


ORIGINAL RESEARCH OPEN ACCESS

PPDAMEGCN: Predicting piRNA-Disease Associations Based on Multi-Edge Type Graph Convolutional Network

Yinglong Peng¹ | Shuang Chu² | Xindi Huang²  | Yan Cheng²
¹School of Information and Intelligence, XiangXi Vocational and Technical College for Nationalities, Jishou, China | ²School of Informatics, Hunan University of Chinese Medicine, Changsha, China

Correspondence: Xindi Huang (xindi.huang@hnucm.edu.cn)

Received: 4 November 2024 | **Revised:** 16 January 2025 | **Accepted:** 6 March 2025

Handling Editor: Fang-Xiang Wu (Lead GE)

Funding: This work is supported in part by the National Natural Science Foundation of China (No. 62473149), Natural Science Foundation of Hunan Province of China (No. 2022JJ30428) and Excellent youth funding of Hunan Provincial Education Department (No. 22B0372).

Keywords: association prediction | multi-edge type graph convolutional network | piRNA-disease associations | similarity

ABSTRACT

Recently, many studies have proven that Piwi-interacting RNAs (piRNAs) play key roles in various biological processes and also associate with human complicated diseases. Therefore, in order to accelerate the traditional biomedical experimental methods for determining piRNA-disease associations, many computational approaches have been proposed. However, piRNA-disease associations can be classified into known and unknown associations, each of which may provide distinct types of information. Traditional graph convolutional networks (GCNs) typically treat all edges in a graph as identical, overlooking the fact that different edge types may carry different signals and influence the learning process in unique ways. In this study, we also provide a new piRNA-disease association prediction method, called PPDAMEGCN, based on a multi-edge type graph convolutional network. First, we calculate the piRNA sequence similarity based on the piRNA sequence information and Smith–Waterman method. The disease semantic similarity is also computed by disease ontology (DO). In addition, we calculate the Gaussian interaction profile (GIP) kernel similarities of piRNA and diseases through the known piRNA-disease associations. Then, we construct the piRNA similarity network by integrating the piRNA's sequence similarity and GIP similarity. We also construct the disease similarity network by integrating disease's semantic similarity and GIP similarity. Finally, we obtain the piRNA and disease embeddings by the multi-edge type graph convolutional network model on the heterogeneous piRNA-disease association network. The piRNA-disease pair association probability score is calculated by a multilayer perceptron (MLP) with its concatenated embedding. We also compare PPDAMEGCN to other piRNA-disease prediction methods. The experimental results show that our method outperforms compared methods.

1 | Introduction

PIWI-interacting RNAs (piRNAs) are well-known small RNAs of 19–35 nucleotides in length, slightly longer than miRNAs. They silence transposable elements, regulate gene expression and fight viral infections [1]. Many recent studies have

confirmed the role of piRNAs in regulating transposon mobility and activity in mammals [2, 3]. Like miRNAs, piRNAs also have a strong preference for 5' uridine [4]. In addition, due to the variety of source transposons, piRNA sequences are much more diverse than other known classes of cellular RNAs, constituting the largest class of noncoding RNAs [5]. The regulation of

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *IET Systems Biology* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

transposons by piRNAs can achieve ‘self’ and ‘nonself’ recognition, conceptually similar to that in immune systems. PiRNAs can effectively select and regulate the ‘nonself’ genes through a complex mechanism [6].

Furthermore, recent studies on association between the aberrant expression of piRNAs and diverse diseases have shown that piRNAs may serve as potential noninvasive biomarkers for disease diagnosis and prognosis. For example, piRNAs have been found to be differentially expressed in cardiovascular diseases (CVDs) [7]. A total of 21 piRNAs or piRNA clusters were differentially expressed in chronic thromboembolic pulmonary hypertension (CTEPH) patients compared with controls [8]. Four genes (CYCS, LIN7C, KPNA6 and RAB11 A) have been found to be downregulated in Alzheimer’s disease (AD) versus Huntington’s disease (HD) control tissues by RNA-seq, which are target by four piRNAs (piR-38240, piR-51810, piR-40666 and piR-34393) that are found to be reciprocally co-expressed also via RNA-seq [9]. There is some evidence suggesting that piRNAs also contribute to the development and progression of cancers of breast, liver, lung and so on [10]. The dysregulation of piRNA-14633 in cervical cancer tissues has shown that its dysregulation contributes to disease mechanisms rather than as a consequence [11]. The decreased expression of piR-55490 in lung cancer specimens and cell lines was associated with the decreased survival of lung cancer [12]. In addition, piR-4987 and piR-20365 were found to be upregulated in breast cancer [13]. Therefore, by considering piRNAs’ important role(s) in a wide range of diseases, some piRNA-related databases have been created, such as piRNABank [4], piRBase [14], piRDisease [15], MNDR [16–19], ncRPheno [20] and ncRNAvar [21]. piRNABank stores empirically known sequences and other related information on piRNAs reported in human, mouse and rat, which also supports organism and chromosome-wise comprehensive search features including accession numbers, localisation on chromosomes and so on. The latest piRBase release (v2.0, 2018) was more focused on the comprehensive annotation of piRNA sequences and potential information of piRNA targets and disease-related piRNA. piRDisease is the first piRNA-disease association database that incorporates experimentally supported data for piRNAs’ association with a wide range of diseases, which includes 7939 manually curated associations between 4796 piRNAs and 28 diseases. In addition, it also provides the detailed information of the piRNA in respective diseases, such as brief description, sequence and location information. MNDR is a diverse ncRNA-disease database, which not only integrates piRNA-disease associations but also integrates miRNA-disease and lncRNA-disease associations. The new version MNDR v4.0 contains 3,428,058 ncRNA-disease associations covering 18 RNA types, 117 species and 4090 diseases over 40,000 published studies. In addition, ncRPheno and ncRNAvar are also association databases between noncoding RNA and diseases. Based on these benchmark datasets of piRNA-disease associations, great effort has been made to discover biomarkers or to develop computational methods for accurately predicting piRNA-disease associations. iPiDA-sHN is the first piRNA-disease association predictor whose features are extracted by a convolutional neural network (CNN), which apply support vector machines (SVMs) with selected high quality negative samples and positive samples [22]. piRDA is also a deep learning-based piRNA-disease association prediction

method, which further improves the prediction performance by extracting the most significant and abstract information from raw sequences represented in a simplified piRNA-disease pair [23]. iPiDA-GCN is a graph convolutional network-based (GCN) piRNA-disease association prediction method. The main feature is that iPiDA-GCN constructs the graphs based on piRNA sequence information, disease semantic information and known piRNA-disease associations to extract the features of both piRNAs and diseases [24]. In addition, PUTransGCN [25], iPiDA-PUL [26] and iPiDA-SWGCN [27] are GCN-based methods, which obtain the piRNA and disease embeddings by attention encoding mechanism, supplementarily weight and embedding transformation GCN, respectively. However, although these methods have obtained good results in piRNA-disease association prediction, the prediction performance should be improved. In the field of piRNA-disease association prediction, many methods are inspired by approaches from other areas of bioinformatics, such as ncRNA-ncRNA association prediction, ncRNA-disease association prediction and molecular property prediction. For example, GCNCRF [28], SSCRb [29], SNMGCDa [30] and MPGK-LMI [31] employ GCN and the attention mechanism to predict ncRNA-ncRNA associations. LDAP [32], KBMF-MDI [33], LDICDL [34], KGANCDa [35] and IGNSCDa [36] predict ncRNA-disease associations by integrating multiple types of biological data and improved GCN. DMFGAM [37], DCAMCP [38], ATPFGT-muli [39] and DMFPGA [40] utilise multi-feature fusion and attention mechanisms for molecular property prediction. We are also deeply inspired by these studies.

In this study, we also provide a new multi-edge type GCN-based piRNA-disease association prediction method (called PPDA-MEGCN). First, we calculate the piRNA sequence similarity based on the Smith–Waterman scores between piRNA sequences. The disease semantic similarity is also computed by disease ontology (DO) information. In addition, we also calculate the Gaussian interaction profile (GIP) kernel similarities of piRNA and disease based on the known piRNA-disease associations. Then, the integrated piRNA similarity is obtained by the average of piRNA sequence similarity and piRNA GIP similarity. We also computed the integrated disease similarity by the average of disease semantic similarity and disease GIP similarity. Furthermore, in order to further remove the noise of the disease(piRNA) similarity network, we select top k neighbours for each disease(piRNA) to construct the final disease(piRNA) similarity network. Based on the disease(piRNA) similarity network, we apply GCN to extract the initial feature of disease(piRNA). We further obtain the piRNA and disease embeddings through the multi-edge type GCN model with the known piRNA-disease association network and initial feature of them. Finally, the association probability scores of piRNA-disease pairs are calculated by the multilayer perceptron (MLP) with their embeddings. In order to assess the prediction performance of our method and compare it with other piRNA-disease association prediction methods, we conduct 5-fold cross validation (5-CV) on piRDisease v1.0 and MNDR v4.0. In 5-CV, our method achieves an average area under the receiver operating characteristics curve (AUC) of 0.944 and area under the precision-recall curve (AUPR) of 0.485 on piRDisease v1.0. In addition, the AUC and AUPR values also reach 0.931 and 0.506 on MNDR v4.0, which are significantly higher than the

compared methods. The case studies also illustrate that PPDA-MEGCN is an effective piRNA-disease association prediction method and can further provide help for disease diagnosis and prognosis.

2 | Materials and Methods

2.1 | Dataset

In this study, we also use the well-curated MNDR and piR-Disease as benchmark datasets. They were also widely used in previous piRNA-disease association prediction methods. MNDR 4.0 is a comprehensive ncRNA-disease database that covers 18 RNA types, 117 species and 4090 diseases from over 40,000 published studies and 23 other experimentally validated databases. After removing the duplicate piRNA-disease associations, we obtain 9502 associations between 8178 piNRAs and 15 diseases. Similarly, after sorting and projecting the downloaded data from piRDisease v1.0 database, we also obtain 4981 piRNA-disease associations over 4349 piRNAs and 21 diseases. The detail of two benchmark datasets is shown in Table 1.

In order to calculate the piRNA sequence similarity, we download the piRNA sequence information that contains the attribute information of them from piRBase. The disease ontology (DO) information is also used to compute disease semantic similarity, and we also obtain it from the latest vision (revision 2702, 6 October 2014) [41].

2.2 | piRNA Sequence Similarity

Base on the attribute information of the piRNA sequence, we calculate the piRNA sequence similarity. For piRNA p_i and piRNA p_j , the sequence similarity of them can be calculated by Equation (1) and as follows:

$$PSS(p_i, p_j) = \frac{SW(p_i, p_j)}{\sqrt{SW(p_i, p_i) \times SW(p_j, p_j)}} \quad (1)$$

in which $SW(p_i, p_j)$ is the sequence alignment value of them, which is calculated by the Smith-Waterman alignment algorithm [42].

2.3 | Disease Semantic Similarity

Based on the directed acyclic graph (DAG) representation of each disease on a DO database, we also calculate the disease semantic similarity which is also widely used in noncoding

RNA-disease association prediction filed [11, 43, 44]. For a disease node d_i , we define $DAG(d_i) = (T(d_i), E(d_i))$ in which $T(d_i)$ and $E(d_i)$ are the node set and edge set, respectively. $T(d_i)$ contains disease nodes and its ancestor disease nodes and $E(d_i)$ contains the direct link between the parent disease nodes and the child disease nodes. For disease $T(d_i)$ and disease d_j , their semantic similarity is calculated by Equation (2) and as follows:

$$DSS(d_i, d_j) = \frac{\sum_{d_t \in T(d_i) \cap T(d_j)} (D1_{d_i}(d_t) + D1_{d_j}(d_t))}{DS1(d_i) + DS1(d_j)} \quad (2)$$

In which $DS1(d_i)$ represents the semantic value of disease d_i , and it is calculated by $DS1(d_i) = \sum_{d_k \in T(d_i)} D1_{(d_i)}(d_k)$. $D1_{(d_i)}(d_t)$ is the semantic contribution of disease d_t to disease d_i , which is computed by Equation (3) and as follows:

$$D1_{d_i}(d_t) = \begin{cases} 1, & d_t = d_i \\ \max \{ \Delta \times D1_{d_i}(d'_t) | d'_t \in d_t \}, & d_t \neq d_i \end{cases} \quad (3)$$

where Δ is the semantic contribution factor between disease d_i and its direct ancestor diseases, which is also set to be 0.5 [45].

2.4 | piRNA and Disease GIP Similarities

In this study, by considering the similar piRNA (disease) related with a similar disease (piRNA), we also further calculate the GIP similarities based on the known piRNA-disease associations. We denote that $P = \{p_1, p_2, p_3, \dots, p_N\}$ is the set of piRNAs and $D = \{d_1, d_2, d_3, \dots, d_{N_d}\}$ is the set of diseases, respectively. The adjacency matrix $Y \in \mathbb{N}_p \times \mathbb{N}_d$ represents the known piRNA-disease associations in which the value of y_{ij} is 1 if there is a known piRNA-disease association, and 0 otherwise. For piRNA p_i and piRNA p_j , the GIP similarity $PGS(p_i, p_j)$ is calculated by Equations (4) and (5), and as follows:

$$PGS(p_i, p_j) = \exp\left(-\gamma_p \|IP(p_i) - IP(p_j)\|^2\right) \quad (4)$$

$$\gamma_p = \frac{\gamma'_p}{\frac{1}{N_p} \sum_{i=1}^{N_p} \|IP(p_i)\|^2} \quad (5)$$

In which $IP(p_i) = \{y_{i1}, y_{i2}, \dots, y_{iN_d}\}$ is the association profile of piRNA p_i . The parameter γ_p controls the kernel bandwidth and is computed by γ'_p , which is set to be 1.

Similarly, we also compute the disease GIP similarity by Equations (6) and (7), and as follows:

$$DGS(d_i, d_j) = \exp\left(-\gamma_d \|IP(d_i) - IP(d_j)\|^2\right) \quad (6)$$

$$\gamma_d = \frac{\gamma'_d}{\frac{1}{N_d} \sum_{i=1}^{N_d} \|IP(d_i)\|^2} \quad (7)$$

In which $IP(d_i) = \{y_{1i}, y_{2i}, \dots, y_{N_p i}\}$ is also the association profile of disease d_i , and γ'_d is also set to be 1.

TABLE 1 | The statistics of the benchmark dataset.

Dataset	piRNA	Disease	Associations	Sparsity
MNDR v4.0	8178	15	9502	7.75%
piRDisease v1.0	4349	21	4981	5.42%

3 | Methodology

According to the successful application of GCN in association prediction tasks, we also provide a new multi-edge type GCN-based piRNA-disease association prediction method (called PPDAMEGCN). PPDAMEGCN composes of three parts (see Figure 1: (i) construct the integrated piRNA similarity network, the integrated disease similarity network and the piRNA-disease association network and initialise the features; (ii) apply GCN and MEGCN models to learn the feature embedding and topology embedding of piRNAs and diseases, respectively; and (iii) utilise attention mechanism and MLP to predict piRNA-disease associations.

3.1 | Integrating piRNA and Disease Similarities

After computing piRNA sequence similarity, disease semantic similarity and their GIP similarities, we further obtain their integrated similarities. For piRNA p_i and piRNA p_j , their final similarity can be computed by Equation (8) and as follows:

$$FPS(p_i, p_j) = \frac{PSS(p_i, p_j) + PGS(p_i, p_j)}{2} \quad (8)$$

Similarly, we also obtain the disease final similarity with the same method based on disease semantic similarity and disease GIP similarity. For disease d_i and disease d_j , the computation process is defined by Equation (9) and as follows:

$$FDS(p_i, p_j) = \frac{DSS(p_i, p_j) + DGS(p_i, p_j)}{2} \quad (9)$$

3.2 | Constructing piRNA and Disease k -Nearest Neighbourhoods (KNN) Graph

In order to obtain more effective similarity, we remove the noise by selecting k -nearest neighbourhoods of each piRNA and disease. Therefore, we construct the isomorphic network KNN graphs of piRNA and disease based on their final similarities. $A^p \in \mathbb{R}^{N_p \times N_p}$ and $A^d \in \mathbb{R}^{N_d \times N_d}$ represent the adjacency matrix of piRNA and disease, respectively. For piRNA p_i and piRNA p_j , the $A^p(i, j)$ of the computation process is defined by Equation (10) and as follows:

$$A^p(i, j) = \begin{cases} 1, & \text{if } p_j \in \tilde{N}_k(p_i) \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

In which $\tilde{N}_k(p_i) = \{p_i\} \cup \{N_k(p_i)\}$ is the set of piRNA p_i and its k -nearest neighbourhoods and is the k -nearest neighbourhoods of piRNA p_i . Similarly, the calculation process of $A^d \in \mathbb{R}^{N_d \times N_d}$ is also defined by Equation (11) and as follows:

$$A^d(i, j) = \begin{cases} 1, & \text{if } d_j \in \tilde{N}_k(d_i) \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

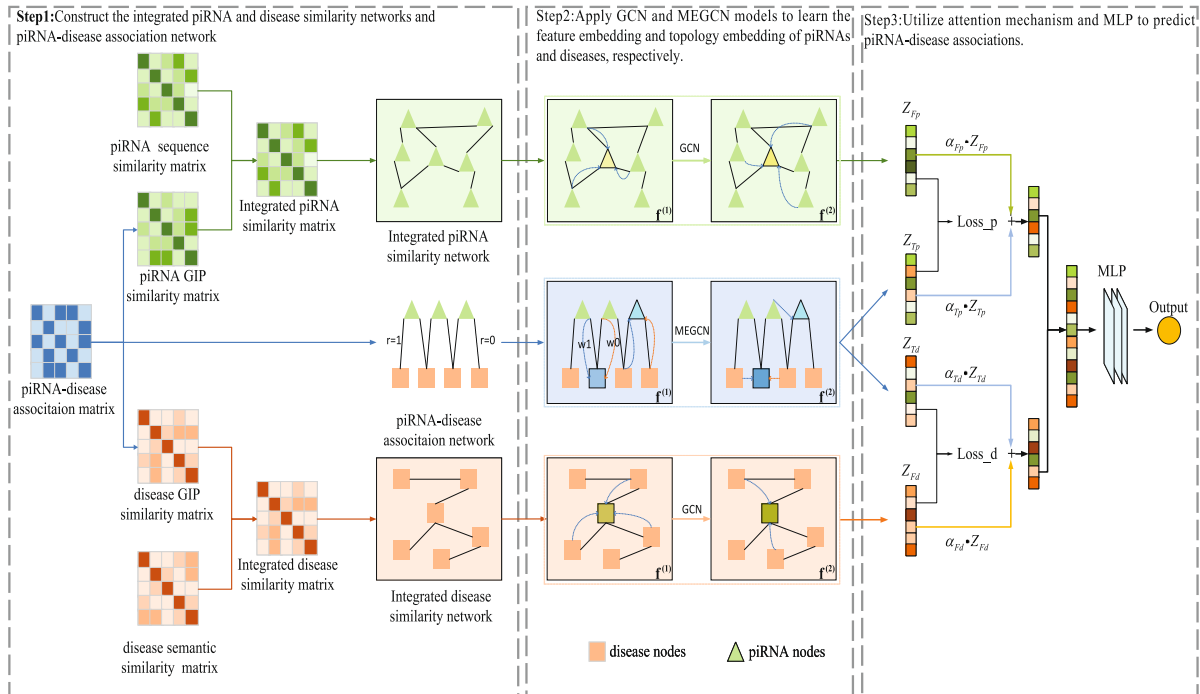


FIGURE 1 | Overview of PPDAMEGCN.

In which $\tilde{N}_k(d_i) = \{d_i\} \cup \{N_k(d_i)\}$ is the set of disease d_i and its k -nearest neighbourhoods and is the k -nearest neighbourhoods of disease d_i .

3.3 | Feature Embedding Based on Graph Convolutional Network

After obtaining the final similarity and KNN graph of piRNA and disease, we further apply the typical GCN model [46, 47] to extract their feature embeddings. The iteration process of l th layer is calculated by Equations (12) and (13), and as follows:

$$Z_p^{(l)} = \text{ReLU} \left(D_p^{\left(\frac{1}{2}\right)} A^p D_p^{\left(\frac{1}{2}\right)} Z_p^{(l-1)} W_p^{(l)} \right) \quad (12)$$

$$Z_d^{(l)} = \text{ReLU} \left(D_d^{\left(\frac{1}{2}\right)} A^d D_d^{\left(\frac{1}{2}\right)} Z_d^{(l-1)} W_d^{(l)} \right) \quad (13)$$

In which D_p and D_d are the diagonal degree matrices of A^p and A^d , respectively. W_p and W_d are the weight matrices of the l -th layer in GCN, respectively. In addition, the initial $Z_p^{(0)}$ is set to be FPS, and $Z_d^{(0)}$ is also set to be FDS. After this process, we obtain piRNA feature embedding Z_{fp} and disease feature embedding Z_{fd} .

3.4 | Topology Embedding Based on Multi-Edge Type Graph Convolutional Network

In order to obtain the topology embedding of piRNA and disease, we apply a multi-edge type graph convolutional network model on the piRNA-disease heterogeneous network. First, we consider the known and unknown piRNA-disease associations as different edge types and denote these edge types by $t \in \{0, 1\}$. Therefore, in the multi-edge type graph convolutional network (MEGCN) model, each edge type of graph convolution can be seen as a form of message passing [48]. For example, in edge-type t , the edge-type specific message $\text{MP}(\mu_{i-j,t})$ from piRNA p_i to disease d_j is defined by Equation (14) and as follows:

$$\text{MP}(\mu_{i-j,t}) = \frac{1}{C_{ij}} W_t x_j \quad (14)$$

In which C_{ij} is a normalisation constant $\sqrt{|N(p_i)| \cdot |N(d_j)|}$, where $N(p_i)$ and $N(d_j)$ are the sets of neighbours of piRNA p_i and disease d_j , respectively. W_t is the edge-type parameter matrix and x_i is the feature vector of piRNA p_i . Similarly, the message $\text{MP}(\mu_{i-j,t})$ from disease d_j to piRNA p_i is also calculated. Therefore, we can accumulate incoming messages of each node by summing all its neighbours $N_t(p_i)$ connected by a specific edge type, and the calculation process is defined by Equation (15) and as follows:

$$h_i = \sigma \left[\sum_{j \in N_t(p_i)} \text{MP}(\mu_{j-i,t}) \right] \quad (15)$$

where \sum is the accumulation operation, and σ is the activation function (tanh). Finally, we obtain the final representation z_i of piRNAs by a linear operator by Equation (16):

$$z_i = W h_i \quad (16)$$

where W is the parameter matrix. Similarly, we also compute the final representation of diseases. Based on the transformation to all nodes in the piRNA-disease heterogeneous graph, we can extract the piRNA topology embedding Z_{tp} and disease topology embedding Z_{td} .

3.5 | Attention Mechanism for Adaptive Learning

After obtaining the feature embeddings and topology embeddings of piRNA and diseases, we further apply the attention mechanism to adaptively learn the corresponding importance of their different embeddings. The specific computation process is defined by Equations (17) and (18), and as follows:

$$(\alpha_{fp}, \alpha_{tp}) = \text{att}(Z_{fp}, Z_{tp}) \quad (17)$$

$$(\alpha_{fd}, \alpha_{td}) = \text{att}(Z_{fd}, Z_{td}) \quad (18)$$

In which att is the attention operation. $\alpha_{fp} \in \mathbb{R}^{N_p \times 1}$ and $\alpha_{tp} \in \mathbb{R}^{N_p \times 1}$ are the attention weight values of piRNAs, respectively. $\alpha_{fd} \in \mathbb{R}^{N_d \times 1}$ and $\alpha_{td} \in \mathbb{R}^{N_d \times 1}$ are also the attention weight values of diseases, respectively.

Specifically, taking piRNA p_i as an example, its final weight values α_{fp}^i and α_{tp}^i can be calculated by normalisation with Equations (19) and (20), and as follows:

$$\alpha_{fp}^i = \text{softmax}(\omega_{fp}^i) = \frac{\exp(\omega_{fp}^i)}{\exp(\omega_{fp}^i) + \exp(\omega_{tp}^i)} \quad (19)$$

$$\alpha_{tp}^i = \text{softmax}(\omega_{tp}^i) = \frac{\exp(\omega_{tp}^i)}{\exp(\omega_{fp}^i) + \exp(\omega_{tp}^i)} \quad (20)$$

in which ω_{fp}^i and ω_{tp}^i are the attention values of piRNA p_i in feature embedding Z_{fp} and topology embedding Z_{tp} . Their calculation processes are defined by Equations (21) and (22) as follows:

$$\omega_{fp}^i = q^T \cdot \tanh \left(W_{fp} \cdot (z_{fp}^i)^T + b_{fp} \right) \quad (21)$$

$$\omega_{tp}^i = q^T \cdot \tanh \left(W_{tp} \cdot (z_{tp}^i)^T + b_{tp} \right) \quad (22)$$

In which q is the share vector, and W_{fp} and b_{fp} are the weight matrix and bias vector for piRNA feature embedding, respectively. In addition, W_{tp} and b_{tp} are also the weight matrix and

bias vector for piRNA topology embedding, respectively. Finally, we can obtain the piRNA final embedding Z_p by Equation (23):

$$Z_p = \alpha_{Fp} \cdot Z_{Fp} + \alpha_{Tp} \cdot Z_{Tp} \quad (23)$$

Similarly, we can also obtain the disease final embedding Z_d by Equation (24):

$$Z_d = \alpha_{Fd} \cdot Z_{Fd} + \alpha_{Td} \cdot Z_{Td} \quad (24)$$

3.6 | piRNA-Disease Association Prediction

After obtaining the piRNA final embeddings and disease final embeddings, we concatenate the two embeddings to represent a piRNA-disease pair. Then, we input this concatenated vector into a three-layer multilayer perceptron (MLP) to calculate the prediction result. For piRNA p_i and disease d_j , the calculation process of the association probability score \hat{y}_{ij} can be defined by Equation (25) and as follows:

$$\hat{y}_{ij} = \text{MLP}(z_p^i || z_d^j) \quad (25)$$

In this study, we also apply binary cross-entropy (BCE) loss as the main loss. Mathematically, it is described by Equation (26) and as follows:

$$l_{\text{BCE}} = -\sum_{(i,j)} \left[y_{ij} \cdot \log(\hat{y}_{ij}) + (1 - y_{ij}) \cdot \log(1 - \hat{y}_{ij}) \right] \quad (26)$$

in which (i, j) denotes the pair of piRNA p_i and disease d_j , and y_{ij} is the truth label. In addition, we also further consider the consistency of the feature space and topology space. We apply a consistency constraint by using L_2 -normalisation to normalise the embedding matrix. For piRNA, the constraint is defined by Equation (27) and as follows:

$$l_{(C_p)} = \|S_F^p - S_T^p\|_F^2 \quad (27)$$

in which S_F^p and S_T^p are the similarities of piRNA in the feature space and topology space, respectively. The specific process of them can be defined by Equations (28) and (29), and as follows:

$$S_F^p = Z_{Fp} \cdot Z_{Fp}^T \quad (28)$$

$$S_T^p = Z_{Tp} \cdot Z_{Tp}^T \quad (29)$$

Similarly, we can also calculate the disease consistency constraint l_{C_d} . Therefore, we can compute the final loss by weighted combing the BCE loss l_{BCE} , the consistency constraint losses l_{C_p} and l_{C_d} with Equation (30):

$$l = l_{\text{BCE}} + \lambda l_{(C_p)} + \lambda l_{(C_d)} \quad (30)$$

where λ is the hyperparameter to control the three loss terms.

4 | Experiments and Results

4.1 | Performance Evaluation

To evaluate the prediction performance of our method and compare it with other piRNA-disease association prediction methods, we conduct 5-fold cross validation (5-CV) on two datasets. In addition, we take the following three metrics for comparison, that is, area under curve (AUC), area under the precision-recall curve (AUPR) values and the rank of positive predictions (rank index) [26]. For both AUC and AUPR, higher values indicate better performance. On the contrary, a lower value of the rank index indicates a better performance of the model. The prediction scores of all the test piRNA-disease pairs are ranked in descending order and measured using the rank index. The rank index is calculated by Equation (31) and as follows:

$$\text{Rankindex} = \frac{1}{|S_{\text{test}}^+|} \sum_{\alpha \in S_{\text{test}}^+} \frac{r_{\alpha}}{|S_{\text{test}}|} \quad (31)$$

where $|S_{\text{test}}^+|$ is the number of all known piRNA-disease associations in the test subset S_{test}^+ , and $|S_{\text{test}}|$ is the number of all piRNA-disease pairs in the test set S_{test}^+ . α is an association in a positive test subset S_{test}^+ , and r_{α} represents its rank position of all positive test subsets.

We compare our method with other five typical piRNA-disease prediction methods, which include iPiDA-SWGCN [27], iPiDi-PUL [26], piRDA [23], iPiDA-GCN [24] and PUTransGCN [25]. iPiDA-PUL was the first piRNA-disease association prediction method employing positive unlabelled learning to select negative samples from all unlabelled piRNA-disease associations. piRDA is the deep learning-based piRNA-disease association prediction method. iPiDA-SWGCN, PUTransGCN and iPiDA-GCN are the GCN-based piRNA-disease association prediction methods. The validation experiments of our method and compared methods are conducted on the same cross-validation method.

4.2 | Compared it With Other Methods

1. 5-Fold Cross Validation on MNDR v4.0: As shown in Figures 2 and 3, the AUCs of PPDAMEGCN under 5-fold experiments are 0.932, 0.932, 0.930, 0.933 and 0.927, respectively. The AUPRs are 0.511, 0.505, 0.511, 0.518 and 0.484, respectively.

Table 2 shows the prediction performances of PPDAMEGCN and compared methods on MNDR v4.0. We can see from Table 2 that the AUC value of PPDAMEGCN reaches 0.931, which is higher than compared methods (iPiDASWGCN:0.919, piRDA:0.910, iPiDA-GCN:0.831, iPiDA-PUL(RF): 0.784, iPiDA-PUL(SVM): 0.725, iPiDA-PUL(DT):0.569 and PUTransGCN:0.566). In addition, PPDAMEGCN also outperforms other methods in terms of AUPR and the rank index, which reach 0.505 and 0.103, respectively. Figures 4 and 5 also show AUC and AUPR curves of PPDAMEGCN and compared methods on MNDR v4.0.

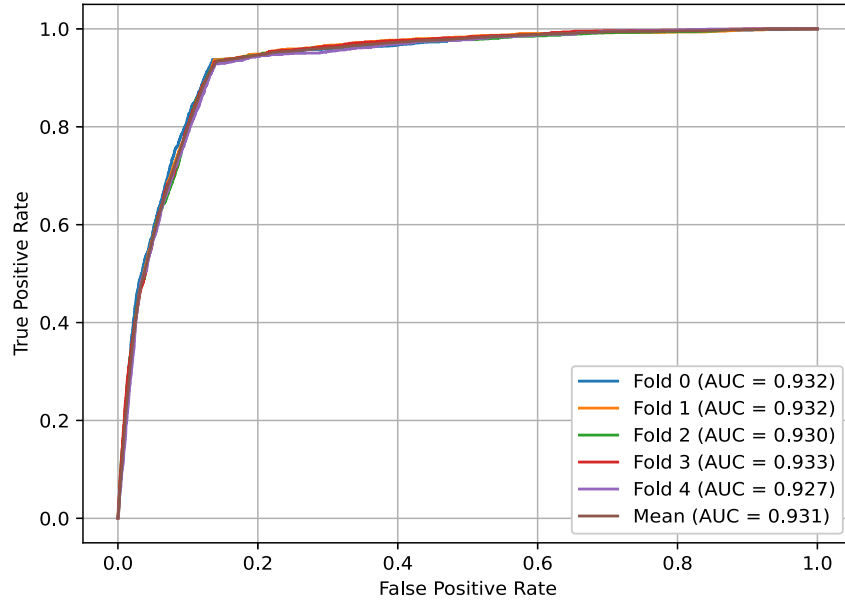


FIGURE 2 | The ROC curve of PPDAMEGCN on MNDR v4.0.

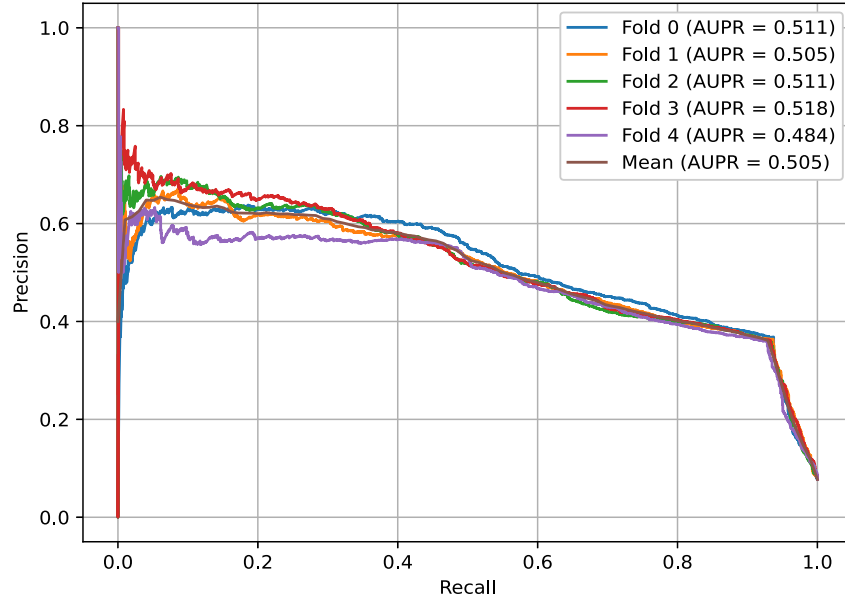


FIGURE 3 | The P-R curve of PPDAMEGCN on MNDR v4.0.

TABLE 2 | The results of PPDAMEGCN and compared methods on MNDR v4.0.

Methods	AUC	AUPR	Rank index
PPDAMEGCN	0.931	0.505	0.103
iPiDA-SWGCN	0.919	0.469	0.113
piRDA	0.910	0.343	0.122
iPiDA-GCN	0.831	0.477	0.198
iPiDA-PUL(RF)	0.784	0.166	0.238
iPiDA-PUL(SVM)	0.725	0.134	0.292
iPiDA-PUL(DT)	0.569	0.118	0.444
PUTransGCN	0.566	0.247	0.439

2. 5-Fold Cross Validation on piRDisease v1.0

As shown in Figure 6, the AUCs of PPDAMEGCN under 5-fold experiments are 0.937, 0.943, 0.949, 0.947 and 0.944, respectively. The AUPRs are 0.457, 0.497, 0.498, 0.506 and 0.467 as shown in Figure 7.

Table 3 shows the prediction performances of PPDAMEGCN and compared methods on piRDisease v1.0. Our method obtains AUC, AUPR and rank index values of 0.944, 0.484 and 0.080, respectively. The experimental results also illustrate that PPDAMEGCN is superior to compared methods. Figures 8 and 9 also show AUC and AUPR curves of PPDAMEGCN and compared methods on piRDisease v1.0.

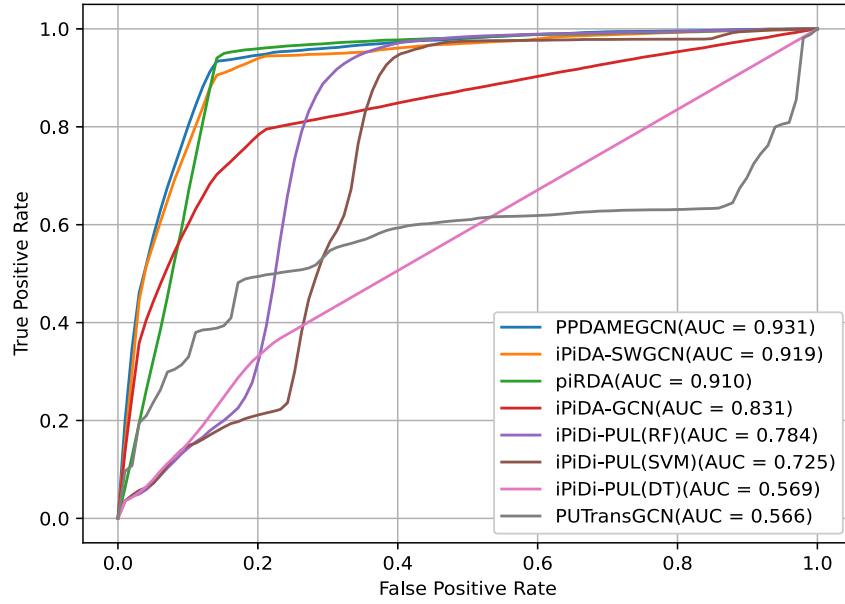


FIGURE 4 | The ROC curve of PPDAMEGCN and compared methods on MNDR v4.0.

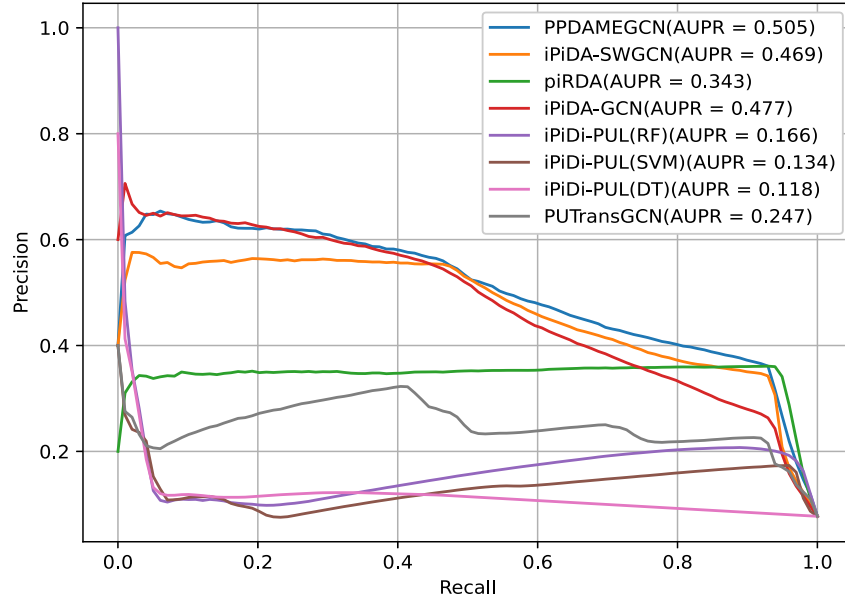


FIGURE 5 | The P-R curve of PPDAMEGCN and compared methods on MNDR v4.0.

5 | Ablation and Hyperparameters Analysis

In order to further evaluate the robustness and generalisability of our methods, we also conduct the ablation and hyperparameters analysis on MNDR v4.0 with 5-CV.

5.1 | Ablation Analysis

In this study, we conduct compared experiments on different strategies for the training method on MNDR v4.0 to assess their effectiveness. We analyse four cases which are defined as follows:

- **PPDAMEGCN-w/o-l**: PPDAMEGCN without l_{C_p} and l_{C_d} .

- **PPDAMEGCN-w/o-a**: PPDAMEGCN without the attention mechanism.
- **PPDAMEGCN-w/o-f**: PPDAMEGCN without using the feature embedding.
- **PPDAMEGCN-w/o-t**: PPDAMEGCN without using the topology embedding.

Table 4 shows the prediction performances of our methods on different strategies for the training method on MNDR v4.0. We can see from Table 4 that PPDAMEGCN obtains highest AUC and AUPR values of 0.929 and 0.501 when it applies all strategies. The experiment results also illustrate that all the strategies used can effectively improve the prediction performances of PPDAMEGCN.

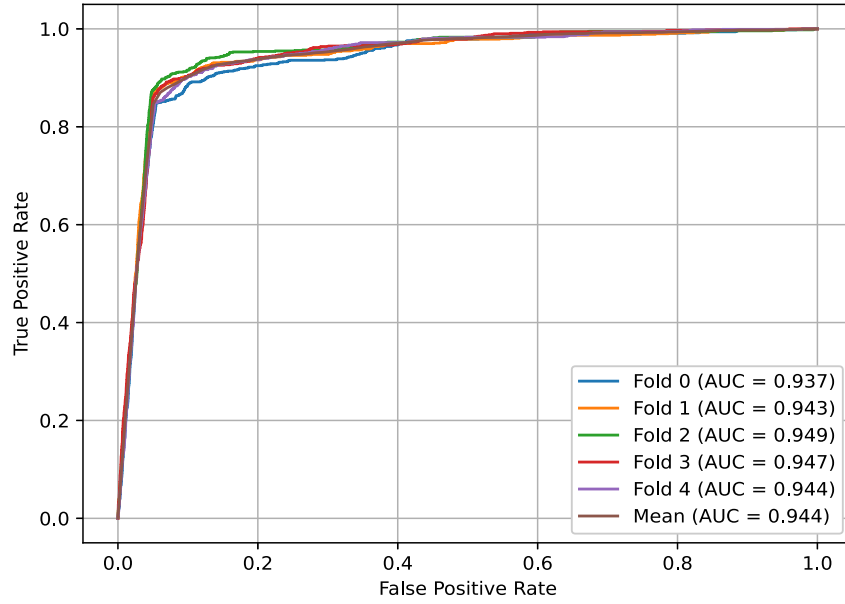


FIGURE 6 | The ROC curve of PPDAMEGCN on piRDisease v1.0.

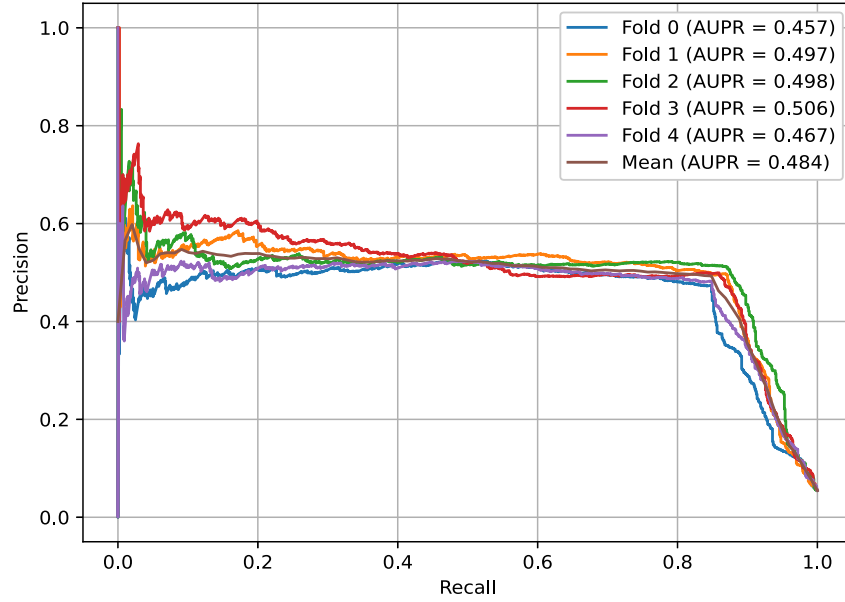


FIGURE 7 | The P-R curve of PPDAMEGCN on piRDisease v1.0.

TABLE 3 | The results of PPDAMEGCN and compared methods on piRDisease v1.0.

Methods	AUC	AUPR	Rank index
PPDAMEGCN	0.944	0.484	0.080
iPiDA-SWGCN	0.926	0.434	0.097
piRDA	0.567	0.071	0.436
iPiDA-GCN	0.844	0.457	0.173
iPiDA-PUL(RF)	0.500	0.377	0.500
iPiDA-PUL(SVM)	0.500	0.527	0.500
iPiDA-PUL(DT)	0.622	0.091	0.386
PUTransGCN	0.898	0.369	0.123

5.2 | Hyperparameters Analysis

In this study, we analyse the prediction performances of hyperparameters λ and k in our method. They are used to control the three loss terms and construct piRNA and disease k -nearest neighbourhood graphs, respectively. We set λ to be the default value when we analyse parameter k . Similarly, we also set k to be the default value when we analyse parameter λ .

We select the value of λ from [0.001, 0.01, 0.1, 1, 10, 100]. Table 5 shows the variation of AUC and AUPR for MNDR v4.0 on 5-CV. We can see from Table 5 that our method obtains the best prediction performances when λ is set to 0.001. In addition, it exhibits a declining trend as λ is increased from 0.001 to 100.

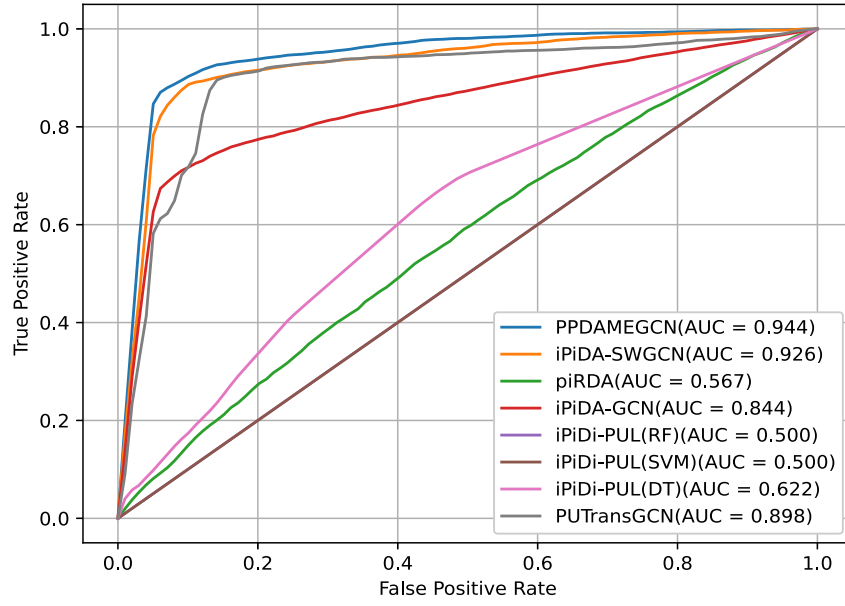


FIGURE 8 | The ROC curve of PPDAMEGCN and compared methods on piRDisease v1.0.

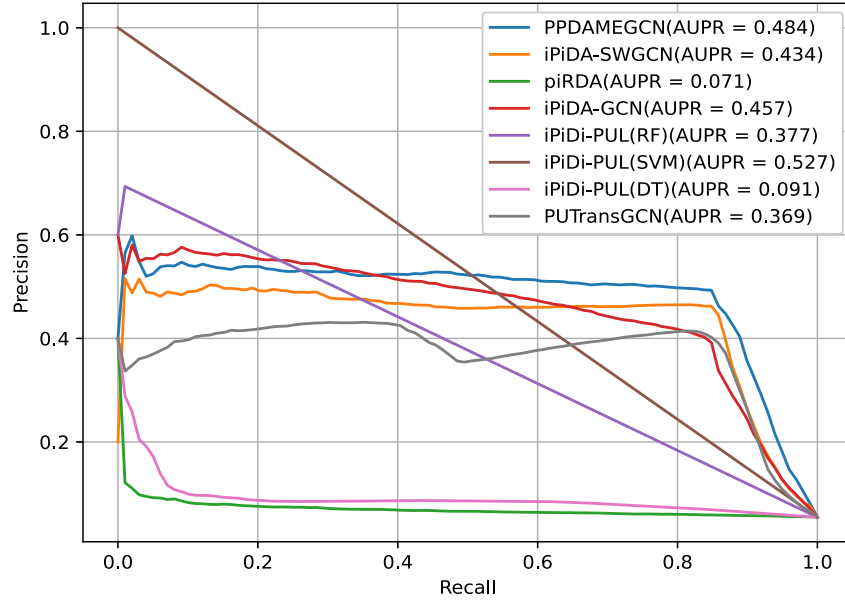


FIGURE 9 | The P-R curve of PPDAMEGCN and compared methods on piRDisease v1.0.

TABLE 4 | This is a sample table caption.

Methods	AUC	AUPR
PPDAMEGCN	0.931	0.505
PPDAMEGCN-w/o-l	0.656	0.269
PPDAMEGCN-w/o-a	0.774	0.404
PPDAMEGCN-w/o-f	0.427	0.067
PPDAMEGCN-w/o-t	0.774	0.498

In this study, we select the value of k from [1, 2, 4, 8 and 12]. Table 6 shows the variation of AUC and AUPR for MNDR v4.0 on 5-CV. We can see from Table 6 that our method obtains best prediction performances with AUC and AUPR values of 0.931 and 0.505, respectively. Therefore, we set the default value of k to 2.

TABLE 5 | The prediction performances of PPDAMEGCN with different λ on MNDR v4.0.

λ	0.001	0.01	0.1	1	10	100
AUC	0.931	0.929	0.927	0.923	0.915	0.911
AUPR	0.505	0.495	0.498	0.490	0.482	0.469

TABLE 6 | The prediction performances of PPDAMEGCN with different λ on MNDR v4.0.

k	1	2	4	8	12
AUC	0.929	0.931	0.930	0.925	0.650
AUPR	0.501	0.505	0.498	0.483	0.248

To further evaluate the performance of PPDAMEGCN for identifying new piRNA-disease associations, we conduct a case study on Parkinson disease and the top 20 scoring piRNA-disease associations are selected to be verified. Specifically, all known associations in MNDR v4.0 are as the training set, and potential associations are scored by PPDAMEGCN to find possible biomarkers. As shown in Table 7, the top 20 scoring piRNA-disease associations have been verified by the biological literature. For example, piR-hsa-28607 and piR-hsa-23337 are downregulated in Parkinson's disease-derived neuronal Cells. On the contrary, piR-hsa-26768 and piR-hsa-26343 are upregulated in Parkinson's disease-derived neuronal cells.

Similar to the case study on Parkinson disease, we also conduct a case study on lung cancer. As shown in Table 8, 9 of the top 20 potential piRNAs linked to lung cancer have been verified by the biological literature. For example, piR-hsa-15865 and piR-hsa-9277 are downregulated in lung cancer cells. Compared to the case study on Parkinson disease, there are few validated piRNAs associated with lung cancer. This may be due to the fact that the validation of piRNA involvement in lung cancer is still at an early stage. We believe these unverified piRNAs are worth for further investigation. We expect that as more research is conducted, the number of validated piRNAs associated with lung cancer will increase. In future work, we anticipate that more biological literature will become available to support the piRNA-disease associations, particularly for diseases like lung cancer.

With the development of high-throughput sequencing technology, many experimental studies have uncovered that piRNAs play important roles in some human complicated diseases, which also indicates that the piRNAs may serve as potential noninvasive biomarkers for disease diagnosis and prognosis. Therefore, in order to reveal the pathogenic mechanism of these diseases and provide help to diagnosis and treatment of diseases, many computational methods have developed to identify potential piRNA-disease associations. Although these methods have obtained some good results, the prediction performance should still be further improved.

In this study, we also provide a computational method, PPDAMEGCN, to predict potential piRNA-disease associations. PPDAMEGCN integrates piRNA sequence similarity and GIP similarity and disease semantic similarity and GIP similarity. We also extract the feature embedding and topology embedding of piRNA and disease based on GCN and MEGCN models, respectively. By the attention mechanism for the adaptive learning model, we extract final embeddings of piRNA-disease pairs and input them into MLP to compute the association probability scores. The compared experiment results and case studies show that our method outperforms compared methods and can provide help to systemically understand the pathogenic mechanism of diseases. However, there are several limitations in the current work that should be explored in future research. Firstly, the current model does not fully incorporate the complex biological data associated with piRNA expression profiles, disease-gene

TABLE 7 | The top 20 associated piRNAs of Parkinson disease and relevant evidences.

Disease	Rank	piRNA	piRNA ID	Evidence
Parkinson disease	1	piR-hsa-28607	DQ598392	PMID:29986767
	2	piR-hsa-26767	DQ596551	PMID:29986767
	3	piR-hsa-26765	DQ596549	PMID:29986767
	4	piR-hsa-26768	DQ596552	PMID:29986767
	5	piR-hsa-26343	DQ596111	PMID:29986767
	6	piR-hsa-23337	DQ593059	PMID:29986767
	7	piR-hsa-27885	DQ597640	PMID:29986767
	8	piR-hsa-21184	DQ590905	PMID:29986767
	9	piR-hsa-11768	DQ581483	PMID:29986767
	10	piR-hsa-11765	DQ581480	PMID:29986767
	11	piR-hsa-11778	DQ581494	PMID:29986767
	12	piR-hsa-11780	DQ581496	PMID:29986767
	13	piR-hsa-3689	DQ573396	PMID:29986767
	14	piR-hsa-3690	DQ573397	PMID:29986767
	15	piR-hsa-14509	DQ584287	PMID:29986767
	16	piR-hsa-14510	DQ584288	PMID:29986767
	17	piR-hsa-3693	DQ573400	PMID:29986767
	18	piR-hsa-19688	DQ589420	PMID:29986767
	19	piR-hsa-19804	DQ589536	PMID:29986767
	20	piR-hsa-10425	DQ580185	PMID:29986767

TABLE 8 | The top 20 associated piRNAs of lung cancer and relevant evidences.

Disease	Rank	piRNA	piRNA ID	Evidence
Lung cancer	1	piR-hsa-20256	DQ590003	PMID:28817650
	2	piR-hsa-15865	DQ585570	PMID:28817650
	3	piR-hsa-9277	DQ578964	None
	4	piR-hsa-9278	DQ578965	PMID:28817650
	5	piR-hsa-12743	DQ582520	PMID:28817650
	6	piR-hsa-18971	DQ588674	PMID:28817650
	7	piR-hsa-20084	DQ589816	PMID:28817650
	8	piR-hsa-21757	DQ591493	PMID:28817650
	9	piR-hsa-26769	DQ596553	None
	10	piR-hsa-26764	DQ596548	None
	11	piR-hsa-22696	DQ592491	PMID:28817650
	12	piR-hsa-21096	DQ590800	None
	13	piR-hsa-15341	DQ585030	None
	14	piR-hsa-17334	DQ587028	PMID:28817650
	15	piR-hsa-24108	DQ593860	None
	16	piR-hsa-28607	DQ598392	None
	17	piR-hsa-18669	DQ589881	None
	18	piR-hsa-26767	DQ596551	None
	19	piR-hsa-26768	DQ596552	None
	20	piR-hsa-26765	DQ596549	None

associations or other relevant molecular data. Integrating these additional layers of information could potentially improve the prediction accuracy of piRNA-disease associations. Secondly, consider applying a contrastive learning model to better integrate and utilise these diverse biological data. Contrastive learning has shown promise in effectively handling multi-modal data and could enhance the model's ability to differentiate between positive and negative associations in a more robust manner. In future work, we aim to address these limitations by incorporating these biological data sources and exploring the potential of contrastive learning models to develop a more accurate and effective piRNA-disease association prediction method.

Author Contributions

Yinglong Peng: conceptualisation, data curation, methodology, writing – original draft. **Shuang Chu:** data curation, formal analysis, writing – review and editing. **Xindi Huang:** conceptualisation, methodology, supervision, writing – review and editing. **Yan Cheng:** conceptualisation, funding acquisition.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China (No. 62473149), Natural Science Foundation of Hunan Province of China (No. 2022JJ30428) and Excellent youth funding of Hunan Provincial Education Department (No. 22B0372).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The original data presented in the study are openly available in the public domain: piRDisease v1.0 at <http://www.piwirna2disease.org/index.php>, RNADisease v4.0 at <http://www.rnadisease.org> or <http://www.rna-society.org/mndr/>. The code and data that support the findings of this study are available at <https://github.com/27167199/PPDAMEGCN>.

References

1. D. M. Ozata, I. Gainetdinov, A. Zoch, D. Ó Carroll, and P. D. Zamore, "PIWI-Interacting RNAs: Small RNAs With Big Functions," *Nature Reviews Genetics* 20, no. 2 (2019): 89–108, <https://doi.org/10.1038/s41576-018-0073-3>.
2. K. A. O'Donnell and J. D. Boeke, "Mighty Piwis Defend the Germline Against Genome Intruders," *Cell* 129, no. 1 (2007): 37–44, <https://doi.org/10.1016/j.cell.2007.03.028>.
3. H. Lin, "piRNAs in the Germ Line," *Science* 316, no. 5823 (2007): 397, <https://doi.org/10.1126/science.1137543>.
4. S. Lakshmi and S. Agrawal, "piRNABank: A Web Resource on Classified and Clustered Piwi-Interacting RNAs," *Nucleic Acids Research* 36, no. suppl_1 (2008): D173–D177, <https://doi.org/10.1093/nar/gkm696>.
5. M. C. Siomi, K. Sato, D. Pezic, and A. A. Aravin, "Piwi-interacting Small RNAs: The Vanguard of Genome Defence," *Nature Reviews Molecular Cell Biology* 12, no. 4 (2011): 246–258, <https://doi.org/10.1038/nrm3089>.
6. C. D. Malone and G. J. Hannon, "Small RNAs as Guardians of the Genome," *Cell* 136, no. 4 (2009): 656–668, <https://doi.org/10.1016/j.cell.2009.01.045>.
7. M. Li, Y. Yang, Z. Wang, et al., "Piwi-interacting RNAs (piRNAs) as Potential Biomarkers and Therapeutic Targets for Cardiovascular Diseases," *Angiogenesis* 24, no. 1 (2021): 19–34, <https://doi.org/10.1007/s10456-020-09750-w>.

8. C. Lipps, P. Northe, R. Figueiredo, et al., "Non-invasive Approach for Evaluation of Pulmonary Hypertension Using Extracellular Vesicle-Associated Small Non-coding RNA," *Biomolecules* 9, no. 11 (2019): 666, <https://doi.org/10.3390/biom9110666>.
9. Y. H. Sun, J. Zhu, L. H. Xie, et al., "Ribosomes Guide Pachytene piRNA Formation on Long Intergenic piRNA Precursors," *Nature Cell Biology* 22, no. 2 (2020): 200–212, <https://doi.org/10.1038/s41556-019-0457-4>.
10. Y. Liu, M. Dou, X. Song, et al., "The Emerging Role of the piRNA/piwi Complex in Cancer," *Molecular Cancer* 18 (2019): 1–15, <https://doi.org/10.1186/s12943-019-1052-9>.
11. Q. Xie, Z. Li, X. Luo, et al., "piRNA-14633 Promotes Cervical Cancer Cell Malignancy in a METTL14-dependent m6A RNA Methylation Manner," *Journal of Translational Medicine* 20, no. 1 (2022): 51, <https://doi.org/10.1186/s12967-022-03257-2>.
12. L. Peng, L. Song, C. Liu, et al., "piR-55490 Inhibits the Growth of Lung Carcinoma by Suppressing mTOR Signaling," *Tumor Biology* 37, no. 2 (2016): 2749–2756, <https://doi.org/10.1007/s13277-015-4056-0>.
13. G. Huang, H. Hu, X. Xue, et al., "Altered Expression of piRNAs and Their Relation With Clinicopathologic Features of Breast Cancer," *Clinical and Translational Oncology* 15, no. 7 (2013): 563–568, <https://doi.org/10.1007/s12094-012-0966-0>.
14. J. Wang, Y. Shi, H. Zhou, et al., "piRBase: Integrating piRNA Annotation in All Aspects," *Nucleic Acids Research* 50, no. D1 (2022): D265–D272, <https://doi.org/10.1093/nar/gkab1012>.
15. A. Muhammad, R. Waheed, N. A. Khan, H. Jiang, and X. Song, "piRDisease v1. 0: A Manually Curated Database for piRNA Associated Diseases," *Database* 2019 (2019): baz052, <https://doi.org/10.1093/database/baz052>.
16. Y. Wang, L. Chen, B. Chen, et al., "Mammalian ncRNA-Disease Repository: A Global View of ncRNA-Mediated Disease Network," *Cell Death & Disease* 4, no. 8 (2013): e765, <https://doi.org/10.1038/cddis.2013.292>.
17. T. Cui, L. Zhang, Y. Huang, et al., "MNDR v2. 0: An Updated Resource of ncRNA-Disease Associations in Mammals," *Nucleic Acids Research* 46, no. D1 (2018): D371–D374.
18. L. Ning, T. Cui, B. Zheng, et al., "MNDR v3. 0: Mammal ncRNA-Disease Repository With Increased Coverage and Annotation," *Nucleic Acids Research* 49, no. D1 (2021): D160–D164, <https://doi.org/10.1093/nar/gkaa707>.
19. J. Chen, J. Lin, Y. Hu, et al., "RNADisease v4.0: An Updated Resource of RNA-Associated Diseases, Providing RNA-Disease Analysis, Enrichment and Prediction," *Nucleic Acids Research* 51, no. D1 (2023): D1397–D1404, <https://doi.org/10.1093/nar/gkac814>.
20. W. Zhang, G. Yao, J. Wang, M. Yang, H. Zhang, and W. Li, "ncRPheno: A Comprehensive Database Platform for Identification and Validation of Disease Related Noncoding RNAs," *RNA Biology* 17, no. 7 (2020): 943–955, <https://doi.org/10.1080/15476286.2020.1737441>.
21. W. Zhang, B. Zeng, M. Yang, et al., "ncRNAVar: A Manually Curated Database for Identification of Noncoding RNA Variants Associated With Human Diseases," *Journal of Molecular Biology* 433, no. 11 (2021): 166727, <https://doi.org/10.1016/j.jmb.2020.166727>.
22. H. Wei, Y. Ding, and B. Liu, "iPiDi-sHN: Identification of Piwi-Interacting RNA-Disease Associations by Selecting High Quality Negative Samples," *Computational Biology and Chemistry* 88 (2020): 107361, <https://doi.org/10.1016/j.compbiolchem.2020.107361>.
23. S. D. Ali, H. Tayara, and K. T. Chong, "Identification of piRNA Disease Associations Using Deep Learning," *Computational and Structural Biotechnology Journal* 20 (2022): 1208–1217, <https://doi.org/10.1016/j.csbj.2022.02.026>.
24. J. Hou, H. Wei, and B. Liu, "iPiDi-GCN: Identification of piRNA-Disease Associations Based on Graph Convolutional Network," *PLoS Computational Biology* 18, no. 10 (2022): e1010671, <https://doi.org/10.1371/journal.pcbi.1010671>.
25. Q. Chen, L. Zhang, Y. Liu, Z. Qin, and T. Zhao, "PUTransGCN: Identification of piRNA-Disease Associations Based on Attention Encoding Graph Convolutional Network and Positive Unlabelled Learning," *Briefings in Bioinformatics* 25, no. 3 (2024): bbae144, <https://doi.org/10.1093/bib/bbae144>.
26. H. Wei, Y. Xu, and B. Liu, "iPiDi-PUL: Identifying Piwi-Interacting RNA-Disease Associations Based on Positive Unlabeled Learning," *Briefings in Bioinformatics* 22, no. 3 (2021): bbaa058, <https://doi.org/10.1093/bib/bbaa058>.
27. J. Hou, H. Wei, and B. Liu, "iPiDi-SWGCN: Identification of piRNA-Disease Associations Based on Supplementarily Weighted Graph Convolutional Network," *PLoS Computational Biology* 19, no. 6 (2023): e1011242, <https://doi.org/10.1371/journal.pcbi.1011242>.
28. W. Wang, L. Zhang, J. Sun, Q. Zhao, and J. Shuai, "Predicting the Potential Human lncRNA-miRNA Interactions Based on Graph Convolution Network With Conditional Random Field," *Briefings in Bioinformatics* 23, no. 6 (2022): bbac463, <https://doi.org/10.1093/bib/bbac463>.
29. L. Liu, Y. Wei, Q. Zhang, and Q. Zhao, "SSCRB: Predicting circRNA-RBP Interaction Sites Using a Sequence and Structural Feature-Based Attention Model," *IEEE Journal of Biomedical and Health Informatics* 28, no. 3 (2024): 1762–1772, <https://doi.org/10.1109/jbhi.2024.3354121>.
30. S. Yin, P. Xu, Y. Jiang, et al., "Predicting the Potential Associations Between circRNA and Drug Sensitivity Using a Multisource Feature-Based Approach," *Journal of Cellular and Molecular Medicine* 28, no. 19 (2024): e18591, <https://doi.org/10.1111/jcmm.18591>.
31. J. Xie, P. Xu, Y. Lin, et al., "LncRNA-miRNA Interactions Prediction Based on Meta-Path Similarity and Gaussian Kernel Similarity," *Journal of Cellular and Molecular Medicine* 28, no. 19 (2024): e18590, <https://doi.org/10.1111/jcmm.18590>.
32. W. Lan, M. Li, K. Zhao, et al., "LDAP: A Web Server for lncRNA-Disease Association Prediction," *Bioinformatics* 33, no. 3 (2017): 458–460, <https://doi.org/10.1093/bioinformatics/btw639>.
33. W. Lan, J. Wang, M. Li, J. Liu, F. Wu, and Y. Pan, "Predicting microRNA-Disease Associations Based on Improved microRNA and Disease Similarities," *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 15, no. 6 (2018): 1774–1782, <https://doi.org/10.1109/tcbb.2016.2586190>.
34. W. Lan, D. Lai, Q. Chen, et al., "LDICDL: lncRNA-Disease Association Identification Based on Collaborative Deep Learning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 19, no. 3 (2022): 1715–1723, <https://doi.org/10.1109/tcbb.2020.3034910>.
35. W. Lan, Y. Dong, Q. Chen, R. Zheng, J. Liu, and Y. Pan, "KGANCA: Predicting circRNA-Disease Associations Based on Knowledge Graph Attention Network," *Briefings in Bioinformatics* 23, no. 1 (2022): bbab494, <https://doi.org/10.1093/bib/bbab494>.
36. W. Lan, Y. Dong, Q. Chen, J. Liu, J. Wang, and S. Pan, "IGNSCDA: Predicting circRNA-Disease Associations Based on Improved Graph Convolutional Network and Negative Sampling," *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 19, no. 6 (2022): 3530–3538, <https://doi.org/10.1109/tcbb.2021.3111607>.
37. T. Wang, J. Sun, and Q. Zhao, "Investigating Cardiotoxicity Related With hERG Channel Blockers Using Molecular Fingerprints and Graph Attention Mechanism," *Computers in Biology and Medicine* 153 (2023): 106464, <https://doi.org/10.1016/j.compbiomed.2022.106464>.
38. Z. Chen, L. Zhang, J. Sun, R. Meng, S. Yin, and Q. Zhao, "DCAMCP: A Deep Learning Model Based on Capsule Network and Attention Mechanism for Molecular Carcinogenicity Prediction," *Journal of Cellular and Molecular Medicine* 27, no. 20 (2023): 3117–3126, <https://doi.org/10.1111/jcmm.17889>.

39. X. Yang, J. Sun, B. Jin, et al., "Multi-task Aquatic Toxicity Prediction Model Based on Multi-Level Features Fusion," *Journal of Advanced Research* 68 (2024): 477–489, <https://doi.org/10.1016/j.jare.2024.06.002>.
40. J. Wang, L. Zhang, J. Sun, et al., "Predicting Drug-Induced Liver Injury Using Graph Attention Mechanism and Molecular Fingerprints," *Methods* 221 (2024): 18–26, <https://doi.org/10.1016/j.ymeth.2023.11.014>.
41. W. A. Kibbe, C. Arze, V. Felix, et al., "Disease Ontology 2015 Update: An Expanded and Updated Database of Human Diseases for Linking Biomedical Knowledge through Disease Data," *Nucleic Acids Research* 43, no. D1 (2015): D1071–D1078, <https://doi.org/10.1093/nar/gku1011>.
42. T. F. Smith and M. S. Waterman, "Identification of Common Molecular Subsequences," *Journal of Molecular Biology* 147, no. 1 (1981): 195–197, [https://doi.org/10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5).
43. C. Yan and G. Duan, "PMDAGS: Predicting miRNA-Disease Associations With Graph Nonlinear Diffusion Convolution Network and Similarities," *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 21, no. 3 (2024): 394–404, <https://doi.org/10.1109/tcbb.2024.3366175>.
44. S. Chu, G. Duan, and C. Yan, "PGCNMDA: Learning Node Representations along Paths With Graph Convolutional Network for Predicting miRNA-Disease Associations," *Methods* 229 (2024): 71–81, <https://doi.org/10.1016/j.ymeth.2024.06.007>.
45. C. Yan, J. Wang, P. Ni, W. Lan, F. Wu, and Y. Pan, "DNRLMF-MDA: Predicting microRNA-Disease Associations Based on Similarities of microRNAs and Diseases," *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 16, no. 1 (2017): 233–243, <https://doi.org/10.1109/tcbb.2017.2776101>.
46. Kipf T. N., Welling M., "Semi-supervised Classification With Graph Convolutional Networks." arXiv preprint arXiv:1609.02907 2016.
47. F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "Simplifying Graph Convolutional Networks," in *International Conference on Machine Learning* (PMLR, 2019), 6861–6871.
48. B. Rvd, Kipf T. N., Welling M., "Graph Convolutional Matrix Completion." arXiv preprint arXiv:1706.02263 2017.