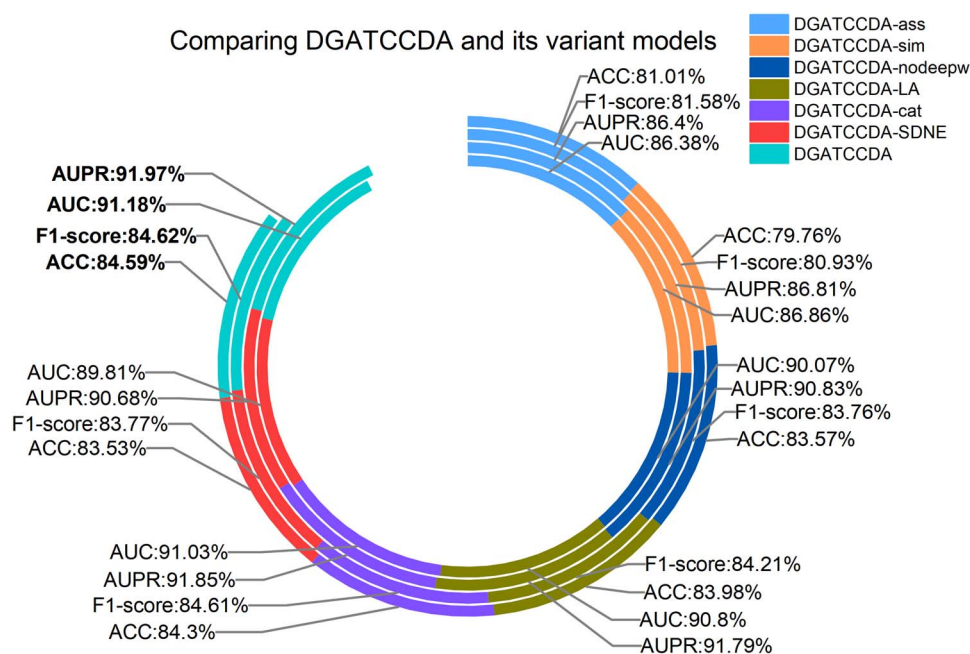
### The dimension $d$ of DeepWalk embedding

DeepWalk generates low-dimensional node representations by learning the neighbor relationships of the nodes in the input graph. If the embedding dimension is set too low, the vector of node representations may not capture the complex features and relationships of the nodes, and if the embedding dimension is set too high, the vector of node representations may overfit the training data. The effect of dimensional changes on the model can be seen in Figure 6, and, finally, we select 128 as the value of  $d$ .

**Figure 5:** Parameter analysis regarding the dimension  $F$  of the GAT layer embedding.**Figure 6:** Parameter analysis regarding the dimension  $d$  of DeepWalk embedding.

### Ablation experiments

Because the embedding representation of our model learning nodes employ a multimodal approach to extract features, we first



**Figure 7:** Ablation experiments results of the seven different models.

evaluate the impacts of the circRNA–drug association network and integrated similarity network on our model. We propose two variant models DGATCCDA-ass and DGATCCDA-sim. DGATCCDA-ass solely extracts features from the association network, but not from the similarity network of circRNA–drug fusion, while DGATCCDA-sim extracts features only from the similarity network of circRNA–drug fusion.

In addition, the DeepWalk part of our model is used to extract the topological features of the graph. DeepWalk plays an important role in the model feature extraction process, and to evaluate its impact, we develop a variant model called DGATCCDA-nodewp, indicating that the DGATCCDA model does not use the DeepWalk module to extract the topological features of the graph. The 2D-CNN of the attention mechanism is used in our model to fuse the views of different GAT layers, which include the similarity feature view and the DeepWalk feature view. We use a general layer attention module instead of the 2D-CNN, which is similar to the attention mechanism in MKGAT [30], and represent this variant as DGATCCDA-LA. DGATCCDA fuses the features that are finally extracted from the circRNA–drug association network and the fused similarity network using layer attention, so we propose DGATCCDA-cat, which replaces the layer attention mechanism with a splicing operation to determine the effect of layer attention. To further reflect the advantage of DeepWalk, we replace DeepWalk with another graph embedding algorithm SDNE [38], and this variant model is called DGATCCDA-SDNE.

The final 5-fold cross-validation results of the original models, DGATCCDA-ass, DGATCCDA-sim, DGATCCDA-nodewp, DGATCCDA-LA, DGATCCDA-cat and DGATCCDA-SDNE are compared with each other and represented in Figure 7. Based on the results, we are able to find that both the circRNA–drug association network and the fused similarity network have significant impacts on the prediction of the models, indicating that the multimodal learning approach does have an important impact on the resulting prediction performance. In addition from the comparison between DGATCCDA-nodewp and the original model, it can be found that DeepWalk extracted graph topological features play a role in the final prediction effect of the model. The

performance of DGATCCDA-LA and DGATCCDA-cat is slightly lower than that of the original model in terms of the AUC, AUPR and accuracy, indicating that the attentional 2D-CNN is effective in terms of fusing different layer views and that the multimodal views from the final layer with attentional fusion are effective. From the results of DGATCCDA-SDNE, it can be seen that the AUC, AUPR, F1-score and accuracy values of DGATCCDA-SDNE are lower than those of the DGATCCDA, which indicates the suitability of DeepWalk for us in extracting information about the global structure of the graph.

## Case studies

To make the experimental results of our model more credible, we take all known associations in the GDSC dataset for training and finally obtain a prediction matrix. After that, we find the association evidence corresponding to the model prediction matrix from another dataset, CTRP [39], to demonstrate the accuracy of the prediction results. Specifically, we conduct a case study involving the two drugs Piperlongumine and Linifanib, taking the top twenty circRNAs with the highest prediction scores for Piperlongumine and Linifanib in the predicted circRNA and drug association matrix, and finally validating them on CTRP.

Piperlongumine is an alkaloid [40]. Recent research has demonstrated that Piperlongumine can specifically kill a variety of tumor cells, including colon, ovarian and liver cancer cells, without harming normal cells or having obvious toxic side effects. It also has a number of pharmacological activities, including anti-pathogenic microorganism, sedation and anticonvulsant properties [41]. From the data in Table 3, 17 of the top 20 predicted circRNA and Piperlongumine associations have been confirmed in CTRP, and all of these confirmed associations satisfied the Wilcoxon tests with a false discovery rate (FDR) < 0.05. FDR > 0.05 represents a nonsignificant association. Along the same lines, Linifanib (VEGFR inhibitor) is a multitargeted VEGF and PDGFR receptor family inhibitor that has shown potent antiangiogenic and antitumor effects in preclinical studies [42–44]. From Table 4, the predictions of 16 of the top 20 circRNAs related to Linifanib have been validated in CTRP.



**Table 3.** Prediction of top 20 circRNAs related to Piperlongumine

Rank	circRNAs	Evidences	Rank	circRNAs	Evidences
1	COL3A1	Nonsignificant	11	MUC1	CTRP
2	EFEMP1	CTRP	12	FBN1	CTRP
3	COL6A1	CTRP	13	CSRP1	Nonsignificant
4	PEA15	CTRP	14	SERPINH1	CTRP
5	FBLN1	CTRP	15	ASPH	CTRP
6	POLR2A	CTRP	16	CTTN	CTRP
7	LTBP3	CTRP	17	ECI2	Nonsignificant
8	MUC16	CTRP	18	PSAP	CTRP
9	PTMS	CTRP	19	KRT7	CTRP
10	AHNAK	CTRP	20	ANP32B	CTRP

**Table 4.** Prediction of top 20 circRNAs related to Linifanib

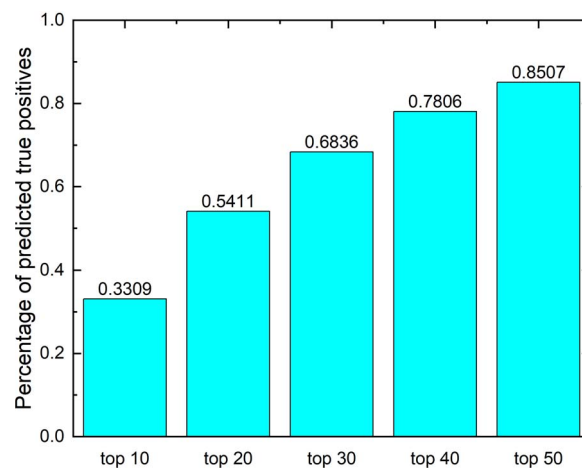
Rank	circRNAs	Evidences	Rank	circRNAs	Evidences
1	CRIM1	CTRP	11	AHNAK	CTRP
2	ANXA2	CTRP	12	HSP90B1	CTRP
3	DCBLD2	CTRP	13	COL7A1	CTRP
4	ANP32B	CTRP	14	KRT19	CTRP
5	FBLN1	Nonsignificant	15	COL8A1	CTRP
6	CTTN	CTRP	16	FND3B	Nonsignificant
7	CALD1	CTRP	17	PTMS	CTRP
8	ANKRD36	CTRP	18	LTBP3	CTRP
9	VIM	CTRP	19	SPINT2	Nonsignificant
10	LINC01089	Nonsignificant	20	CPSF6	CTRP

We choose these two drugs, which have just one connection with circRNAs, from the dataset to conduct *de novo* testing to further assess the prediction power of the proposed model for circRNAs and new drug sensitivity. To validate these circRNAs in the CTRP database, we first remove the special relationship between Crizotinib and MG-132, train the model with other known associations, and then choose the top 10 circRNAs with Crizotinib and MG-132 prediction scores from the final generated prediction results. Crizotinib is an orally administered receptor tyrosine kinase inhibitor that is primarily used for the treatment of patients diagnosed with locally advanced or metastatic non-small-cell lung cancer who have a positive expression of mesenchymal lymphoma kinase (ALK) expression. Additionally, it is indicated for patients with advanced non-small-cell lung cancer who test positive for ROS1 expression. It has been demonstrated in clinical trials to show high activity in lung cancer patients with ALK mutations [45]. MG-132 is a potent, selective, cell-permeable inhibitor of the peptide aldehyde proteasome [46]. As indicated in Table 5, all of the top 10 circRNAs related to Crizotinib have been validated and confirmed through CTRP. Additionally, out of the top 10 circRNAs related to MG-132, four of them have been verified through CTRP.

We further count up the number of known circRNA–drug sensitivity associations that are correctly identified under different top-ranked thresholds to estimate the performance of DGATCCDA. As shown in Figure 8, among the 4134 true positives, DGATCCDA can identify 3227 (or 78.06%) known associations in the top 40 predictions. The above results further demonstrate that DGATCCDA is effective in identifying circRNA–drug interactions.

## CONCLUSION

In recent years, an increasing number of studies have demonstrated the substantial role of circRNAs in drug sensitivity.

**Figure 8:** The percentage of predicted true positives by DGATCCDA under different rankings.

Identifying the associations between drug sensitivity and circRNAs has great potential for facilitating the discovery of novel drugs and advancing disease treatment strategies. In this work, we introduce a computational approach based on deep learning for circRNA–drug sensitivity association identification. We adopt a multimodal strategy to construct the drug and circRNA association network, the integrated drug network and integrated circRNA network respectively, and use DeepWalk-aware GAT for feature extraction; this is followed by an attention CNN to incorporate the views of these three networks, and finally the inner product decoder is used for prediction. Case studies and experimental results show that DGATCCDA efficiently utilizes the available feature information to accurately explore latent circRNAs and drug sensitivity associations and outperforms

**Table 5.** Prediction of the top 10 circRNAs related to the new drugs Crizotinib and MG-132

Crizotinib			MG-132		
Rank	circRNAs	Evidences	Rank	circRNAs	Evidences
1	POLR2A	CTRP	1	POLR2A	Nonsignificant
2	THBS1	CTRP	2	CTTN	CTRP
3	CRIM1	CTRP	3	THBS1	CTRP
4	ASPH	CTRP	4	CRIM1	Nonsignificant
5	VIM	CTRP	5	ASPH	CTRP
6	SPINT2	CTRP	6	SPINT2	Nonsignificant
7	CTTN	CTRP	7	ANP32B	Nonsignificant
8	EFEMP1	CTRP	8	ANXA2	CTRP
9	KRT19	CTRP	9	FBLN1	Nonsignificant
10	ANP32B	CTRP	10	PTMS	Nonsignificant

other existing computational methods. It is noted that the existing circRNA and drug sensitivity association and biological information data are not sufficient, which may lead to less accurate prediction results. Therefore, to further improve the predictive performance of the model, we plan to gather additional circRNA and drug sensitivity association data and incorporate a wider range of biomedical data to construct more comprehensive similarity information in our subsequent study. The method proposed here adopts the inner product as a decoder. It is flexible, and the inner product can be replaced by some machine learning-based methods [47], such as the Random Forest [48] and eXtreme gradient boosting [49]. In addition, we will also consider applying graph Transformer [50] to further capture the global structural information and improve the predictive performance of the model. The existing studies on computational methods for predicting the correlations between circRNA and drug sensitivity are very limited, and further studies in this field are worthwhile.

#### Key Points

- We present a novel method, called DGATCCDA, to predict unobserved circRNA–drug sensitivity associations, in which a multimodal network is constructed to extract multi-source information.
- DGATCCDA sufficiently extracts the global and local information of graph structures from the multimodal networks by the DeepWalk-aware graph attention networks and then uses the attention-based CNN to further fuse feature information.
- The ultimate experimental results obtained under 5-fold cross-validation show that DGATCCDA is better than the five current state-of-the-art calculation methods and the case study also shows that DGATCCDA is an effective computational method for exploring latent circRNA–drug sensitivity associations.

## FUNDING

This work is supported by the National Natural Science Foundation of China (grant numbers 62362034, 61862025, 61873089) and the Natural Science Foundation of Jiangxi Province of China (grant numbers 20232ACB202010, 20212BAB202009, 20181BAB211016).

## AUTHOR CONTRIBUTIONS

Guanghui Li conceived the study, analyzed the results and drafted the article. Youjun Li collected the data, designed and performed the experiments, and drafted the article. Cheng Liang revised the article. Jiawei Luo supervised the study and revised the article. All authors read and approved the final manuscript.

## DATA AVAILABILITY

Implementations of DGATCCDA can be obtained at <https://github.com/ghli16/DGATCCDA>.

## REFERENCES

1. Jeck WR, Sorrentino JA, Wang K, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 2013;**19**:141–57.
2. Chen L-L, Yang L. Regulation of circRNA biogenesis. *RNA Biol* 2015;**12**:381–8.
3. Li X, Yang L, Chen L-L. The biogenesis, functions, and challenges of circular RNAs. *Mol Cell* 2018;**71**:428–42.
4. Chen X, Fan S, Song E. Noncoding RNAs: new players in cancers. *Adv Exp Med Biol* 2016;**927**:1–47.
5. Xie J, Ning Y, Zhang L, et al. Overexpression of hsa\_circ\_0006470 inhibits the malignant behavior of gastric cancer cells via regulation of miR-1234/TP53I11 axis. *Eur J Histochem* 2022;**66**:3477.
6. Wang C, Jiang H, Peng J, et al. Circular RNA circ\_SKA3 enhances gastric cancer development by targeting miR-520h. *Histol Histopathol* 2023;**38**:317–28.
7. Kristensen LS, Andersen MS, Stagsted LVW, et al. The biogenesis, biology and characterization of circular RNAs. *Nat Rev Genet* 2019;**20**:675–91.
8. Kristensen LS, Hansen TB, Venø MT, et al. Circular RNAs in cancer: opportunities and challenges in the field. *Oncogene* 2017;**37**:555–65.
9. Xu X, Zhang J, Tian Y, et al. CircRNA inhibits DNA damage repair by interacting with host gene. *Mol Cancer* 2020;**19**:128.
10. Liang G, Ling Y, Mehrpour M, et al. Autophagy-associated circRNA circCDYL augments autophagy and promotes breast cancer progression. *Mol Cancer* 2020;**19**:65.
11. Wei L, Sun J, Zhang N, et al. Noncoding RNAs in gastric cancer: implications for drug resistance. *Mol Cancer* 2020;**19**:62.
12. Wang C-C, Han C, Zhao Q, et al. Circular RNAs and complex diseases: from experimental results to computational models. *Brief Bioinform* 2021;**22**:286.

13. Deng L, Liu Z, Qian Y, et al. Predicting circRNA-drug sensitivity associations via graph attention auto-encoder. *BMC Bioinform* 2022;**23**:1–15.
14. Yang B, Chen H. Predicting circRNA-drug sensitivity associations by learning multimodal networks using graph auto-encoders and attention mechanism. *Brief Bioinform* 2023;**24**:bbac596.
15. Veličković P, Cucurull G, Casanova A, et al. Graph attention networks. arXiv preprint arXiv:1710.10903. 2017.
16. Chen Y, Wang J, Wang C, et al. Deep learning models for disease-associated circRNA prediction: a review. *Brief Bioinform* 2022;**23**:bbac364.
17. Zhang Y, Lei X, Fang Z, et al. CircRNA-disease associations prediction based on metapath2vec++ and matrix factorization. *Big Data Min Anal* 2020;**3**:280–91.
18. Niu M, Zou Q, Wang C. GMNN2CD: identification of circRNA-disease associations based on variational inference and graph Markov neural networks. *Bioinformatics* 2022;**38**:2246–53.
19. Mudiyansele TB, Lei X, Senanayake N, et al. Predicting CircRNA disease associations using novel node classification and link prediction models on graph convolutional networks. *Methods* 2022;**198**:32–44.
20. Zeng X, Wang W, Deng G, et al. Prediction of potential disease-associated microRNAs by using neural networks. *Mol Ther Nucleic Acids* 2019;**16**:566–75.
21. Zhao H, Kuang L, Feng X, et al. A novel approach based on a weighted interactive network to predict associations of miRNAs and diseases. *Int J Mol Sci* 2018;**20**:110.
22. Ning Q, Zhao Y, Gao J, et al. AMHMDA: attention aware multi-view similarity networks and hypergraph learning for miRNA-disease associations identification. *Brief Bioinform* 2023;**24**:bbad094.
23. Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, New York, NY, USA, 2014, 701–10.
24. Yang W, Soares J, Greninger P, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2012;**41**:D955–61.
25. Ruan H, Xiang Y, Ko J, et al. Comprehensive characterization of circular RNAs in ~1000 human cancer cell lines. *Genome Med* 2019;**11**:55.
26. Rangwala SH, Kuznetsov A, Ananiev V, et al. Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). *Genome Res* 2021;**31**:159–69.
27. Wang Y, Bryant SH, Cheng T, et al. Pubchem bioassay: 2017 update. *Nucleic Acids Res* 2017;**45**:D955–63.
28. Landrum G. RDKit: a software suite for cheminformatics, computational chemistry, and predictive modeling. 2013. [http://www.rdkit.org/RDKit\\_Overview.pdf](http://www.rdkit.org/RDKit_Overview.pdf).
29. Van Laarhoven T, Nabuurs SB, Marchiori EJB. Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics* 2011;**27**:3036–43.
30. Wang W, Chen H. Predicting miRNA-disease associations based on graph attention networks and dual Laplacian regularized least squares. *Brief Bioinform* 2022;**23**:bbac292.
31. Jin T, Dai H, Cao L, et al. Deepwalk-aware graph convolutional networks. *Sci China Inf Sci* 2022;**65**:152104.
32. Li G, Luo J, Wang D, et al. Potential circRNA-disease association prediction using DeepWalk and network consistency projection. *J Biomed Inform* 2020;**112**:103624.
33. Kouhsar M, Khashaninia E, Mardani B, et al. CircWalk: a novel approach to predict CircRNA-disease association based on heterogeneous network representation learning. *BMC Bioinform* 2022;**23**:1–15.
34. Yang L, Li LP, Yi HC. DeepWalk based method to predict lncRNA-miRNA associations via lncRNA-miRNA-disease-protein-drug graph. *BMC Bioinform* 2021;**22**:1–15.
35. Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. 2013.
36. Dai Q, Liu Z, Wang Z, et al. GraphCDA: a hybrid graph representation learning framework based on GCN and GAT for predicting disease-associated circRNAs. *Brief Bioinform* 2022;**23**:bbac379.
37. Lou Z, Cheng Z, Li H, et al. Predicting miRNA–disease associations via learning multimodal networks and fusing mixed neighborhood information. *Brief Bioinform* 2022;**23**:bbac159.
38. Wang D, Cui P, Zhu W. Structural deep network embedding. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, New York, NY, USA, 2016, 1225–34.
39. Rees MG, Seashore-Ludlow B, Cheah JH, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol* 2016;**12**:109–16.
40. Li D, Yang YH, Lai RZ, et al. Status of chemical constituents and pharmacological activities of *Piper longum* L. *Chin J Clin Pharmacol* 2017;**33**:565–9.
41. Tripathi SK, Biswal BK. Piperlongumine, a potent anticancer phytotherapeutic: perspectives on contemporary status and future possibilities as an anticancer agent. *Pharmacol Res* 2020;**156**:104772.
42. Dai Y, Hartandi K, Ji Z, et al. Discovery of N-(4-(3-Amino-1 H-indazol-4-yl) phenyl)-N'-(2-fluoro-5-methylphenyl) urea (ABT-869), a 3-aminoindazole-based orally active multitargeted receptor tyrosine kinase inhibitor. *J Med Chem* 2007;**50**:1584–97.
43. Albert DH, Tapang P, Magoc TJ, et al. Preclinical activity of ABT-869, a multitargeted receptor tyrosine kinase inhibitor. *Mol Cancer Ther* 2006;**5**:995–1006.
44. Shankar DB, Li J, Tapang P, et al. ABT-869, a multitargeted receptor tyrosine kinase inhibitor: inhibition of FLT3 phosphorylation and signaling in acute myeloid leukemia. *Blood* 2007;**109**:3400–8.
45. Dagogo-Jack I, Shaw AT. Crizotinib resistance: implications for therapeutic strategies. *Ann Oncol* 2016;**27**:iii42–50.
46. Tarjányi O, Haerer J, Vecsernyés M, et al. Prolonged treatment with the proteasome inhibitor MG-132 induces apoptosis in PC12 rat pheochromocytoma cells. *Sci Rep* 2022;**12**:5808.
47. Ding Y, Lei X, Liao B, et al. Machine learning approaches for predicting biomolecule-disease associations. *Brief Funct Genom* 2021;**20**:273–87.
48. Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32.
49. Chen T, He T, Benesty M, et al. Xgboost: extreme gradient boosting. R package version 0.4–2. 2015;**1**:1–4. <https://cran.ms.unimelb.edu.au/web/packages/xgboost/vignettes/xgboost.pdf>.
50. Bo D, Shi C, Wang L, et al. Specformer: spectral graph neural networks meet transformers. arXiv preprint arXiv:2303.01028. 2023.