



# Using Graph Attention Network and Graph Convolutional Network to Explore Human CircRNA–Disease Associations Based on Multi-Source Data

Guanghui Li<sup>1\*</sup>, Diancheng Wang<sup>1</sup>, Yuejin Zhang<sup>1</sup>, Cheng Liang<sup>2</sup>, Qiu Xiao<sup>3</sup> and Jiawei Luo<sup>4\*</sup>

<sup>1</sup>School of Information Engineering, East China Jiaotong University, Nanchang, China, <sup>2</sup>School of Information Science and Engineering, Shandong Normal University, Jinan, China, <sup>3</sup>College of Information Science and Engineering, Hunan Normal University, Changsha, China, <sup>4</sup>College of Computer Science and Electronic Engineering, Hunan University, Changsha, China

## OPEN ACCESS

### Edited by:

Lihong Peng,  
Hunan University of Technology,  
China

### Reviewed by:

Min Chen,  
Hunan Institute of Technology, China  
Aiping Yao,  
Lanzhou University, China  
Fei Guo,  
Tianjin University, China

### \*Correspondence:

Guanghui Li  
ghli16@hnu.edu.cn  
Jiawei Luo  
luojiawei@hnu.edu.cn

### Specialty section:

This article was submitted to  
RNA,  
a section of the journal  
Frontiers in Genetics

Received: 06 December 2021

Accepted: 10 January 2022

Published: 07 February 2022

### Citation:

Li G, Wang D, Zhang Y, Liang C, Xiao Q  
and Luo J (2022) Using Graph  
Attention Network and Graph  
Convolutional Network to Explore  
Human CircRNA–Disease  
Associations Based on Multi-  
Source Data.  
Front. Genet. 13:829937.  
doi: 10.3389/fgene.2022.829937

Cumulative research studies have verified that multiple circRNAs are closely associated with the pathogenic mechanism and cellular level. Exploring human circRNA–disease relationships is significant to decipher pathogenic mechanisms and provide treatment plans. At present, several computational models are designed to infer potential relationships between diseases and circRNAs. However, the majority of existing approaches could not effectively utilize the multisource data and achieve poor performance in sparse networks. In this study, we develop an advanced method, GATGCN, using graph attention network (GAT) and graph convolutional network (GCN) to detect potential circRNA–disease relationships. First, several sources of biomedical information are fused *via* the centered kernel alignment model (CKA), which calculates the corresponding weight of different kernels. Second, we adopt the graph attention network to learn latent representation of diseases and circRNAs. Third, the graph convolutional network is deployed to effectively extract features of associations by aggregating feature vectors of neighbors. Meanwhile, GATGCN achieves the prominent AUC of 0.951 under leave-one-out cross-validation and AUC of 0.932 under 5-fold cross-validation. Furthermore, case studies on lung cancer, diabetes retinopathy, and prostate cancer verify the reliability of GATGCN for detecting latent circRNA–disease pairs.

**Keywords:** circRNA-disease associations, deep learning, graph attention network, graph convolutional network, centered kernel alignment

## INTRODUCTION

Circular RNA (circRNA) is a novel endogenous non-coding RNA forming a covalently closed loop structure, which lacks a 50-end cap and a 30-end ployA tail (Memczak et al., 2013; Meng et al., 2017). This structure is beneficial for circRNA to develop resistance to RNA exonuclease degradation and provides a more stable biological expression (Li et al., 2015). As a result, in most species, the average half-life of circRNAs is substantially increased than their linear equivalent. When circRNAs were first found as early as 1970s, they had been regarded as the abnormal shear or product of “shear noise,”

limited to the level of technology and knowledge at that time. In previous studies, multiple circRNAs were verified to be widespread in eukaryotes and play an essential role in biological functions with the advancement of biology and sequencing technologies. Currently, the biological functions of circRNA are reflected as follows (Rong et al., 2017): regulation of alternative splicing or transcription, miRNA sponges, regulation of protein binding, and generation of pseudogenes.

CircRNA has become a new biomarker due to its abundance, structural stability, developmental stage specificity, and tissue specificity (Zhang Z. et al., 2018), which can be discovered in saliva, blood, and exosomes. Cumulative research studies have confirmed that multiple circRNAs are significant to the expression of various pathological conditions (Han et al., 2018; Zhu et al., 2017; Zhang S. et al., 2018), especially cancer (Vo et al., 2019), cardiovascular, cerebrovascular, and nervous system diseases. For instance, circRNA hsa\_circ\_0027599 is overexpressed in gastric cancer (Wang L. et al., 2018), thereby regulating the expression of the gene PHLDA1 and promoting tumorigenesis. In cardiovascular and cerebrovascular diseases, circRNA circWDR77Z targets and regulates miRNA miR-124/FGF-2 through the “sponge” function (Chen et al., 2017), which affects the migration and proliferation for vascular smooth muscle cells, thereby promoting atherosclerosis development. For myocardial infarction, overexpression of circRNA CDR1 leads to the upregulation of downstream corresponding enzymes and proteins (Zhang et al., 2016), thereby aggravating myocardial infarction. In neurological diseases, the expression of circRNA in brain tissue is different, and its distribution in the brain is uneven (Zhang et al., 2021b).

Although circRNA is commonly expressed in various cell lines and tissues with strong tissue specificity and development stage specificity, the pathogenic mechanism of circular RNA and how it interacts with other biological molecules remain unknown. In recent years, researchers have established many experimentally verified or reported databases on relationships between circRNAs and diseases, such as circBase (Glažar et al., 2014), circRNADb (Chen et al., 2016), circR2Disease (Fan et al., 2018b), circRNADisease (Zhao et al., 2018), circ2Disease (Yao et al., 2018) and circ2Traits databases (Ghosal et al., 2013). Considering that conventional biological studies are cost-ineffective and time-consuming, several computational approaches have been designed to detect relationships between diseases and circRNAs efficiently (Xiao et al., 2022; Lei et al., 2021). At present, the proposed computational models for discovering relationships between diseases and circRNAs are mainly classified into the following groups:

Network propagating methods have been widely applied to detect correlations between diseases and various biological entities, including circRNAs, due to the efficient use of network structure information (Peng et al., 2018). Zhang et al. designed a linear neighbor marker propagation approach named CD-LNLP *via* neighbor similarity to reveal relationships between diseases and circRNAs (Zhang et al., 2019). Li et al. presented the DWNPCDA using DeepWalk and network consistency projection (Chen et al., 2018) to detect unobserved associations between diseases and circRNAs (Li G. et al.,

2020). Lei et al. constructed a prediction model named RWRKNN, which combined the k-nearest neighbor and RWR to calculate weighted features for diseases and circRNAs (Lei and Bian, 2020).

Path-based methods are widely adopted to calculate potential interactions between diseases and circRNAs by measuring the weight of paths in different networks. Lei et al. presented a path-weighted method named PWEDA, which predicted the circRNA–disease relationships by calculating the probability value for each circRNA–disease pair *via* path information (Lei et al., 2018). Fan et al. presented the model named KATZHEDA *via* the circRNA expression profile, the similarity of the disease phenotype, and the nuclear similarity of the Gaussian interaction profile using the KATZ method to detect potential interactions between diseases and circRNAs through the heterogenous network (Fan et al., 2018a). Zhao et al. revealed a computed method named IBNPKATZ using the bipartite network projection model and the KATZ (Zhang et al., 2021a) model to discover circRNA–disease interactions (Zhao et al., 2019).

Matrix factorization–based methods have been carried out for detecting circRNA–disease relationships by constructing a low-dimensional matrix to represent the initial input features (Wang P. et al., 2018; Peng et al., 2020a). Wei et al. used weight-based nearest neighbor nodes to reconstruct the association matrix and designed a graph regularized non-negative matrix factorization algorithm iCircDA-MF to detect relationships between diseases and circRNAs (Wei and Liu, 2020). Lu et al. constructed a model named DMFCDA with deep matrix factorization, which infers potential circRNA–disease interactions by mapping features of diseases and circRNAs into low-dimensional spaces (Lu et al., 2021). Yan et al. used the Kronecker product kernel to design a regularized least squares algorithm called DWNN-RLS to detect relationships (Yan et al., 2018). Li et al. presented an advanced approach named SIMCCDA by regarding predicting associations as a recommendation system task, which achieves outstanding performance for discovering circRNA–disease associations (Li M. et al., 2020).

Deep learning integrates low-level features to construct high-level representations of features or attribute categories through the deep non-linear network structure (Peng et al., 2021; Zhou et al., 2021). Wang et al. designed a model to reveal interactions between diseases and circRNAs using deep convolutional neural networks and deep generative adversarial networks (Wang et al., 2020a). Wang et al. designed an approach named GCNCDA to identify disease-related circRNAs, which extracts high-level features contained in the circRNA–disease heterogenous network through graph convolutional networks to calculate association scores (Wang et al., 2020b). GATCDA is a novel model for discovering the correlation between diseases and circRNAs, which learns the latent representation of nodes by assigning corresponding weights to each neighbor node (Bian et al., 2021). Xiao et al. designed a computational model named NSL2CD that adopts network embedding by adaptive subspace learning (Xiao et al., 2021).

Although the abovementioned approaches have achieved excellent predictive performance, there are still several

limitations given as follows: First, network-based methods achieve poor performance in sparse networks due to a small amount of network structure information. Second, path-based methods fail to dynamically calculate weights based on known associations, which makes it unable to efficiently detect relationships between diseases and circRNAs with new diseases or circRNAs. Third, matrix factorization-based methods could not discover a non-linear interaction between diseases and circRNAs. Last, current deep learning-based methods could not effectively utilize the multisource data and only pay more attention to features of the neighbor nodes or the node itself, respectively.

To solve the abovementioned challenges, we develop an advanced method GATGCN *via* graph attention network (GAT) and graph convolutional network (GCN) to detect potential circRNA–disease relationships. The complete process could be summarized as four steps: First, multisource similarity data for circRNAs and diseases are fused by the centered kernel alignment model (CKA) (Cristianini et al., 2006). Second, we adopt the graph attention network to learn the dense representation of nodes on fused disease similarity network and fused circRNA similarity network. Third, we construct the heterogenous network by connecting circRNA–disease interaction network, feature matrix of diseases, and feature matrix of circRNAs. Finally, the graph convolutional network is adopted to get prediction scores based on the heterogenous network. According to reliable computer experiments, GATGCN outperforms several state-of-the-art methods with a prominent AUC of 0.932.

## MATERIALS

### Human CircRNA–Disease Associations

The circR2Disease provides verified relationships between diseases and circRNAs, which is a manually curated database including 739 known relationships between 100 diseases and 676 circRNAs. We eventually extract 661 associations between 88 diseases and 585 circRNAs for humans after removing the associations unrelated to human species and duplicate associations.

### Human Disease–MiRNA Associations

MiRNAs are significant to pathogenesis and treatment of diseases as the important regulatory molecule for genes. On dataset, we collect 1,883 experimentally verified disease–miRNA relationships between 462 miRNAs and 88 diseases from the HMDD (Li et al., 2014), which provides disease-associated miRNAs and their target genes, including 8,802 known relationships between 350 diseases and 32281 miRNAs.

### Human Disease–Gene Associations

Due to gene mutation and expression affecting diseases, diseases are closely related to genes. On the dataset, 74 experimentally verified disease–gene associations between 61 genes and 88 diseases are filtered out, downloaded from <http://cssb2.biology.gatech.edu/knowgene/>.

### Human CircRNA–MiRNA Associations

With plenty miRNA binding sites (Hansen et al., 2013; Peng et al., 2020b), circRNAs actively affect the expression of miRNA's downstream genes as miRNA sponges (Peng et al., 2017; Zeng et al., 2020). We obtain 17844 known circRNA–miRNA associations between 640 miRNAs and 585 circRNAs from ENCORI (available at <http://starbase.sysu.edu.cn/agoClipRNA.php?source=circRNA>).

### Human CircRNA–Gene Associations

According to the previous research, circRNAs are verified to be significant in regulating the expression of genes. On the dataset, 487 known circRNA–gene associations between 418 genes and 585 circRNAs are extracted from <http://cssb2.biology.gatech.edu/knowgene/search.html>.

### Disease Semantic Similarity

The semantic information of the diseases has been wildly adopted to measure the similarity of diseases because of its effectiveness and stability. In this study, we obtain the related annotation terms for each disease from MeSH.

In MeSH, the directed acyclic graph (DAG) is applied to represent the semantic relationship among diseases, in which nodes denote corresponding disease information and directed edges denote the relationship among diseases. Specifically, disease  $d_i$  can be described as three items  $DAG_i = [d_i, T(d_i), E(d_i)]$ , where  $T(d_i)$  represents  $d_i$  itself and its ancestor nodes and  $E(d_i)$  is relationships between  $d_i$  and all diseases. The contribution of disease  $d_i$  in  $DAG_i$  is formulated as follows:

$$\begin{cases} D_{d_i}(n) = 1 & \text{if } n = d \\ D_{d_i}(n) = \max\{\sigma \cdot D_{d_i}(n') \mid n' \in \text{children of } n\} & \text{if } n \neq d \end{cases} \quad (1)$$

where  $\sigma$  denotes the attenuation factor for semantic contribution, which is defined as the optimal value of 0.5 according to Wang's experience Wang et al. (2010);  $n'$  represents the child node of the node  $n$ . Therefore, the overall semantic score of the disease  $d_i$  is measured by accumulating the contribution scores from its ancestor diseases and itself as follows:

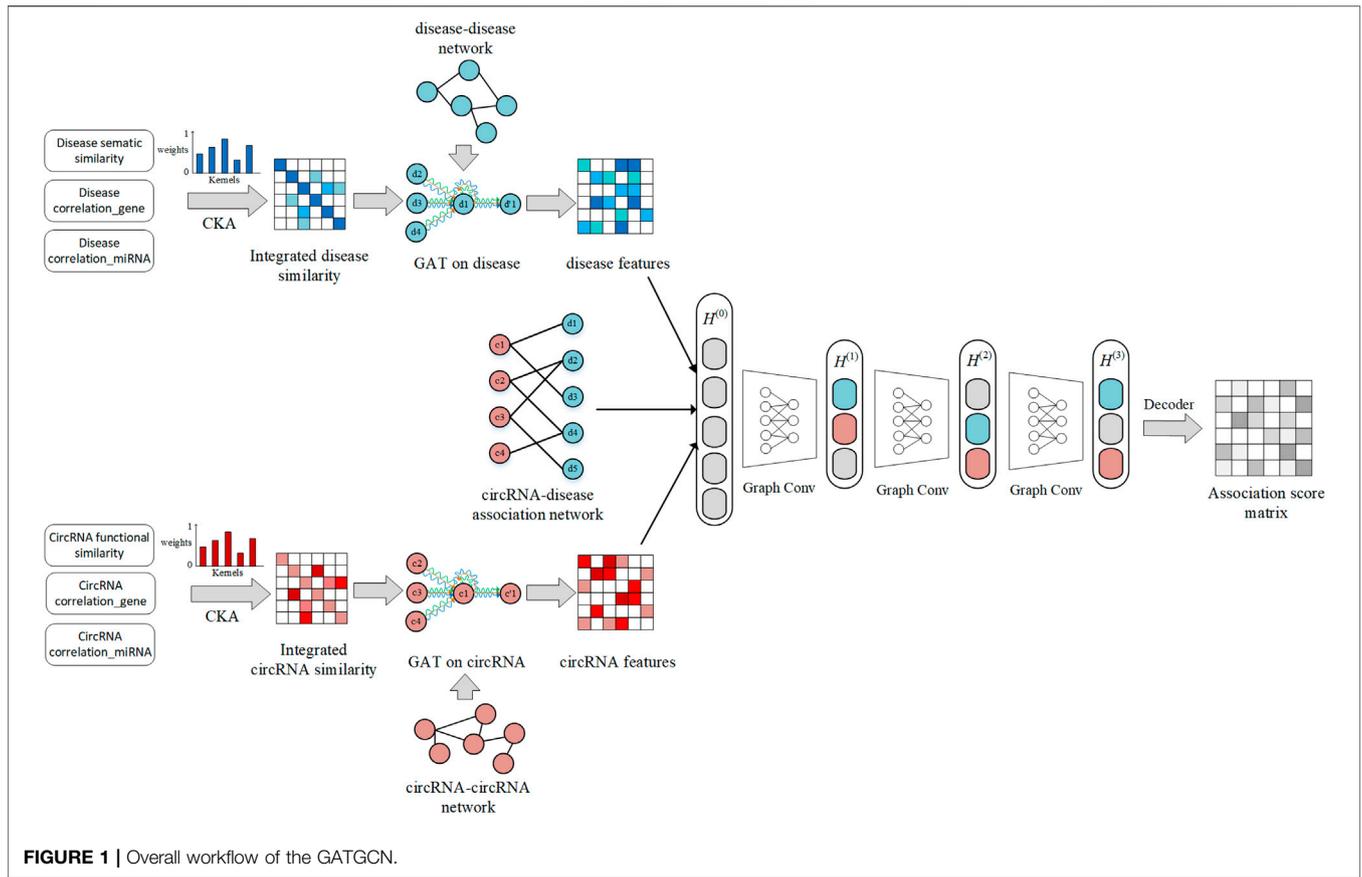
$$D(d_i) = \sum_{n \in T(d_i)} D_{d_i}(n). \quad (2)$$

In general, diseases with more common parts shared in the DAG achieve higher semantic similarities. Based on this hypothesis, the value of disease semantic similarity between disease  $d_i$  and disease  $d_j$  is formulated *via* Eq.3:

$$DS(d_i, d_j) = \frac{\sum_{n \in T_{d_i} \cap T_{d_j}} (D_{d_i}(n) + D_{d_j}(n))}{D(d_i) + D(d_j)}. \quad (3)$$

### CircRNA Functional Similarity

According to previous studies, circRNAs that are relevant to more similar diseases are prone to be more similar in functions (Li et al., 2019). Then, the BMA method is deployed to measure the functional similarity score among different circRNAs according to relevant disease sets. Given a specific disease  $d_i$



**FIGURE 1 |** Overall workflow of the GATGCN.

and  $D = (d_1, d_2, \dots, d_t)$ , the score of functional similarity between circRNA  $c_i$  and circRNA  $c_j$  is measured via Eqs 4, 5:

$$FS(c_i, c_j) = \frac{\sum_{m=1}^{|D_i|} S(d_m, D_j) + \sum_{n=1}^{|D_j|} S(d_n, D_i)}{|D_i| + |D_j|}, \quad (4)$$

$$S(d_m, D_j) = \max_{1 \leq t \leq |D_j|} (S(d_m, d_t)), \quad (5)$$

where  $D_j$  represents the collection of diseases associated with circRNA  $c_j$ .  $S(d_m, D_j)$  represents the similarity between disease  $d_m$  associated with circRNA  $c_i$  and disease collection  $D_j$  associated with circRNA  $c_j$ .

### Pearson’s Correlation Coefficient Similarity

Since the circRNA functional similarity network and the disease semantic similarity network are prone to be sparse, we adopt Pearson’s correlation coefficient approach to enrich multisource similarity data by calculating the linear correlation among different variables. To be specific, the value of Pearson’s correlation between variable  $M$  and variable  $N$  is measured as follows:

$$Cor(M, N) = \frac{cov(M, N)}{\sqrt{var(M)var(N)}} \quad (6)$$

where  $var(M)$  measures the variance of  $M$ ;  $cov(M, N)$  calculates the covariance between  $M$  and  $N$ ; the value of  $Cor(M, N)$  ranges

from  $-1$  to  $1$ , which reflects the strength of the linear correlation between  $M$  and  $N$ .

Four binary networks have been built including the disease–gene network, circRNA–miRNA network, circRNA–gene network, and disease–miRNA network. Then, Pearson’s correlation coefficient approach is adopted to compute disease similarity and circRNA similarity via corresponding bipartite networks. The equation is computed as follows:

$$Cor(n_i, n_j) = \frac{cov(IP(n_i), IP(n_j))}{\sqrt{var(IP(n_i))var(IP(n_j))}} \quad (7)$$

where  $IP(n_i)$  denotes the  $i$ th row of the corresponding association network.  $Cor(n_i, n_j)$  denotes the value of Pearson’s correlation similarity between node  $n_i$  and node  $n_j$  based on the corresponding association network.

## METHODS

In this work, we develop an advanced method GATGCN via the graph attention network and graph convolutional network to detect potential circRNA–disease relationships. As shown in Figure 1, the complete process could be summarized in four

steps: First, the CKA-based model is adopted to fuse multisource similarity data for circRNAs and diseases. Second, we adopt the graph attention network to calculate the dense representation of nodes on the fused disease similarity network and fused circRNA similarity network. Third, we construct the heterogenous network, including circRNA–disease interactions network, feature matrix of diseases, and feature matrix of circRNAs. Eventually, the graph convolutional network is adopted to get prediction scores based on the constructed heterogenous network.

## Centered Kernel Alignment

In previous studies, multisource data are usually fused by calculating the average value, which ignores the importance among different kernels. Thus, the centered kernel alignment (CKA) model (Wang et al., 2021) is adopted to fuse several kinds of similarities for diseases and circRNAs based on different weights. We consider  $K_d = \{K_{d_1}^1, \dots, K_{d_v}^v\}$  and  $K_c = \{K_{c_1}^1, \dots, K_{c_u}^u\}$  as different kernels for disease space and circRNA space. The  $v$  and  $u$  denote the number of kernels from disease space and circRNA space, respectively. Meanwhile, the basic CKA model (Cristianini et al., 2006) is used as the objective of MKL (Ding et al., 2019) to measure the corresponding weight of each kernel.

To be specific, the kernels  $K_c^*$  and  $K_d^*$  based on optimal weight are calculated as follows:

$$K_c^* = \sum_{p=1}^u \alpha_c^p K_c^p, \quad K_c^p \in R^{m \times m}, \quad (8)$$

$$K_d^* = \sum_{p=1}^v \alpha_d^p K_d^p, \quad K_d^p \in R^{n \times n}, \quad (9)$$

where  $\alpha_c = \{\alpha_c^1, \dots, \alpha_c^u\}$  and  $\alpha_d = \{\alpha_d^1, \dots, \alpha_d^v\}$ .

Basic CKA (Cristianini et al., 2006) is adopted to calculate the weights of each kernel on the training set. The kernel alignment score between the two kernels is formulated as follows:

$$U(E, I) = \frac{\langle E, I \rangle_F}{\|E\|_F \|I\|_F}, \quad (10)$$

where  $E, I$  denotes the corresponding similarity matrix,  $\|E\|_F$  denotes the Frobenius norm, and  $\langle E, I \rangle = \text{Trace}(E^T I)$  denotes the Frobenius inner product. The kernel alignment score represents the similarity among different kernels. Specifically, the kernel alignment score between the similarity kernel (disease kernel or circRNA kernel) and the ideal kernel matrix is measured as follows:

$$\max_{\beta \geq 0} CU(K^*, K_{ideal}) = \max_{\beta \geq 0} \frac{\langle Z_N K^* Z_N, K_{ideal} \rangle_F}{\|Z_N K^* Z_N\|_F \|K_{ideal}\|_F}, \quad (11)$$

$$\text{subject to } K^* = \sum_{p=1}^N \beta^p K^p \quad \beta \geq 0, p = 1, 2, \dots, N, \quad (12)$$

$$\sum_{p=1}^N \beta^p = 1, \quad (13)$$

where  $K_{ideal}$  denotes a label kernel constructed by known associations;  $K_{ideal, d} = Y_{train}^T Y_{train} \in R^{n \times n}$  and  $K_{ideal, c} = Y_{train} Y_{train}^T \in R^{m \times m}$  denote the ideal kernel of diseases and circRNAs, respectively.

## Attention Mechanism on Similarity

Considering that current methods did not capture potential features on the similarity network, we adopt the graph attention method to learn latent representation of diseases and circRNAs, which assigns corresponding weights to different node features based on the local graph structure to ignore noise and redundancy. The advantage of the attention mechanism is to directly evaluate which features are preferred embedding for specific downstream tasks by calculating the weights. First, we obtain the corresponding association matrix by setting a threshold on the similarity network for diseases and circRNAs. Then, the GAT (Veličković et al., 2017) is applied to learn dense representation for diseases and circRNAs as follows:

The input layer of the graph attention network is formulated as follows:

$$f = \{f_1, f_2, \dots, f_N\}, f_i \in R^F, \quad (14)$$

where  $F$  denotes the dimension of features, and  $N$  represents the number of nodes in the corresponding similarity network.  $f \in R^{N \times F}$  is constructed by the features of nodes in the corresponding similarity network. The output layer of the graph attention network is defined as follows:

$$f' = \{f'_1, f'_2, \dots, f'_i\}, f'_i \in R^{F'}, \quad (15)$$

where  $F'$  denotes the length of learned features, and  $f' \in R^{N \times F'}$  represents the learned latent representations of nodes in the network. The first step is to calculate the weight of the corresponding neighbor node. The importance of the given nodes is computed by the self-attention mechanism. For each association pair between node  $n_i$  and node  $n_j$ , the attention coefficient  $e_{ij}$  is calculated as follows:

$$e_{ij}(n_i, n_j) = \text{att}(W f_i, W f_j), \quad (16)$$

where  $\text{att}$  represents a mapping function transforming high-level features to a real number for association pair between node  $n_i$  and node  $n_j$  based on input features, and  $W \in R^{F \times F'}$  denotes a trainable weight matrix. To avoid the influence of dimension between different attention coefficients,  $e_{ij}$  is further normalized as follows:

$$\theta_{ij} = \text{soft max}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{t \in N_i} \exp(e_{it})}, \quad (17)$$

where  $N_i$  represents the collection of neighbor nodes of node  $n_i$ .  $\theta_{ij}$  denotes the normalized weight representing the importance between node  $n_i$  and node  $n_j$  in the network.

From the abovementioned formula, we obtain the combined attention mechanism as follows:

$$\theta_{ij} = \frac{\exp(\text{leakyRelu}(a^T [W f_i \| W f_j]))}{\sum_{t \in N_i} \exp(\text{leakyRelu}(a^T [W f_i \| W f_t]))}, \quad (18)$$

where  $\text{leakyRelu}$  denotes a non-saturated activation function, which can solve the vanishing gradients and accelerate convergence.  $a \in R^{2F'}$  denotes the weight matrix, which maps features to a real number. The second step is to aggregate the features of all neighbors for a given node by integrating the

corresponding weight. The aggregation between the given node and neighbors is formulated as follows:

$$f'_i = \sigma\left(\sum_{t \in N_i} \theta_{it} W f_t\right) \quad (19)$$

where  $\sigma$  denotes a non-saturated activation function. Multi-head attention mechanism is applied in GAT to integrate features and prevent overfitting. The output with the multi-head attention mechanism contains the features in different representation subspaces, which enhances the expressive capacity of the model. To be specific, the multi-head attention model based on the combination of  $K$ -independent attention mechanisms learns latent features as follows:

$$f'_i = \sigma\left(\frac{1}{K} \sum_{K=1}^K \sum_{t \in N_i} \theta_{it}^k \cdot W^K f_t\right), \quad (20)$$

where  $K$  represents the number of self-attention models.  $W^k$  denotes the trained weight matrix of the  $k$ th attention model.

## Heterogenous Network

The heterogenous network is constructed as initial features of GCN, including circRNA–disease associations, learned feature matrix of circRNAs, and learned feature matrix of diseases. The binary matrix  $A$  is constructed, and  $A_{ij} = 1$  if the interaction between circRNA  $c_i$  and disease  $d_j$  has been verified; otherwise  $A_{ij} = 0$ . The learned feature matrix of circRNAs and learned feature matrix of diseases based on GAT are denoted as matrix  $S_c$  and matrix  $S_d$ , respectively. The heterogenous network  $A_H$  is defined as follows:

$$A_H = \begin{bmatrix} S_c & A \\ A^T & S_d \end{bmatrix} \in R^{(M+N) \times (M+N)}. \quad (21)$$

## Graph Convolutional Network on Heterogenous Network

In recent years, GCN has achieved superior performance in node prediction, node classification, and edge prediction tasks (Kipf and Welling, 2016). In order to discover potential relationships between diseases and circRNAs, GCN models (Wang et al., 2020b) are designed to effectively extract features of circRNA–disease relationships based on the global graph structure by aggregating feature vectors of neighbors. To be specific, given a network  $G$ , each layer of the GCN model embedding is formulated as follows:

$$H^{(l+1)} = \sigma\left(D^{-\frac{1}{2}} G D^{-\frac{1}{2}} H^{(l)} W^{(l)}\right), \quad (22)$$

where  $H^{(l)}$  denotes the propagation of features at the  $l$ th layer,  $\sigma(\cdot)$  represents a nonlinear activation function,  $D = \text{diag}(\sum G_{ij})$  denotes the degree matrix of  $G$ , and  $W^{(l)}$  is the trained weight matrix at the  $l$ th layer. GCN integrates low-level features to construct high-level representations of nodes on the constructed heterogenous network  $A_H$ . In addition, we adjust the number of graph convolutional network layers and set node dropout to avoid overfitting, which can reduce excessive

parameters and improve the generalization ability of the GATGCN. The penalty factor  $\mu$  is set to regulate the contribution of learned similarity features in the embedding of graph convolutional layers. Specifically, the input heterogenous network  $G$  is defined as follows:

$$G = \begin{bmatrix} \mu \cdot S_c & A \\ A^T & \mu \cdot S_d \end{bmatrix}. \quad (23)$$

Then, the initial embedding is defined as follows:

$$H^{(0)} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}. \quad (24)$$

The first layer of the GCN model embedding is calculated as follows:

$$H^{(1)} = \sigma\left(D^{-\frac{1}{2}} G D^{-\frac{1}{2}} H^{(0)} W^{(0)}\right), \quad (25)$$

where  $W^{(0)} \in R^{(M+N) \times k}$  represents an input-to-hidden trained weight matrix,  $H^{(1)} \in R^{(M+N) \times k}$  represents the first-layer propagation of features, including circRNAs and diseases.  $k$  denotes the embedding dimension in graph conventional layers. We adopt the exponential linear unit (Clevert et al., 2016) as the nonlinear activation function to enhance noise robustness and expressive capacity of the model in all graph convolutional layers. Eventually, the bilinear decoder  $A'$  proposed by Huang et al., (2020) is deployed to reconstruct the circRNA–disease association matrix as follows:

$$A' = \text{sigmoid}(H_C W' H_D^T), \quad (26)$$

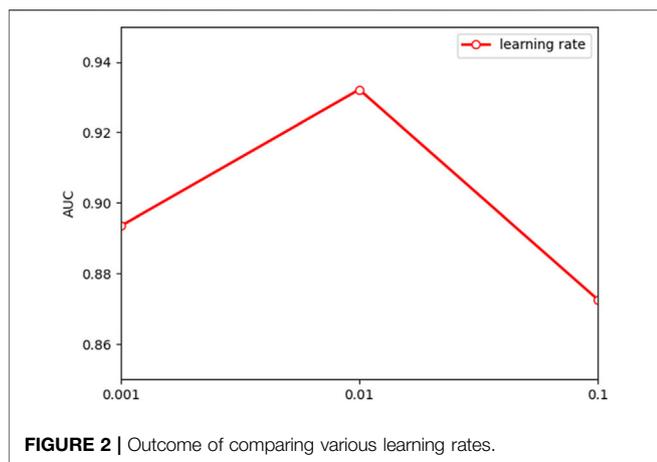
where  $W' \in R^{k \times k}$  denotes a trained weight matrix.  $H_D \in R^{N \times k}$  and  $H_C \in R^{M \times k}$  represent the last embedding for diseases and circRNAs, respectively. The final predicted relationship score  $a'_{ij}$  between circRNA  $c_i$  and disease  $d_j$  is obtained according to the corresponding  $(i, j)$ th entry of  $A'$ .

## RESULTS

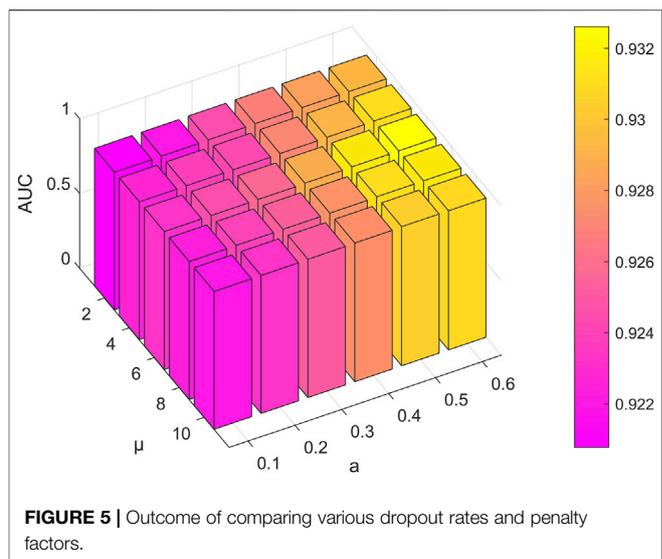
In this section, several verification experiments are deployed to assess the predictive capacity of GATGCN. First, we assess the influence of different parameters setting on GATGCN. Second, we introduce the evaluation metrics under leave-one-out cross-validation and 5-fold cross-validation to analyze the predictive capacity of GATGCN. Third, we design the ablation study to assess the impact of each part on GATGCN. Fourth, we discuss and compare GATGCN with state-of-the-art models on the same dataset. Last, case studies are deployed to further assess the performance in detecting potential relationships on GATGCN.

### Parameter Setting

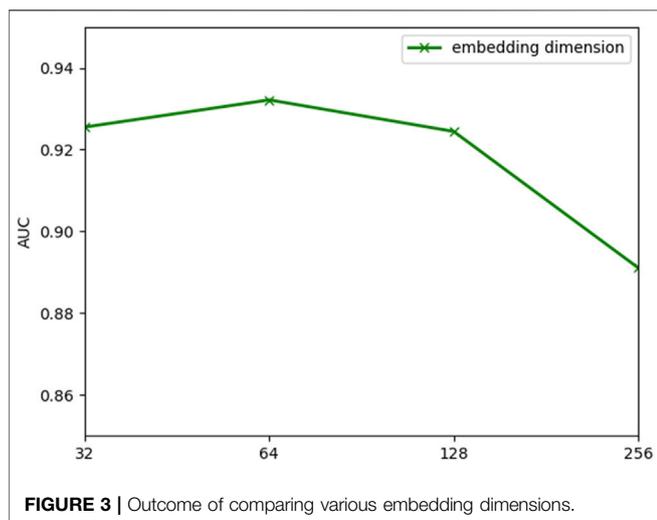
The performance of the model is frequently impacted by hyperparameter settings. Analysis of the parameters can quantitatively evaluate the stability of the model and provide a reference for parameter selection. The learning rate is significant to the convergence of the gradient descent algorithm in the model. **Figure 2** indicates that the model will converge slowly



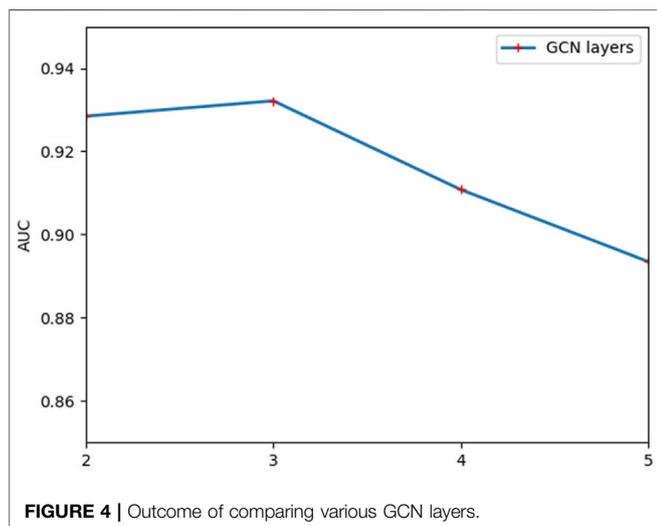
**FIGURE 2** | Outcome of comparing various learning rates.



**FIGURE 5** | Outcome of comparing various dropout rates and penalty factors.



**FIGURE 3** | Outcome of comparing various embedding dimensions.



**FIGURE 4** | Outcome of comparing various GCN layers.

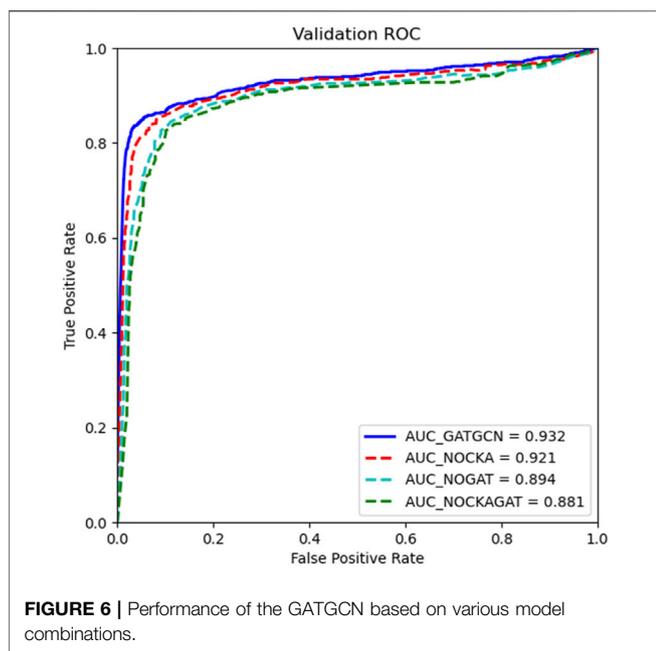
with too small a learning rate, while too large a learning rate makes it hard to converge. According to the results in **Figure 3**, the embedding dimension within a certain size range has less impact on the convergence of our model. However, when the embedding dimension is too large, the model is prone to overfitting due to plenty of parameters. As shown in **Figure 4**, the model performs better with small layers of the graph convolutional network, and the performance drops significantly when the number of layers of GCN is  $l > 4$ . The reason is that the GCN with more layers not only captures more global prior information but also captures a lot of noise at the same time. Meanwhile, the penalty factor  $\mu$  is set to regulate the contribution of learned similarity features in the propagation of convolutional layers, and the dropout rate  $a$  is adopted to avoid overfitting. As shown in **Figure 5**, the model achieves best performance at  $\mu = 6$  and  $a = 0.6$ .

### Evaluation Metrics

Cross-validation is a self-consistent testing approach widely adopted to demonstrate the predictive capacity of a method. The basic idea is to carry out the resampling method to select a portion of the benchmark data set as the training set to train the model, and the remaining samples to verify the model. Five-fold cross-validation and leave-one-out cross-validation are deployed to assess the predictive capacity of GATGCN. For five-fold cross-validation, the whole samples in the dataset are randomly separated into five roughly identical sections, four of which are adopted to train the GATGCN and the other is used to test the GATGCN. In order to decrease the bias produced by sample segmentation, the five-fold cross-validation is repeated 30 times to calculate the average result as the ultimate output. For leave-one-out cross-validation, each time only one sample in the dataset is selected among all recorded circRNA–disease relationships to test the model, and the remaining known relationships are utilized as training samples. In this study, since circRNA functional similarity relies on known associations; we recalculate the circRNA functional similarity in each repetition of the experiment.

**TABLE 1** | Results generated by the GATGCN under five-fold CV and LOOCV.

Test set	Accu	Rec	Spe	F1	AUC
5-fold CV_1	0.988	0.682	0.989	0.437	0.956
5-fold CV_2	0.987	0.568	0.991	0.361	0.918
5-fold CV_3	0.987	0.644	0.988	0.373	0.922
5-fold CV_4	0.990	0.627	0.991	0.414	0.931
5-fold CV_5	0.991	0.647	0.990	0.402	0.934
Average	0.9886 ± 0.0024	0.6336 ± 0.0656	0.9898 ± 0.0012	0.3974 ± 0.0396	0.9322 ± 0.0238
LOOCV	0.987	0.782	0.992	0.542	0.951



In this study, the area under the curve (AUC) is applied as the primary metric to assess our model, which can visually show the predictive ability of GATGCN under each decision threshold. The basic principle is to treat the false-positive rate (FPR) and the true rate (TPR) as a two-dimensional coordinate point in a Cartesian coordinate system with FPR as the abscissa and TPR as the ordinate under different discrimination thresholds. Besides, several threshold-based metrics are adopted to further evaluate the predictive performance of the GATGCN including recall, specificity, accuracy, and F1. The detailed results of five-fold cross-validation and leave-one-out cross-validation are summarized in **Table 1**.

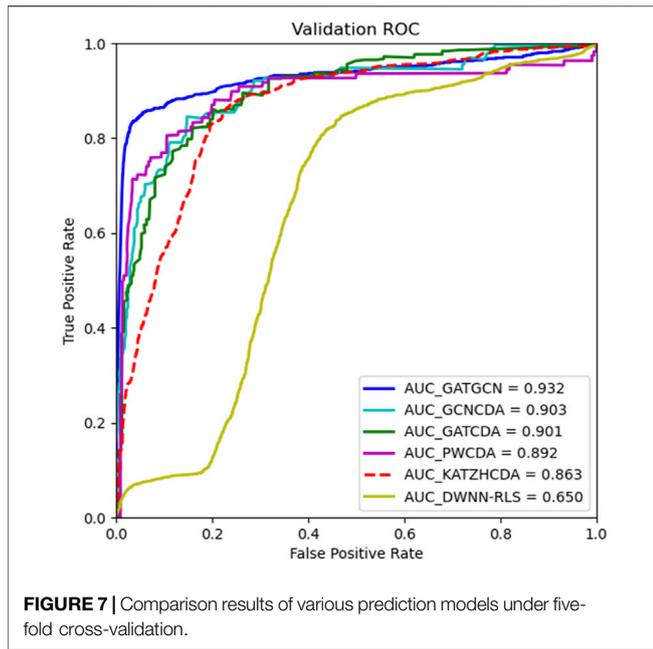
## Ablation Study

The model GATGCN is used to detect potential relationships between diseases and circRNAs based on the centered kernel alignment model (CKA), graph attention network (GAT), and graph convolutional network (GCN). In order to verify the importance of CKA, GAT, and GCN in our model, we apply the ablation study to our model. In this part, we replace the CKA model with calculated average to fuse multisource similarity as NOCKA. Meanwhile, we only combine the

CKA model and GCN model as NOGAT to calculate association scores. In addition, we only adopt the GCN to predict associations between diseases and circRNAs as NOCKAGAT. According to the results in **Figure 6**, the complete model GATGCN is compared with NOCKA, NOGAT, and NOCKAGAT with five-fold cross-validation, which achieves the best AUC of 0.932. In general, using the the graph attention network on the similarity network is beneficial to learn the latent representation of nodes. The AUC of GATGCN and NOCKA is significantly higher than that of the other two models, which indicates that GAT is significant to detect relationships between diseases and circRNAs. Moreover, the comparison between GATGCN and NOCKA suggests that the fusion of multisource similarity based on weights can improve performance in circRNA–disease relationship prediction.

## Comparison With Other Methods

To confirm the advantage of GATGCN, we compare it with several classic prediction methods with five-fold cross-validation. Since these methods adopt various datasets and evaluation metrics, we apply the same dataset and AUC as the metrics to compare the predictive capacity of models fairly and reasonably. In this part, the GATGCN is compared with several state-of-the-art methods, including KATZHCDA (Fan et al., 2018a), DWNN-RLS (Yan et al., 2018), PWCDA (Lei et al., 2018), GCNCDA (Wang et al., 2020b), and GATCDA (Bian et al., 2021). KATZHCDA is a graph-based method that uses the walking lengths and number of walks among nodes to measure the similarity among nodes in the heterogenous network. The DWNN-RLS measures initial relational values of new diseases and circRNAs *via* the decreasing weight k-nearest neighbor model and adopts the Kronecker product kernel to predict associations between diseases and circRNAs. The PWCDA predicts the circRNA–disease relationships by searching the paths without repeating for all circRNA–disease pairs based on the constructed heterogenous network. The GCNCDA extracts high-level features in the heterogenous network through graph convolutional neural networks and predicts the correlation between circRNAs and diseases *via* Forest by Penalizing Attributes. GATCDA learns the latent representation of nodes by assigning corresponding weights to each neighbor node, which efficiently aggregates the information of neighbor nodes and utilizes the local features of the graph. The results in **Figure 7** indicate that



GATGCN achieves the best AUC of 0.932, which is substantially greater than that of other models, and produces 7.9%, 43.3%, 4.5%, 3.2%, and 3.4% improvement in the AUC compared with KATZHCDA, DWNN-RLS, PWCDA, GCNCDA, and GATCDA respectively.

Furthermore, the number of known interactions between diseases and circRNAs in the dataset can greatly affect the performance of the method, which also indicates the

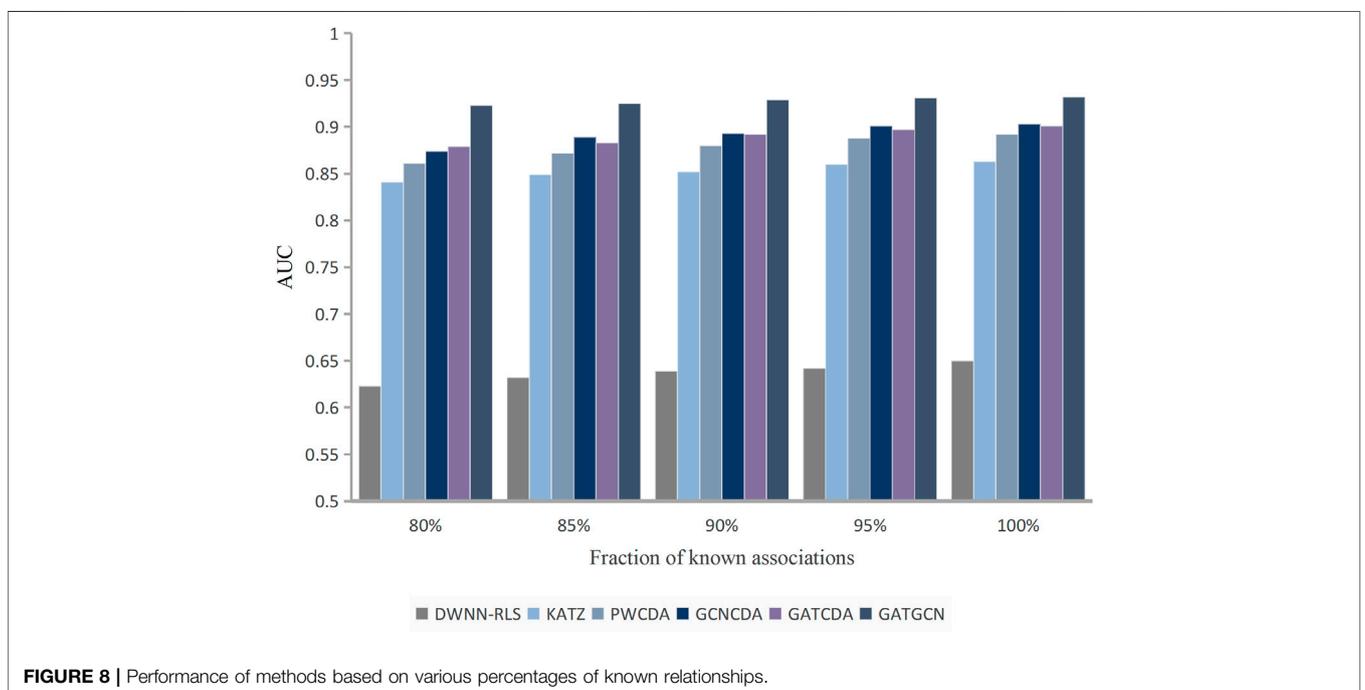
**TABLE 2 |** Top 10 candidate circRNAs related to lung cancer.

Rank	circRNA	Evidence (PMID)
1	hsa_circ_0007385	29372377
2	hsa_circ_0014130	29440731
3	hsa_circ_0016760	33416186
4	hsa_circ_0043256	28958934
5	hsa_circ_0012673	32141533
6	hsa_circRNA_404833	unconfirmed
7	hsa_circRNA_006411	unconfirmed
8	hsa_circRNA_401977	unconfirmed
9	hsa_circ_0013958	28685964
10	hsa_circ_0006404	unconfirmed

robustness of the method. Thus, we randomly remove a part of known associations between diseases and circRNAs at a ratio  $r \in \{80\%, 85\%, 90\%, 95\%, \text{ and } 100\%\}$  with five-fold cross-validation. As shown in **Figure 8**, the performance of GATGCN improves with increasingly known associations. Meanwhile, the GATGCN achieves the best result across different data richness among KATZ, DWNN-RLS, PWCDA, GCNCDA, and GATCDA.

### Case Studies

In this part, two kinds of case studies are utilized to further assess the reliability of the GATGCN for detecting potential circRNA–disease associations, which calculated the predicted probability matrix *via* a candidate set comprising unproven circRNAs. For the first kind of case study, all known circRNA–disease relationships are selected as training samples, and all unknown circRNA–disease relationships are prioritized



**TABLE 3** | Top 10 candidate circRNAs related to diabetes retinopathy.

Rank	circRNA	Evidence (PMID)
1	hsa_circRNA_063981	28817829
2	hsa_circRNA_404457	28817829
3	hsa_circRNA_100750	28817829
4	hsa_circRNA_406918	28817829
5	hsa_circRNA_104387	28817829
6	hsa_circRNA_103410	28817829
7	hsa_circRNA_100192	28817829
8	hsa_circ_0013509	unconfirmed
9	circSLC8A1-1	unconfirmed
10	hsa_circ_101396	unconfirmed

**TABLE 4** | Top 10 candidate circRNAs related to prostate cancer.

Rank	circRNA	Evidence (PMID)
1	circHIPK3	32547085
2	hsa_circ_0004383	unconfirmed
3	circ-Foxo3	31733095
4	hsa_circRNA_2149	unconfirmed
5	circR-284	unconfirmed
6	circDLGAP4	unconfirmed
7	hsa_circ_0008887	unconfirmed
8	hsa_circ_0044516	31625175
9	CDR1as	23900077
10	Cir-ITCH	32904490

according to the corresponding prediction scores. We select the top 10 scores by sorting the scores of the probability matrix in descending order and verified those predicted candidates through validated databases and literature, such as CircR2Disease, CircBase, and PubMed. Eventually, we adopt case studies on lung cancer, diabetes retinopathy, and prostate cancer.

Lung cancer occurs in the bronchial mucosa or glands with the highest incidence and the highest number of deaths in the world. The results in **Table 2** show that six associations are verified by experiments among top 10 predicted candidate circRNAs for lung cancer. For example, the hsa\_circ\_0007385 (top 1) knockdown resulted in considerable inhibition of the proliferation, invasion, and migration of lung cancer cells (Jiang et al., 2018). Zhang et al. discovered that hsa\_circ\_0014130 (top 2) exhibited substantially overexpression in NSCLC tissues (Zhang S. et al., 2018). Zhu et al. indicated that hsa\_circ\_0016760 (top 3) accelerated the malignant growth of NSCLC by sponging miR-145-5p/FGF5 (Zhu et al., 2021).

Diabetes retinopathy is a microvascular complication caused by diabetes, which can be divided into proliferative diabetic retinopathy and non-proliferative diabetic retinopathy. As shown in **Table 3**, the predictive results contain seven experimentally verified associations among the top 10 ranked candidate circRNAs. For instance, hsa\_circRNA\_063981 (top 1), hsa\_circRNA\_404457 (top 2), and hsa\_circRNA\_100750 (top 3) are considerably elevated in the serum of T2DR patients compared to T2DM and control patients (Gu et al., 2017).

Prostate cancer refers to malignant tumors produced by the epithelial cells of the prostate under the action of a variety of

**TABLE 5** | Top 10 candidate circRNAs related to cholangiocarcinoma.

Rank	circRNA	Evidence (PMID)
1	hsa_circ_000438	unconfirmed
2	circHIPK3	31654054
3	ciRS-7	33390857
4	circR-284	unconfirmed
5	circDLGAP4	unconfirmed
6	circSMARCA5	31880360
7	hsa_circ_0008887	unconfirmed
8	hsa_circ_0006404	unconfirmed
9	hsa_circRNA_000585	34182814
10	hsa_circ_0000673	33221765

**TABLE 6** | Top 10 candidate circRNAs related to clear cell renal cell carcinoma.

Rank	circRNA	Evidence (PMID)
1	circHIPK3	32409849
2	circR-284	unconfirmed
3	circDLGAP4	unconfirmed
4	hsa_circ_0004383	unconfirmed
5	Cir-ITCH	unconfirmed
6	hsa_circRNA_003251	unconfirmed
7	circPVT1	33453148
8	hsa_circ_0001451	30271486
9	ciRS-7	32496306
10	circZFR	31571906

carcinogenic factors, which causes bone pain, pathological fractures, and paraplegia. Using the GATGCN, we successfully predict five of 10 top candidate circRNAs for prostate cancer (**Table 4**). The results in the literature indicate that circHIPK3 (top 1) expression is upregulated in prostate cancer cells and prostate cancer tissues (Liu et al., 2020). Kong et al. found that circFOXO3 (top 3) acted as a sponge for miR-29a-3p, exhibiting oncogenic activity in prostate cancer (Kong et al., 2020). Li et al. revealed that hsa\_circ\_0044516 (top 8) downregulation suppressed prostate cancer cell metastasis and growth (Li T. et al., 2020).

In order to further assess the capacity of GATGCN for detecting new diseases, two common diseases, that is, clear cell renal cell carcinoma and cholangiocarcinoma are chosen for case studies. Specifically, all known associations about clear cell renal cell carcinoma and cholangiocarcinoma are reset to unknown and all candidate circRNAs are prioritized according to corresponding prediction scores. Eventually, we select the top 10 scores to assess the performance of GATGCN for detecting new circRNAs and diseases.

Cholangiocarcinoma is a malignant tumor that originates from the extrahepatic bile duct. The result in **Table 5** shows that five associations are verified among the top 10 ranked candidate circRNAs. For example, Louis et al. demonstrated that the expression of circHIPK3 (top 2) was specifically elevated in cholangiocarcinoma cell lines (Louis et al., 2019). Chen et al. discovered that in cholangiocarcinoma, ciRS-7 (top 3) acts as an oncogene and promotes tumor development by competitively inhibiting miR-7. (Chen et al., 2021). Lu et al.

indicated that circSMARCA5 (top 6) expression was lower in ICC tumor tissues than surrounding tissues (Lu and Fang, 2020).

Clear cell renal cell carcinoma is derived from adenocarcinoma of renal tubular epithelial cells, which forms hemangioma thrombus or metastasizes to lymph nodes and other organs. As shown in **Table 6**, the predicted results contain five experimental verified associations among the top 10 ranked candidate circRNAs. For example, Li et al. discovered that overexpression of circHIPK3 (top 1) substantially reduced CCRCC cell invasion and migration *in vitro* (Li H. et al., 2020). Zheng et al. discovered that circPVT1 (top 7) promotes progression in CCRCC cells by regulating TBX15 expression and sponging miR-145-5p (Zheng et al., 2021). Wang et al. indicated that hsa\_circ\_0001451 (top 8) upregulation could promote CCRCC cell invasion and proliferation (Wang G. et al., 2018).

The results of the case studies show that GATGCN can efficiently detect the potential circRNA–disease relationships and provide clues for exploring the mechanism between human complex diseases and circRNAs.

## CONCLUSION

Cumulative evidence has proved that the development of powerful calculation methods is significant to infer the interactions between diseases and circRNAs. These calculation models address challenges of high cost and high time consumption in conventional biological experiments. In this study, an advanced calculation method called GATGCN is designed to discover potential circRNA–disease relationships *via* graph attention mechanism and graph convolutional network. First, multisource similarity data for circRNAs and diseases are fused by the centered kernel alignment model. Second, the graph attention network is deployed to learn the dense representation of nodes on the disease–disease similarity network and circRNA–circRNA similarity network. Third, the heterogenous network is constructed by connecting known circRNA–disease associations, feature matrix of diseases, and feature matrix of circRNAs. Finally, the graph convolutional network is applied to get prediction scores based on the constructed heterogenous network. To further confirm the advantage of GATGCN for detecting circRNA–disease interactions, we compare it with several state-of-the-art prediction models under five-fold cross-validation. The results indicate that GATGCN achieves significant performance among compared methods. Meanwhile, the case study substantiates the excellent capability of the GATGCN for detecting potential circRNA–disease relationships. In conclusion, GATGCN is a

powerful and promising approach for detecting circRNA–disease relationships.

Although we have integrated multisource biological information and utilized graph attention network and graph convolutional network to learn latent representation for diseases and circRNAs, there is still room to strengthen the predictive capability of the model. On the one hand, a large number of nonlinear features are extracted to detect circRNA–disease associations, which ignore the importance of linear features. We could further solve this problem by fusing nonlinear features and linear features to enhance the stability of our model. On the other hand, feature aggregation in excessive network layers could affect the expression of initial feature information. In the future, we can splice the representations of nodes in different layers as node features.

## 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 authors. The GATGCN dataset and code can be downloaded from <https://github.com/ghli16/GATGCN>.

## AUTHOR CONTRIBUTIONS

GL and JL conceived, designed, and managed the study. DW developed the GATGCN model and wrote the original manuscript. YZ revised the original draft. CL and QX discussed the GATGCN model and provided further research. All authors read and approved the final manuscript.

## FUNDING

This work has been supported by the National Natural Science Foundation of China (Grant Nos. 61862025, 61873089, 62002116, 11862006, and 92159102), Natural Science Foundation of Jiangxi Province of China (Grant Nos. 20212BAB202009, 20181BAB211016, and 20202BAB205011).

## ACKNOWLEDGMENTS

We would like to thank all authors of the cited references.

## REFERENCES

- Bian, C., Lei, X.-J., and Wu, F.-X. (2021). GATCDA: Predicting circRNA–Disease Associations Based on Graph Attention Network. *Cancers* 13 (11), 2595. doi:10.3390/cancers13112595
- Chen, X., Han, P., Zhou, T., Guo, X., Song, X., and Li, Y. (2016). circRNADb: a Comprehensive Database for Human Circular RNAs with Protein-Coding Annotations. *Sci. Rep.* 6 (1), 1–6. doi:10.1038/srep34985
- Chen, J., Cui, L., Yuan, J., Zhang, Y., and Sang, H. (2017). Circular RNA WDR77 Target FGF-2 to Regulate Vascular Smooth Muscle Cells Proliferation and Migration by Sponging miR-124. *Biochem. Biophys. Res. Commun.* 494 (1–2), 126–132. doi:10.1016/j.bbrc.2017.10.068
- Chen, M., Peng, Y., Li, A., Li, Z., Deng, Y., Liu, W., et al. (2018). A Novel Information Diffusion Method Based on Network Consistency for Identifying Disease Related MicRNAs. *RSC Adv.* 8 (64), 36675–36690. doi:10.1039/C8RA07519K
- Chen, J., Yang, J., Fei, X., Wang, X., and Wang, K. (2021). CircRNA ciRS-7: a Novel Oncogene in Multiple Cancers. *Int. J. Biol. Sci.* 17 (1), 379–389. doi:10.7150/ijbs.54292
- Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2016). “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs),” in *International*

- Conference on Learning Representations (ICLR), San Juan, Puerto Rico, May 2–4, 2016. Available at: <https://arxiv.org/abs/1511.07289>.
- Cristianini, N., Kandola, J., Elisseeff, A., and Shawe-Taylor, J. (2006). “On Kernel Target Alignment,” in *Innovations in Machine Learning* (Berlin, Heidelberg: Springer), 205–256. doi:10.1007/3-540-33486-6\_8
- Ding, Y., Tang, J., and Guo, F. (2019). Identification of Drug-Side Effect Association via Multiple Information Integration with Centered Kernel Alignment. *Neurocomputing* 325, 211–224. doi:10.1016/j.neucom.2018.10.028
- Fan, C., Lei, X., and Wu, F.-X. (2018a). Prediction of CircRNA-Disease Associations Using KATZ Model Based on Heterogeneous Networks. *Int. J. Biol. Sci.* 14 (14), 1950–1959. doi:10.7150/ijbs.28260
- Fan, C., Lei, X., Fang, Z., Jiang, Q., and Wu, F.-X. (2018b). CircR2Disease: a Manually Curated Database for Experimentally Supported Circular RNAs Associated with Various Diseases. *Database* 2018, bay044. doi:10.1093/database/bay044
- Ghosal, S., Das, S., Sen, R., Basak, P., and Chakrabarti, J. (2013). Circ2Traits: a Comprehensive Database for Circular RNA Potentially Associated with Disease and Traits. *Front. Genet.* 4, 283. doi:10.3389/fgene.2013.00283
- Glažar, P., Papavasileiou, P., and Rajewsky, N. (2014). circBase: a Database for Circular RNAs. *Rna* 20 (11), 1666–1670. doi:10.1261/rna.043687.113
- Gu, Y., Ke, G., Wang, L., Zhou, E., Zhu, K., and Wei, Y. (2017). Altered Expression Profile of Circular RNAs in the Serum of Patients with Diabetic Retinopathy Revealed by Microarray. *Ophthalmic Res.* 58 (3), 176–184. doi:10.1159/000479156
- Han, B., Chao, J., and Yao, H. (2018). Circular RNA and its Mechanisms in Disease: from the Bench to the Clinic. *Pharmacol. Ther.* 187, 31–44. doi:10.1016/j.pharmthera.2018.01.010
- Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA Circles Function as Efficient microRNA Sponges. *Nature* 495 (7441), 384–388. doi:10.1038/nature11993
- Huang, Y.-a., Hu, P., Chan, K. C. C., and You, Z.-H. (2020). Graph Convolution for Predicting Associations between miRNA and Drug Resistance. *Bioinformatics* 36 (3), 851–858. doi:10.1093/bioinformatics/btz621
- Jiang, M.-M., Mai, Z.-T., Wan, S.-Z., Chi, Y.-M., Zhang, X., Sun, B.-H., et al. (2018). Microarray Profiles Reveal that Circular RNA Hsa\_circ\_0007385 Functions as an Oncogene in Non-small Cell Lung Cancer Tumorigenesis. *J. Cancer Res. Clin. Oncol.* 144 (4), 667–674. doi:10.1007/s00432-017-2576-2
- Kipf, T. N., and Welling, M. (2016). *Semi-supervised Classification with Graph Convolutional Networks*. arXiv preprint arXiv:1609.02907. Available at: <https://arxiv.org/abs/1609.02907>.
- Kong, Z., Wan, X., Lu, Y., Zhang, Y., Huang, Y., Xu, Y., et al. (2020). Circular RNA circFOXO3 Promotes Prostate Cancer Progression through Sponging miR-29a-3p. *J. Cel Mol Med* 24 (1), 799–813. doi:10.1111/jcmm.14791
- Lei, X., and Bian, C. (2020). Integrating Random Walk with Restart and K-Nearest Neighbor to Identify Novel circRNA-Disease Association. *Sci. Rep.* 10 (1), 1–9. doi:10.1038/s41598-020-59040-0
- Lei, X., Fang, Z., Chen, L., and Wu, F.-X. (2018). PWCDA: Path Weighted Method for Predicting circRNA-Disease Associations. *Ijms* 19 (11), 3410. doi:10.3390/ijms19113410
- Lei, X., Mudiyansele, T. B., Zhang, Y., Bian, C., Lan, W., Yu, N., et al. (2021). A Comprehensive Survey on Computational Methods of Non-coding RNA and Disease Association Prediction. *Brief. Bioinform.* 22 (4), bbaa350. doi:10.1093/bib/bbaa350
- Li, Y., Qiu, C., Tu, J., Geng, B., Yang, J., Jiang, T., et al. (2014). HMDD v2.0: a Database for Experimentally Supported Human microRNA and Disease Associations. *Nucl. Acids Res.* 42 (D1), D1070–D1074. doi:10.1093/nar/gkt1023
- Li, Y., Zheng, Q., Bao, C., Li, S., Guo, W., Zhao, J., et al. (2015). Circular RNA Is Enriched and Stable in Exosomes: a Promising Biomarker for Cancer Diagnosis. *Cel. Res.* 25, 981–984. doi:10.1038/cr.2015.82
- Li, G., Yue, Y., Liang, C., Xiao, Q., Ding, P., and Luo, J. (2019). NCPCDA: Network Consistency Projection for circRNA-Disease Association Prediction. *RSC Adv.* 9 (57), 33222–33228. doi:10.1039/C9RA06133A
- Li, G., Luo, J., Wang, D., Liang, C., Xiao, Q., Ding, P., et al. (2020). Potential circRNA-Disease Association Prediction Using DeepWalk and Network Consistency Projection. *J. Biomed. Inform.* 112, 103624. doi:10.1016/j.jbi.2020.103624
- Li, H., Heng, B., Ouyang, P., Xie, X., Zhang, T., Chen, G., et al. (2020). Comprehensive Profiling of circRNAs and the Tumor Suppressor Function of circHIPK3 in clear Cell Renal Carcinoma. *J. Mol. Hist.* 51 (3), 317–327. doi:10.1007/s10735-020-09882-9
- Li, M., Liu, M., Bin, Y., and Xia, J. (2020). Prediction of circRNA-Disease Associations Based on Inductive Matrix Completion. *BMC Med. Genomics* 13 (5), 1–13. doi:10.1186/s12920-020-0679-0
- Li, T., Sun, X., and Chen, L. (2020). Exosome Circ\_0044516 Promotes Prostate Cancer Cell Proliferation and Metastasis as a Potential Biomarker. *J. Cel. Biochem.* 121 (3), 2118–2126. doi:10.1002/jcb.28239
- Liu, F., Fan, Y., Ou, L., Li, T., Fan, J., Duan, L., et al. (2020). CircHIPK3 Facilitates the G2/M Transition in Prostate Cancer Cells by Sponging miR-338-3p. *Ott* 13, 4545–4558. doi:10.2147/OTT.S242482
- Lihong, P., Wang, C., Tian, X., Zhou, L., and Li, K. (2021). Finding lncRNA-Protein Interactions Based on Deep Learning with Dual-Net Neural Architecture. *Ieee/ acm Trans. Comput. Biol. Bioinf.* 14 (8), 1. doi:10.1109/TCBB.2021.3116232
- Louis, C., Desoteux, M., and Coulouarn, C. (2019). Exosomal circRNAs: New Players in the Field of Cholangiocarcinoma. *Clin. Sci. (Lond).* 133 (21), 2239–2244. doi:10.1042/CS20190940
- Lu, Q., and Fang, T. (2020). Circular RNA SMARCA5 Correlates with Favorable Clinical Tumor Features and Prognosis, and Increases Chemotherapy Sensitivity in Intrahepatic Cholangiocarcinoma. *J. Clin. Lab. Anal.* 34 (4), e23138. doi:10.1002/jcla.23138
- Lu, C., Zeng, M., Zhang, F., Wu, F.-X., Li, M., and Wang, J. (2021). Deep Matrix Factorization Improves Prediction of Human circRNA-Disease Associations. *IEEE J. Biomed. Health Inform.* 25 (3), 891–899. doi:10.1109/JBHI.2020.2999638
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs Are a Large Class of Animal RNAs with Regulatory Potency. *Nature* 495, 333–338. doi:10.1038/nature11928
- Meng, S., Zhou, H., Feng, Z., Xu, Z., Tang, Y., Li, P., et al. (2017). CircRNA: Functions and Properties of a Novel Potential Biomarker for Cancer. *Mol. Cancer* 16 (1), 1–8. doi:10.1186/s12943-017-0663-2
- Peng, L., Chen, Y., Ma, N., and Chen, X. (2017). NARRMDA: Negative-Aware and Rating-Based Recommendation Algorithm for miRNA-Disease Association Prediction. *Mol. Biosyst.* 13 (12), 2650–2659. doi:10.1039/c7mb00499k
- Peng, L.-H., Sun, C.-N., Guan, N.-N., Li, J.-Q., and Chen, X. (2018). HNMDA: Heterogeneous Network-Based miRNA-Disease Association Prediction. *Mol. Genet. Genomics* 293 (4), 983–995. doi:10.1007/s00438-018-1438-1
- Peng, L., Shen, L., Liao, L., Liu, G., and Zhou, L. (2020a). RNMFMDA: A Microbe-Disease Association Identification Method Based on Reliable Negative Sample Selection and Logistic Matrix Factorization with Neighborhood Regularization. *Front. Microbiol.* 11, 592430. doi:10.3389/fmicb.2020.592430
- Peng, L.-H., Zhou, L.-Q., Chen, X., and Piao, X. (2020b). A Computational Study of Potential miRNA-Disease Association Inference Based on Ensemble Learning and Kernel ridge Regression. *Front. Bioeng. Biotechnol.* 8, 40. doi:10.3389/fbioe.2020.00040
- Rong, D., Sun, H., Li, Z., Liu, S., Dong, C., Fu, K., et al. (2017). An Emerging Function of circRNA-miRNAs-mRNA axis in Human Diseases. *Oncotarget* 8 (42), 73271–73281. doi:10.18632/oncotarget.19154
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). *Graph Attention Networks*. arXiv preprint arXiv:1710.10903. Available at: <https://arxiv.org/abs/1710.10903>.
- Vo, J. N., Cieslik, M., Zhang, Y., Shukla, S., Xiao, L., Zhang, Y., et al. (2019). The Landscape of Circular RNA in Cancer. *Cell* 176 (4), 869–881. doi:10.1016/j.cell.2018.12.021
- Wang, D., Wang, J., Lu, M., Song, F., and Cui, Q. (2010). Inferring the Human microRNA Functional Similarity and Functional Network Based on microRNA-Associated Diseases. *Bioinformatics* 26 (13), 1644–1650. doi:10.1093/bioinformatics/btq241
- Wang, L., You, Z.-H., Huang, Y.-A., Huang, D.-S., and Chan, K. C. C. (2020a). An Efficient Approach Based on Multi-Sources Information to Predict circRNA-Disease Associations Using Deep Convolutional Neural Network. *Bioinformatics* 36 (13), 4038–4046. doi:10.1093/bioinformatics/btz825
- Wang, L., You, Z.-H., Li, Y.-M., Zheng, K., and Huang, Y.-A. (2020b). GCNCDA: A New Method for Predicting circRNA-Disease Associations Based on Graph Convolutional Network Algorithm. *Plos Comput. Biol.* 16 (5), e1007568. doi:10.1371/journal.pcbi.1007568
- Wang, H., Tang, J., Ding, Y., and Guo, F. (2021). Exploring Associations of Non-coding Rnas in Human Diseases via Three-Matrix Factorization with

- Hypergraph-Regular Terms on center Kernel Alignment. *Brief. Bioinform.* 22, bbaa409. doi:10.1093/bib/bbaa409
- Wang, G., Xue, W., Jian, W., Liu, P., Wang, Z., Wang, C., et al. (2018). The Effect of Hsa\_circ\_0001451 in clear Cell Renal Cell Carcinoma Cells and its Relationship with Clinicopathological Features. *J. Cancer* 9 (18), 3269–3277. doi:10.7150/jca.25902
- Wang, L., Shen, J., and Jiang, Y. (2018). Circ\_0027599/PHDLA1 Suppresses Gastric Cancer Progression by Sponging miR-101-3p.1. *Cel. Biosci.* 8, 58. doi:10.1186/s13578-018-0252-0
- Wang, P., Zhu, W., Liao, B., Cai, L., Peng, L., Yang, J., et al. (2018). Predicting Influenza Antigenicity by Matrix Completion with Antigen and Antiserum Similarity. *Front. Microbiol.* 9, 2500. doi:10.3389/fmicb.2018.02500
- Wei, H., and Liu, B. (2020). iCircDA-MF: Identification of circRNA-Disease Associations Based on Matrix Factorization. *Brief. Bioinformatics* 21 (4), 1356–1367. doi:10.1093/bib/bbz057
- Xiao, Q., Dai, J., and Luo, J. (2022). A Survey of Circular RNAs in Complex Diseases: Databases, Tools and Computational Methods. *Brief. Bioinformatics* 23(1), bbab444. doi:10.1093/bib/bbab444
- Xiao, Q., Fu, Y., Yang, Y., Dai, J., and Luo, J. (2021). NSL2CD: Identifying Potential circRNA-Disease Associations Based on Network Embedding and Subspace Learning. *Brief. Bioinform.* 22, bbab177. doi:10.1093/bib/bbab177
- Yan, C., Wang, J., and Wu, F.-X. (2018). DWNN-RLS: Regularized Least Squares Method for Predicting circRNA-Disease Associations. *BMC Bioinformatics* 19 (19), 73–81. doi:10.1186/s12859-018-2522-6
- Yao, D., Zhang, L., Zheng, M., Sun, X., Lu, Y., and Liu, P. (2018). Circ2Disease: a Manually Curated Database of Experimentally Validated circRNAs in Human Disease. *Sci. Rep.* 8 (1), 1–6. doi:10.1038/s41598-018-29360-3
- Zeng, X., Zhong, Y., Lin, W., and Zou, Q. (2020). Predicting Disease-Associated Circular RNAs Using Deep Forests Combined with Positive-Unlabeled Learning Methods. *Brief. Bioinformatics* 21 (4), 1425–1436. doi:10.1093/bib/bbz080
- Zhang, Y., Sun, L., Xuan, L., Pan, Z., Li, K., Liu, S., et al. (2016). Reciprocal Changes of Circulating Long Non-Coding RNAs ZFAS1 and CDR1AS Predict Acute Myocardial Infarction. *Sci. Rep.* 6, 22384. doi:10.1038/srep22384
- Zhang, W., Yu, C., Wang, X., and Liu, F. (2019). Predicting CircRNA-Disease Associations through Linear Neighborhood Label Propagation Method. *IEEE Access* 7, 83474–83483. doi:10.1109/ACCESS.2019.2920942
- Zhang, Y., Chen, M., Huang, L., Xie, X., Li, X., Jin, H., et al. (2021a). Fusion of KATZ Measure and Space Projection to Fast Probe Potential lncRNA-Disease Associations in Bipartite Graphs. *PLoS ONE* 16 (11), e0260329. doi:10.1371/journal.pone.0260329
- Zhang, S., Zeng, X., Ding, T., Guo, L., Li, Y., Ou, S., et al. (2018). Microarray Profile of Circular RNAs Identifies Hsa\_circ\_0014130 as a New Circular RNA Biomarker in Non-small Cell Lung Cancer. *Sci. Rep.* 8 (1), 2878. doi:10.1038/s41598-018-21300-5
- Zhang, Y., Gao, T., Li, X., Wen, C.-C., Yan, X.-T., Peng, C., et al. (2021b). Circ\_0005075 Targeting miR-151a-3p Promotes Neuropathic Pain in CCI Rats via Inducing NOTCH2 Expression. *Gene* 767, 145079. doi:10.1016/j.gene.2020.145079
- Zhang, Z., Yang, T., and Xiao, J. (2018). Circular RNAs: Promising Biomarkers for Human Diseases. *EBioMedicine* 34, 267–274. doi:10.1016/j.ebiom.2018.07.036
- Zhao, Z., Wang, K., Wu, F., Wang, W., Zhang, K., Hu, H., et al. (2018). circRNA Disease: a Manually Curated Database of Experimentally Supported circRNA-Disease Associations. *Cel. Death Dis.* 9 (5), 1–2. doi:10.1038/s41419-018-0503-3
- Zhao, Q., Yang, Y., Ren, G., Ge, E., and Fan, C. (2019). Integrating Bipartite Network Projection and KATZ Measure to Identify Novel CircRNA-Disease Associations. *IEEE Trans.on Nanobiosci.* 18 (4), 578–584. doi:10.1109/TNB.2019.2922214
- Zheng, Z., Chen, Z., Zhong, Q., Zhu, D., Xie, Y., Shangguan, W., et al. (2021). CircPVT1 Promotes Progression in clear Cell Renal Cell Carcinoma by Sponging miR-145-5p and Regulating TBX15 Expression. *Cancer Sci.* 112 (4), 1443–1456. doi:10.1111/cas.14814
- Zhou, L., Wang, Z., Tian, X., and Peng, L. (2021). LPI-DeepGBDT: A Multiple-Layer Deep Framework Based on Gradient Boosting Decision Trees for lncRNA-Protein Interaction Identification. *BMC Bioinformatics* 22 (479), 1–24. doi:10.1186/s12859-021-04399-8
- Zhu, L.-P., He, Y.-J., Hou, J.-C., Chen, X., Zhou, S.-Y., Yang, S.-J., et al. (2017). The Role of circRNAs in Cancers. *Biosci. Rep.* 37 (5), BSR20170750. doi:10.1042/BSR20170750
- Zhu, Z., Wu, Q., Zhang, M., Tong, J., Zhong, B., and Yuan, K. (2021). Hsa\_circ\_0016760 Exacerbates the Malignant Development of Non-Small Cell Lung Cancer by Sponging miR-145-5p/FGF5. *Oncol. Rep.* 45 (2), 501–512. doi:10.3892/or.2020.7899

**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.

Copyright © 2022 Li, Wang, Zhang, Liang, Xiao and Luo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.