International Journal of Biology and Life Sciences
ISSN: 2957-9511 | Vol. 12, No. 3, 2025

Prediction of piRNA and Disease Association based on Graph Neural Network

Xiulian Fang

School of College of Electrical Engineering, Southwest Minzu University, Chengdu, Sichuan, 610041, China

Abstract: Piwi interacting RNA (piRNA) is a type of small non coding RNA with a length of 24-32 nucleotides, mainly expressed in germ cells. Its abnormal expression is closely related to various diseases such as cancer and neurodegenerative diseases. Although biological experiments are the gold standard for identifying the association between piRNA and disease (PDA), their high cost and long cycle limit research progress. Therefore, computational models have become an important tool for assisting in predicting PDA. However, existing computational methods generally suffer from issues such as insufficient feature extraction and imbalanced data. This article proposes a prediction model based on the fusion of graph convolutional network and attention mechanism - RandGCN. This model combines piRNA sequence embedding, heterogeneous graph construction based on random walks, multi-layer graph convolution feature extraction, and multi head attention mechanism with gating units, effectively improving the accuracy and robustness of PDA prediction. The experimental results on the MNDR dataset show that RandGCN performs well in both AUC and AUPR values, demonstrating excellent predictive performance and potential applications.

Keywords: piRNA Disease Association; Graph Neural Network; Transformer.

1. Introduction

Piwi-interacting RNA (piRNA) is a class of small non-coding RNA, 24 to 32 nucleotides in length, primarily expressed in germ cells [1,2]. In 2006, Girard et al. identified a group of previously unclassified small atypical RNAs that bind to the MIWI mouse Piwi protein and are highly enriched in testicular tissues, thus named piwi-interacting RNAs (piRNAs) [3-5].

Studies have confirmed that abnormal expression of piRNA is closely associated with the development and progression of various human diseases, including cancer, neurodegenerative disorders, and age-related conditions [6]. PiRNAs are not only deeply involved in the regulatory networks of human diseases but are also increasingly recognized as key initiators and regulators in tumorigenesis and tumor progression [7]. For example, piRNA-14633 is significantly up-regulated in cervical cancer tissues and cells, where it strongly promotes tumor growth [8]. In contrast, piRNA-6426 expression is reduced in patients with heart failure, and its functional restoration can alleviate hypoxia-induced damage in cardiomyocytes and inhibit the progression of heart failure [9]. Thus, accurate identification of disease-related piRNAs is of central importance for developing novel diagnostic biomarkers and therapeutic strategies.

Current piRNA research focuses mainly on its molecular mechanisms and biological functions in various diseases. Although biological experiments—such as gene knockout [10] and RNA interference [11]—remain effective means for identifying piRNA–disease associations, these methods often involve complex, costly, and time-consuming procedures [12]. Moreover, with the continuous growth in the number of known piRNAs and diseases, relying solely on experimental approaches for systematic association mining faces limitations in both time and resources. To address these challenges, researchers have developed various computational models to predict potential piRNA–disease associations (PDA), providing efficient and valuable tools to assist and guide subsequent biological experiments [13]. Identifying disease-related piRNAs and revealing their roles in pathogenesis is currently a core topic in biomedical and clinical research. Drawing on the association patterns between other functional RNAs and diseases, we can reasonably infer the potential relationship between unexplored non coding RNAs and diseases [14-15]. This computational process typically involves two key steps: first, constructing a biological network with RNA and disease as nodes and known associations as edges; Secondly, utilizing the existing information in the network, new associations are predicted through machine learning methods. Although some piRNA disease associations have been experimentally validated, research in this field, like predicting associations between miRNA [16-17], lncRNA [18-19], and circRNA [20], is still generally limited by long experimental cycles and high resource consumption. To address this challenge, an increasing number of efficient computing algorithms have been proposed to accelerate the construction of piRNA disease association networks.

Although predictive models based on computational methods have certain advantages, there are still significant shortcomings in feature design. Firstly, most models overly rely on known correlated samples to construct feature networks, failing to fully integrate the sequence information of piRNA itself, resulting in inaccurate characterization of piRNA and disease features, thereby limiting the predictive performance of the models. Secondly, the common issue of class imbalance in the dataset, where there is a significant disparity in the number of positive and negative samples and reliable negative samples are difficult to obtain, also affects the robustness of the model. In addition, there may be true positive associations hidden in unlabeled samples that have not been discovered due to cost or technical limitations. These factors collectively constrain the accuracy and comprehensiveness of the model in predicting unknown associations.

To address the above issues, this paper proposes a piRNA disease association prediction model based on graph convolution and attention encoding fusion - RandGCN. The main contributions of this method are as follows:

(1) Combining word2vec and TextCNN for embedding representation of piRNA sequences, and using random walk algorithm to construct piRNA disease heterogeneity graph, enhancing the semantic discrimination ability of node features;

(2) Introducing graph convolutional networks as feature extraction modules, by preserving the features of the middle layers of the graph network, multi-level feature representations are constructed to enhance the integrity and expressive power of information;

(3) Introducing gating units in the attention encoding layer to regulate the information flow between attention output and raw input through learnable gating vectors, enhancing the ability to select key information while reducing computational complexity.

2. Method

The flowchart of the RandGCN model is shown in Figure 1, which can be divided into three steps: (1) piRNA sequence processing: splitting the RNA sequence and obtaining feature vectors using word2vec; (2) Construct a similarity matrix between piRNA and diseases using random walk algorithm, and extract feature representations of piRNA and diseases using graph convolutional network; (3) Using attention mechanism to predict the association score between piRNA and diseases.

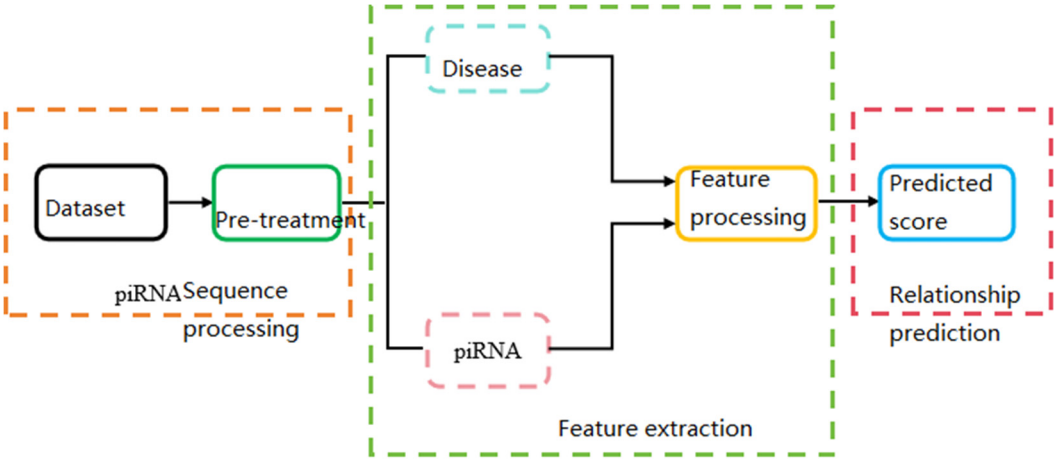

Figure 1. RandGCN flowchart

2.1. Construction of piRNA Disease Association Matrix

This article uses a piRNA disease association database to construct a PDA association matrix, represented as follows:

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} \quad (1)$$

Among them, m and n represent the number of piRNAs and diseases, respectively. If there is an association between the i-th piRNA and the j-th disease, then$a_{i,j} = 1$;Conversely, it is 0.

2.2. Construction of piRNA Sequence Similarity Matrix

To quantify the sequence similarity between piRNA molecules, this study first extracted their sequence information from a designated database. Subsequently, the Smith Waterman algorithm was used to pairwise align all piRNAs, thereby constructing an m-dimensional piRNA sequence similarity matrix, denoted as $PS_{seq} \in R^{m \times m}$, where dimension m corresponds to the number of piRNAs studied. To eliminate the influence of dimensionality, the original similarity score is normalized, and the final calculation formula is defined as follows:

$$S_p^{seq}(p_i, p_j) = \frac{SW(p_i, p_j)}{\sqrt{SW(p_i, p_i) \times SW(p_j, p_j)}} \quad (2)$$

Where $SW(p_i, p_j) \in R^{m \times m}$ represents the sequence similarity between the i-th and j-th piRNAs.

2.3. Construction of Gaussian Interaction Nuclear Similarity in piRNA

In ncRNA disease association prediction, a fundamental assumption is that functionally similar molecules (such as piRNA) tend to be associated with similar diseases. Gaussian interaction spectral kernel similarity is an effective measurement method based on this assumption, which evaluates the correlation between two nodes (such as piRNA) by calculating their interaction spectral similarity in a known association network. Specifically, the GIP nuclear similarity between the i-th and j-th piRNAs is defined as:

$$S_p^{GIP}(p_i, p_j) = \exp\left(-\lambda_p \left\| A(p_i,) - A(p_j,) \right\|^2\right) \quad (3)$$

Among them, $A(p_i,)$ and $A(p_j,)$are the i-th and j-th row vectors of the correlation matrix A, and λ_p is the kernel width coefficient, defined by the following formula:

$$\lambda_p = \frac{1}{\frac{1}{N_p}\sum_{k=1}^{N_p} \|A(p_k,)\|^2} \quad (4)$$

2.4. piRNA Similarity Matrix

The similarity between piRNAs is the average of sequence similarity and GIP kernel similarity, expressed as follows:

$$S_p = \frac{S_p^{seq} + S_p^{GIP}}{2} \quad (5)$$

2.5. Construction of Disease Semantic Similarity Matrix

To quantify the topological relationships between diseases, we introduced a semantic similarity measure based on the MeSH database. The core idea of this method is that if two

diseases share more nodes in their DAG (Directed Acyclic Graph) structure, they are semantically more similar. In this way, the correlation between diseases can be effectively captured. Specifically, the semantic similarity between the i-th and j-th diseases is calculated using the following formula:

$$S_d^{sem}(d_i, d_j) = \frac{\sum_{t \in T_i \cap T_j}\left(S_{d_i}(t)+S_{d_j}(t)\right)}{\sum_{t \in T_i} S_{d_i}(t)+\sum_{t \in T_j} S_{d_j}(t)} \quad (6)$$

Where T_i is the set of all diseases in the DAG that includes the i-th disease, and represents the semantic impact of the disease on the i-th disease. Its calculation is as follows:

$$\begin{cases} S_{d_k}(t) = max\{\theta \cdot S_{d_k}(t') | t' \in children\ of(t)\} & if\ d_k \neq d_j \\ S_{d_k}(t) = 1 & otherwise \end{cases} \quad (7)$$

Where θ is the attenuation parameter, set to 0.5. The less intersection between the parents of two diseases, the lower the semantic similarity.

2.6. Construction of Gaussian Interaction Kernel Similarity Matrix for Diseases

Similar to piRNA, the GIP nuclear similarity between diseases is as follows:

$$S_d^{GIP}(d_i, d_j) = \exp\left(-\lambda_d \left\| A(, d_i) - A(, d_j) \right\|^2\right) \quad (8)$$

2.7. Disease Similarity Matrix

Obtain disease similarity matrix through semantic similarity and GIP kernel similarity:

$$S_d = \frac{S_d^{sem}+S_d^{GIP}}{2} \quad (9)$$

2.8. Construction of piRNA Disease Heterogeneous Network

The piRNA disease heterogeneous network consists of the following three parts:

$$A = \begin{bmatrix} S_p & A \\ A^T & S_d \end{bmatrix} \quad (10)$$

In this model, we define three core matrices: S_p, S_d, and A.$S_p \in R^{m \times m}$ is the piRNA similarity matrix, calculated by formula (5); $S_d \in R^{m \times n}$ is the disease similarity matrix, calculated by formula (9); $A \in R^{m \times n}$ is the piRNA disease adjacency matrix. Among them, the dimensions m and n of the matrix correspond to the total number of piRNAs and diseases, respectively.

3. Model

3.1. piRNA Sequence Processing

To capture the similarity of piRNAs at the sequence level, we introduced an embedding strategy based on natural language processing principles. Firstly, the piRNA sequence is decomposed into a series of non-overlapping 3-mer subsequences to extract its basic compositional information. Next, we treat these 3-mers as a corpus and apply the word2vec model for unsupervised learning to generate dense vector representations for each 3-mer. In this process, we specifically chose the skip gram architecture because it can generate more accurate representations than other models when dealing with infrequently occurring "words" (i.e. low-frequency 3-mers).

Random walk is a fundamental mathematical statistical model used to characterize a path composed of a series of random steps. It can effectively simulate various irregular patterns, such as the unstable walking trajectory of intoxicated individuals. In concept, random walk is closely related to Brownian motion and can be seen as an idealized mathematical representation of Brownian motion in discrete time..

A simple example: The most typical one-dimensional random walk process is as follows. Imagine a particle located at the origin of the number axis. At each moment t, it starts from the current position $x(t)$ and chooses to move forward one step (to reach $x(t+1)$) or backward one step (to reach $x(t)-1$) with the same probability (1/2). Repeatedly, the particle passes through a series of position points $x(0)$, $x(1)$, $x(2)$ The set of... defines a one-dimensional random walk process.

3.2. Graph Convolutional Layer

Graph convolutional network is a neural network model that can effectively extract features from graph data. The core mechanism is to update the representation of the central node by aggregating information from neighboring nodes. As shown in Figure 2, for any node in the graph, GCN first collects its own features and the features of all its neighbors. Subsequently, a new node embedding is generated by normalizing and aggregating these features. This new embedding integrates the topology and attribute information of the node itself and its neighborhood, thereby obtaining a more discriminative depth representation than the original features.

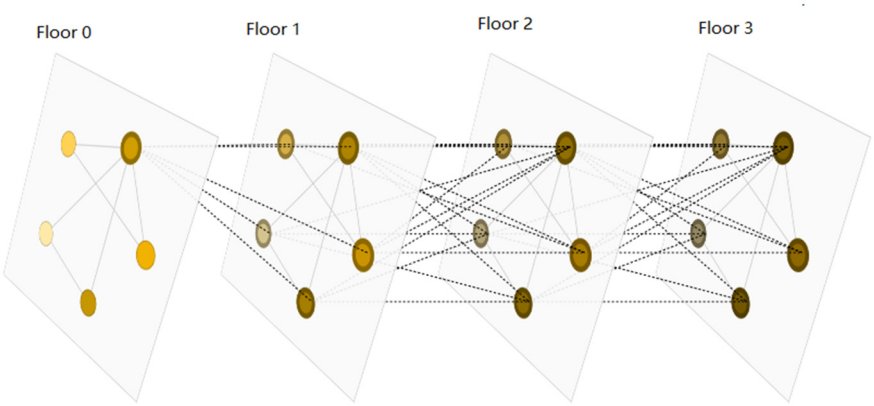

Figure 2. GCN flowchart

Assuming the input feature is $x^{(l)}$, the first layer feature $x^{(l+1)}$ processed by GCN is:

$$x^{(l+1)} = \sigma(D^{-\frac{1}{2}}\hat{A}D^{-\frac{1}{2}}H^l W^l) \quad (11)$$

In this model, the input of GCN is the initial embedding matrix H^0, which is composed of embedding vectors of piRNA and disease concatenated together. Among them, piRNA embedding originates from sequence features, while disease embedding directly uses the similarity matrix S_d. At each layer l of GCN, node embeddings H^l are updated by aggregating neighborhood information, and their matrix structure remains unchanged: the first m rows are piRNA embeddings, and the last n rows are disease embeddings. In the model, σ, $\hat{A}$, D, and W^l represent the ReLU activation function, adjacency matrix, degree matrix, and weight matrix, respectively.

3.3. Attention Mechanism Layer with Gate Control Unit

In order to further model the intricate relationship between piRNA and diseases, we designed an enhancement module that integrates multi head attention and gated linear units (GLU). The core advantage of this module lies in its dual screening mechanism: firstly, the multi head attention mechanism can dynamically evaluate the importance between nodes and assign higher attention weights to key piRNA and disease nodes; Secondly, the GLU gating mechanism further filters information on this basis, acting like an intelligent gate that amplifies effective signals and suppresses noise interference. Through this "focus first, filter later" strategy, the model is able to accurately identify and utilize the most valuable features in PDA prediction tasks, significantly improving its ability to represent complex associations.

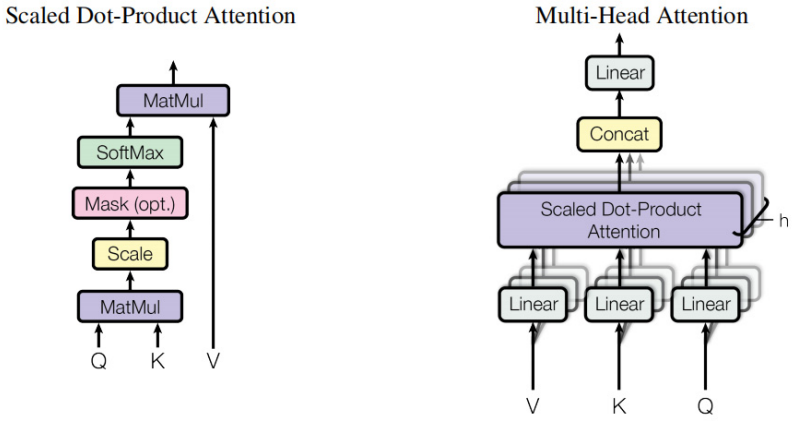

Figure 3. (left) dot product attention; (Right) Multi head Attention

To dynamically evaluate the importance between piRNA and disease nodes, we introduced an attention mechanism, whose structure is shown in Figure 3. The core of this mechanism is to map queries Q, keys K, and values V. In our setting, queries and keys are used to calculate attention scores to quantify the strength of associations between nodes; And the value vector carries the actual features of the node (piRNA or disease). We further adopt multi head attention, which parallel processes multiple representation subspaces to enable the model to capture complex dependency relationships from different perspectives. The specific dot product attention calculation method is as follows:

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (12)$$

Where

$$\begin{aligned} Q' &= QW^Q \\ K' &= KW^K \\ V' &= VW^V \end{aligned} \quad (13)$$

W^Q,W^K and $W^V$are trainable parameters, and d_k is the dimension of the feature vector.

The core advantage of the multi head attention mechanism is that it allows the model to simultaneously learn the dependency relationships between nodes in multiple different representation subspaces. This design enables the model to examine the association between piRNA and disease from multiple perspectives, thereby capturing its complex interaction patterns more comprehensively and deeply:

$$\begin{aligned} X = MultiHead(Q,K,V) &= Concat(head_1,\cdots,head_h)W^O \\ head_i &= Attention(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \quad (14)$$

In this model, h represents the number of attention heads, which we set to 2 by default. The output of the multi head attention mechanism $MultiHead(Q,K,V)$ is an embedding matrix, where each row corresponds to an enhanced feature representation of a node (piRNA or disease). This representation is formed by concatenating and projecting information learned by multiple attention heads in different subspace.

To predict the association between piRNA and disease, we first extracted the final embedding vector of the target piRNA and disease from $MultiHead(Q,K,V)$. Then, we quantify the similarity between these two vectors by calculating their inner product, which is the predicted correlation score.

In standard attention modules, there is usually a residual connection (adding input and output) and layer normalization operation. These steps are used to optimize the training process and feature fusion, but the final correlation score calculation is based on the inner product of the normalized output vector.

The results of traditional attention may lead to the introduction of irrelevant feature noise during fusion. Therefore, this study introduces a gating unit in the output layer of the attention mechanism. The gating unit is an improved variant of MLP enhanced by gating, and its calculation form is:

$$G = \sigma(XW_g + b_g) \quad (15)$$

Among them, X is the original representation of multi head attention that contains interactive information from multiple perspectives, W_g and $b_g$are learnable parameters, and $\sigma$is the Sigmoid function that maps the result to [0,1],

representing the "degree of gate opening and closing". When G is closer to 1, it indicates that the output is biased towards new information; The closer G is to 0, the more biased the output is towards the original input.

The final output is:

$$Y = G \odot X' + (1 - G) \odot X \tag{16}$$

Among them, $\odot$ represents element wise multiplication, and G controls whether each dimension selects more new information X' from attention or retains more original input X.

4. Results and Discussion

4.1. Datasets

To construct the experimental dataset, we extracted raw data from the comprehensive database MNDR [21] (RNADisease v4.0). The database has a large scale, containing over 3.42 million RNA disease associations, covering 18 RNA types, 117 species, and 4090 diseases. After data cleaning (including removing duplicates and fuzzy associations), we ultimately constructed a subset focused on piRNA and disease. This subset contains 9616 known PDAs, consisting of 8205 piRNAs and 15 diseases. It is worth noting that these positive sample associations only account for 7.81% of the entire association space, highlighting the challenge of mining new discoveries from massive unknown associations.

4.2. Evaluation Metric

To evaluate the performance of the model, we constructed the piRNA disease association prediction task as a binary classification problem. Among them, known piRNA disease associations are considered positive samples, while other unknown associations are considered negative samples. The output of the model is a correlation score, and by changing the judgment threshold of this score, a series of classification results can be obtained. Based on this, a confusion matrix can be constructed to calculate the values of TP, FP, FN, and TN. Based on the confusion matrix, we calculate two core metrics: true case rate and false positive case rate. The ROC curve is formed by plotting the (FPR, TPR) points at different thresholds, with FPR on the horizontal axis and TPR on the vertical axis. The definitions of TPR and FPR are as follows:

$$TPR = \frac{TP}{TP+FN} \tag{17}$$

$$FPR = \frac{FP}{FP+TN} \tag{18}$$

To evaluate the performance of the model under imbalanced positive and negative samples, we use the precision recall curve as a supplementary evaluation metric. By continuously adjusting the classification threshold, the model generates a series of prediction results for each sample, and calculates the corresponding accuracy and recall based on this. The PR curve depicts the trajectory formed by these points with recall as the horizontal axis and precision as the vertical axis. The definitions of recall and precision are as follows:

$$Recall = \frac{TP}{TP+FN} \tag{19}$$

$$Precision = \frac{TP}{FP+TP} \tag{20}$$

The F1 score is the harmonic mean of precision and recall, calculated as follows:

$$F1 = \frac{2(Precision \times Recall)}{Precision+Recall} \tag{21}$$

In order to robustly evaluate the performance of the model and address the bias caused by the large difference in the number of positive and negative samples in the raw data, we designed an experimental process based on 5-fold cross validation. Firstly, we construct a balanced dataset through random downsampling: for each positive sample, we randomly select a negative sample from a large pool of unlabeled samples to ensure an equal number of positive and negative samples. Then, randomly divide this balanced dataset into 5 equal parts. In each iteration of cross validation, the model is trained on 4 samples and tested on the remaining 1 sample. This process is repeated 5 times to ensure that each data set has been used as a test set, and the final evaluation result is the average performance of each round.

4.3. Compare with Other Methods

We compared the model with six other methods for 5-fold cross validation on the dataset: ETGPDA, iPiDi PU, iPiDA GCN, iPiDA SWGCN, iPiDA GBNN, and piRDA. The experimental results were obtained by taking the average and variance of the last five iterations of all five cross validations. Table 1 shows the comparison results on the MNDR dataset. It can be seen that RandGCN achieved an AUC of 0.929 and an AUPR value of 0.507, both higher than other methods.

Table 1. Comparison Results with Other Models

	Ranking index	AUC	AUPR
ETGPDA	0.116	0.916	0.417
iPiDi-PUL(DT)	0.444	0.569	0.117
iPiDi-PUL(SVM)	0.292	0.725	0.133
iPiDi-PUL(RF)	0.238	0.784	0.165
iPiDA-GCN	0.145	0.885	0.427
iPiDA-SWGCN	0.113	0.920	0.465
iPiDA-GBNN	0.153	0.879	0.461
RandGCN	0.104	**0.929**	**0.507**

4.4. The Influence of Random Probability

The probability parameter of random walk directly affects the contextual information learned by the model by controlling the generation method of node sequence. Adjusting this parameter essentially changes the intrinsic distribution of the training data, which further shapes the final characteristics of node embedding. Our ablation experiment (see Table 2) validated this point: when the probability parameter is 0.7, the model can learn the node representation that is most conducive to distinguishing positive and negative samples, thus achieving the best performance with an AUC of 0.929 and an AUPR of 0.507. This fully demonstrates the efficiency and accuracy of the RandGCN model in handling piRNA disease association binary classification problems.

Table 2. Results of different random probabilities

Random probability	Ranking index	AUC	AUPR
0.4	0.132	0.898	0.599
0.5	0.136	0.894	0.359
0.6	0.106	0.926	0.499
0.7	0.104	**0.929**	**0.507**
0.8	0.161	0.866	0.451

5. Conclusion

This study addresses the challenges of insufficient feature extraction and imbalanced data in piRNA disease association prediction tasks, and designs and implements a novel computational model called RandGCN. The core innovation of this model lies in the deep fusion of graph convolutional networks and gate enhanced attention mechanisms.

Specifically, the model constructs a heterogeneous graph network by integrating piRNA sequence information, utilizing multi-layer convolution to preserve rich intermediate features, and dynamically focusing on key information through attention mechanisms, thereby significantly enhancing its ability to represent complex associations. A comprehensive experimental evaluation confirms that RandGCN significantly outperforms current mainstream methods in key indicators such as AUC and AUPR. Further ablation studies have also revealed that random walk probability is an important hyperparameter that affects model performance. In summary, this work not only provides an efficient computational framework for piRNA disease association prediction, but also offers new ideas for other non coding RNA related research, with good theoretical significance and application potential.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The source data of RandGCN are available at http:// www. rna-society.org/mndr.

Acknowledgments

This work is supported by the 2025 Graduate Innovative Research Project of Southwest Minzu University.

References

[1] Aravin A, Gaidatzis D, Pfeffer S, et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 2006; 442 (7099): 203–7.

[2] Liu Y, Dou M, Song X, et al. The emerging role of the piRNA/piwi complex in cancer. Mol Cancer 2019;18(1):123.

[3] Iwasaki YW, Siomi MC, Siomi H. PIWI-interacting RNA: its biogenesis and functions. Annu Rev Biochem 2015;84(1):405–33.

[4] Aravin AA, Hannon GJ, Brennecke J. The piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 2007;318(5851):761–4.

[5] Seto AG, Kingston RE,Lau NC. The coming of age for piwi proteins. Mol Cell 2007;26(5):603–9.

[6] Wang K,Wang T,Gao XQ, Chen XZ, Wang F, Zhou LY. Emerging functions of piwi-interacting RNAs in diseases. Journal of Cellular and Molecular Medicine.2021; 25(11): 4893–901.

[7] Zhou JY, Zhou WY, Zhang R. The potential mechanisms of piRNA to induce hepatocellular carcinoma in human. Med Hypotheses.2021;146:110400.

[8] Xie Q, Li Z, Luo X, et al. piRNA-14633 promotes cervical cancer cell malig-nancy in a METTL14-dependent m6A RNA methylation manner. J Transl Med. 2022;20(1):51.

[9] Zhong N, Nong XT, Diao JY, et al. piRNA-6426 increases DNMT3Bmediated SOAT1 methylation and improves heart failure. Aging-Us. 2022; 14:2678–94.

[10] Ernst C, Odom DT,Kutter C. The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. Nat Commun 2017;8(1):10.

[11] Thakker DR, Natt F, Hüsken D, Maier R, Müller M, van der Putten H, et al. Neurochemical and behavioral consequences of widespread gene knockdown in the adult mouse brain by using nonviral RNA interference. Proc Natl Acad Sci USA 2004; 101: 17270–5.

[12] Chen X, Sun Y-Z,Guan N-N, Qu J, Huang Z-A,Zhu Z-X, et al. Computational models for lncRNA function prediction and functional similarity calculation.Brief Funct Genom 2019;18: 58–82.

[13] Bagci H, Sriskandarajah N, Robert A, et al. Ribosomes guide pachytene piRNA formation on long intergenic piRNA precursors. Nat Cell Biol. 2020; 22:353-353.

[14] Wei H, Xu Y, Liu B. iPiDi-PUL:identifying Piwi-interacting RNA-disease associations based on positive unlabeled learning. Brief Bioinforma 2021;22: bbaa058.

[15] You Z-H, Huang Z-A, Zhu Z, Yan G-Y, Li Z-W, Wen Z, et al. PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction. PLoS Comput Biol 2017;13: e1005455.

[16] Chen X, Xie D, Zhao Q, et al. MicroRNAs and complex diseases: from experimental results to computational models. Brief Bioinform 2019; 20:515–39.

[17] Chen L, Heikkinen L, Wang C, et al. Trends in the development of miRNA bioinformatics tools. Brief Bioinform 2019;20: 1836-52.

[18] Chen X, Yan CC, Zhang X, et al. Long non-coding RNAs and complex diseases: from experimental results to computational models. Brief Bioinform 2017; 18:558–76.

[19] Signal B, Gloss BS, Dinger ME. Computational approaches for functional prediction and characterisation of long noncoding RNAs. Trends Genet 2016; 32:620-37.

[20] Chen L, Wang C, Sun H, et al. The bioinformatics toolbox for circRNA discovery and analysis. Brief Bioinform 2021; 22: 1706–28.

[21] Chen J, Lin J, Yongfei H, et al. RNA Disease v4.0: an updated resource of RNA-associated diseases, providing RNA-disease analysis, enrichment and prediction. Nucleic Acids Res 2023;51(D1): D1397–404.