BMC Bioinformatics

RESEARCH Open Access

Multi-granularity transformer contrastive learning and feature reconstruction for prediction of disease-related miRNAs

Ping Xuan[1,2*], Zhicheng Guo[2], Siyuan Lu[3], Hui Cui[4], Jian Ding[5] and Tiangang Zhang[1,3*]

*Correspondence:
Ping Xuan
pingxuan@hainanu.edu.cn
Tiangang Zhang
zhangtg@hainanu.edu.cn
[1]School of Cyberspace Security, Hainan University, Haikou 570228, China
[2]Department of Computer Science and Technology, Shantou University, Shantou 515063, China
[3]School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China
[4]Department of Computer Science and Information Technology, La Trobe University, Melbourne 3083, Australia
[5]Harbin Conservatory of Music, Harbin 150028, China

Abstract

The increasing research confirms the function of microRNAs (miRNAs) play a vital role in the diagnosis, treatment, and prognosis of diseases. Screening the potential candidate miRNAs related to the diseases by the computational methods may be helpful for reducing the experimental cost for discovering the actual disease-miRNA associations. Most of the recent methods exploited the data from multiple sources to infer the association tendencies between miRNAs and diseases. However, these methods did not completely learn the features of the multi-source data from multiple granularities. This paper present an **a**ssociation **p**rediction model based on **m**ulti-granularity transformer contrastive learning and **f**eature reconstruction convolution (APMF). First, the strongly correlated node sequences are constructed by random walks, and then the features of the positive and negative node sequences for the target node are learned. Second, the multi-granularity transformer strategy is designed to refine the features of the miRNA, disease, and lncRNA nodes from fine-granularity and coarse-granularity, respectively. Third, the contrastive learning strategy is presented to maximize the consistency among the features of positive samples and the difference among the features of positive samples and negative ones. Finally, the feature reconstruction convolution is constructed to capture the diverse semantics of features and suppress the noises of the data. The experimental results indicate that APMF outperforms several state-of-the-art methods for prediction of disease-related miRNAs. The ablation experiments demonstrate the effectiveness of the multi-granularity transformer contrastive learning and the feature reconstruction convolution. The case analysis over three diseases also shows APMF's ability in discovering the potential miRNA candidates for the diseases.

Keywords Encoding association characteristics among multiple lncRNAs and diseases, Dynamic attribute learning of hyperedges, Dynamic topology formulation of hypergraph, Pairwise attribute learning enhanced by gated CNN, Attributes of multiple types of nodes

Introduction

Single-stranded noncoding microRNAs (miRNAs) containing about 22 nucleotides, which inhibit the expression of target messenger RNAs (mRNAs) [1–3]. Previous studies have confirmed that miRNAs are crucial in the progression of numerous human diseases [4–7]. For example, dysregulated miR-107 can disrupt BACE1 (secretase 10) function, potentially contributing to the development of Alzheimer's disease [8]. Therefore, it is imperative to get the prediction of disease-related miRNAs in order to explore the molecular pathogenesis of diseases.

Many computational methods have been developed to infer potential disease-miRNA associations. Current approaches can be classified into three categories. The first type of approaches are primarily based on the hypothesis that similar diseases prone to be related to similar miRNAs and vice versa [9]. NCPMDA [10] integrates disease semantic similarity, miRNA family information, miRNA functional similarity, and Jaccard similarity for diseases and miRNAs, then it reconstruct disease–disease and miRNA–miRNA similarity networks to predict disease–miRNA associations. Jiang et al. [11] proposed a model based on hypergeometric probability distribution and similarity network to predict candidate miRNAs associated with diseases. However, this method only considers the information from each node's direct neighbors and ignores other strongly correlated nodes information. Several predictive models are developed using restarted random walk techniques [12–15] and Xuan et al. [16] leverage miRNA family and cluster data to assess miRNA similarities and pinpoint potential disease-associated miRNAs by selecting the top k most similar neighbors. Additionally, they develope predictive models utilizing random walk algorithms on similarity-based networks, but these approaches face challenges when apply to diseases with no known associated miRNAs. In conclusion, approaches based on similarity measures are inadequate in capturing intricate connections and complex nonlinear interactions within network structures.

The second type of methods predicts the association between miRNA and diseases by using models based on machine learning techniques. DNI-MDCAP [17] proposed a method based on network imputation and semi-supervised learning. Liu et al. [18] designed a novel computational approach for predicting MDAs using Autoencoder-based Deep Forest ensemble learning. A matrix decomposition prediction method [19] is used to adaptively capture neighboring information of each node to strengthen its underlying representation. In addition, various approaches utilize data on diseases and miRNAs to predict disease-associated miRNAs through techniques such as support vector machines [20], random walk [21–23] non-negative matrix factorization [24–26]. However, these methods are shallow predictors that fail to derive distinctive embeddings for miRNAs and diseases from networks containing abundant semantic information, which limits their capacity to learn meaningful representations.

The third type establishes a prediction model leveraging deep learning to enhance prediction performance. AMHMDA [27] improves prediction performance by mining potential unknown relationships by utilizing hypergraph learning and attention-aware similarity networks. MAGCN [28] uses GCN with attention mechanism and a CNN encoder for feature learning. MGCNSS [29] is able to automatically capture meta-path relationships of varying lengths in heterogeneous networks by using graph convolution, and adopts a sample selection strategy based on long distance. The methods mentioned above mine deeper semantics in the graph, but they are overly dependent on graph

topology and features. MSGCL [30] and GCLMTP-MLP [31] use graph contrastive learning based on GCN to predict association, but they do not fully utilize the relationship between miRNA and disease. MDformer [32] learns feature information through transformer, but ignores graph topology information. These approaches have attained relatively good prediction effectiveness because deep learning methods are able to capture the underlying relationships between miRNA and disease.

Although similarity-based machine learning and deep learning methods have made progress, challenges remain. Notably, existing methods cannot fully capture multi-source heterogeneous information across miRNA, disease, and lncRNA networks. Second, for a sparse matrix with relatively few known associations, the predictive performance still needs improvement. Finally, most methods lack effective mechanisms to filter redundant features and suppress noise, which limits their generalization ability. To tackle these challenges, the APMF is proposed, which enhances node features through contrastive learning across different granularities and predicts miRNA-disease associations using feature reconstruction convolution module. The main contributions are as follows:

Firstly, we construct a miRNA–disease–lncRNA heterogeneous graph to establish a complex topological structure among lncRNA, disease, and miRNA nodes. Based on the heterogeneous graph we design a random walk module to obtain strongly correlated node sequences for the target node.

Secondly, sampling features of sequence nodes are optimized from both fine-grained and coarse-grained perspectives. Furthermore, a multi-granularity transformer module is proposed to refine and fuse the positive and negative sampling features of miRNA, disease, and lncRNA nodes accordingly.

Thirdly, we introduce a hybrid contrastive learning strategy that combines the InfoNCE loss for fine-grained features and the cosine similarity loss for coarse-grained features, thereby maximizing the consistency among positive samples while enhancing their distinction from negative samples to obtain semantically rich representations.

Finally, we propose a feature reconstruction convolution module to capture comprehensive semantic relationships and suppress noise. It calculates channel-wise feature importance, applies adaptive weights, filters redundant information through thresholding, and performs cross-reconstruction on the optimized features. The effectiveness of APMF is validated through comparisons with seven state-of-the-art models, ablation studies, and case analyses.

Materials and methods

To identify potential disease-associated miRNAs, we propose a miRNA-disease prediction model, APMF (Figs. 1 and 2). A graph is created to combine the similarities, associations and interactions among miRNAs, diseases, and lncRNAs. APMF consists of two components. In the first component (Fig. 1), the multi-granularity transformer learns the dependencies between information of different granularities on the random walk sequence and enhances the features through unsupervised contrastive learning. For the second component (Fig. 2), the feature reconstruction convolutional neural network can enhance informative features and suppress the redundant ones.

Additionally, the model leverages paired sampling features from target miRNA–disease nodes for contrastive learning (Fig. 1) and feature reconstruction (Fig. 2). Initially,

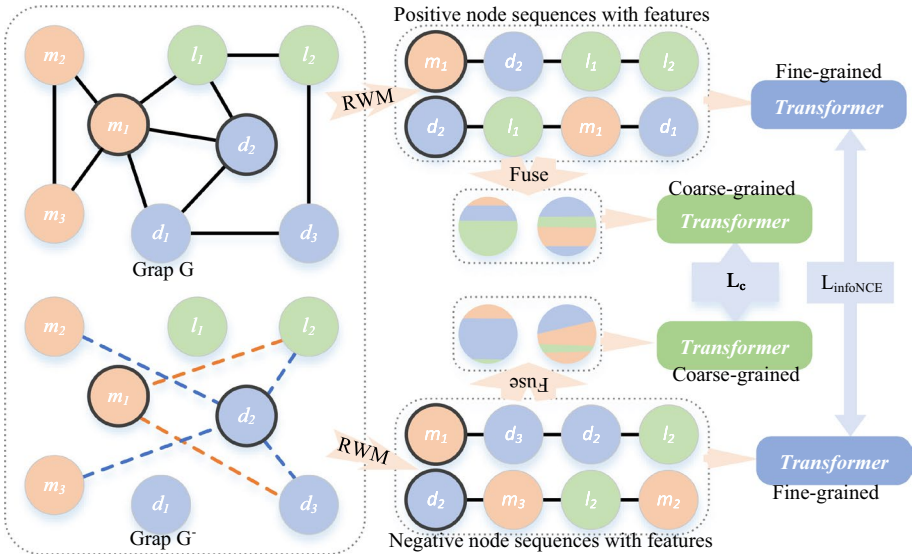

Fig. 1 Framework of multi-granularity transformer contrastive learning

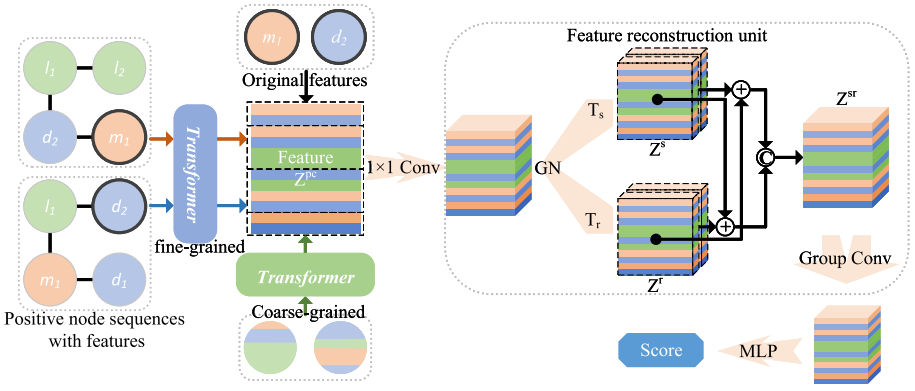

Fig. 2 Framework of feature reconstruction convolution and association prediction

the model does not rely on predefined miRNA–disease associations as input. Subsequently, target miRNA–disease nodes are introduced, enabling the model to incorporate information from both nodes. This incorporation further enhances the feature representations and improves the predictive capability of the model.

Dataset

In this work, the experimental data is extracted from the Human MicroRNA Disease Database (HMDD v4.0) [33], which contains 1245 miRNAs, 2077 diseases, miRNA functional similarities, and a total of 23,337 miRNA-disease associations. The semantic similarities between diseases are calculated according to the Medical Subject Headings (MeSH) [34]. The disease-lncRNA associations and miRNA-lncRNA interactions are obtained from the lncRNASNP data, Yang et al. [35] comprising 1438 miRNA–lncRNA interactions, 320 lncRNA–disease associations, and a total of 557 lncRNAs.

Contrastive learning of multi-granularity transformers

Construction of miRNA-disease-lncRNA graph

First, A graph $G = (V, E)$ is constructed to represent associations, interactions, and similarities among miRNAs, diseases, and lncRNAs. The node set $V = \{V^m \cup V^d \cup V^l\}$ consists of the miRNA node set V^m, disease node set V^d, and lncRNA node set V^l. The node pair $v_i, v_j \in V$ is connected by edge $e_{ij} \in E$. In this graph, O represents the association or interaction matrix and I is used to represent the similarity matrix. The correlation interaction matrix O is defined as,

$$O = \begin{cases} O^{\mathrm{md}} \in \mathbb{R}^{N_{\mathrm{m}} \times N_{\mathrm{d}}} \\ O^{\mathrm{ml}} \in \mathbb{R}^{N_{\mathrm{m}} \times N_{\mathrm{l}}} \\ O^{\mathrm{dl}} \in \mathbb{R}^{N_{\mathrm{d}} \times N_{\mathrm{l}}} \end{cases} \tag{1}$$

where N_d, N_m, and N_l denote the number of diseases, miRNAs, and lncRNAs, respectively within the dataset. O^{ml}, O^{md}, and O^{dl} represent the miRNA-lncRNA interaction matrix, miRNA-disease association matrix, and disease-lncRNA association matrix, respectively. For $m_i \in V^m (l_i \in V^l)$, $d_j \in V^d$, when O_{ij}^{md} $(O_{ij}^{\mathrm{dl}}) = 1$, this indicates that $m_i(l_i)$ has a known association with d_j. When O_{ij}^{md} $(O_{ij}^{\mathrm{dl}}) = 0$, which indicates $m_i(l_i)$ and d_j are not known to be associated. For $m_i \in V^m$, if $O_{ij}^{\mathrm{ml}} = 1$, this indicates that l_j and m_i have an identified interaction relationship, otherwise, $O_{ij}^{\mathrm{ml}} = 0$. The similarity matrix I is defined as,

$$I = \begin{cases} I^{\mathrm{m}} \in \mathbb{R}^{N_{\mathrm{m}} \times N_{\mathrm{m}}} \\ I^{\mathrm{d}} \in \mathbb{R}^{N_{\mathrm{d}} \times N_{\mathrm{d}}} \\ I^{\mathrm{l}} \in \mathbb{R}^{N_{\mathrm{l}} \times N_{\mathrm{l}}} \end{cases} \tag{2}$$

Among them, I^{d}, I^{m}, and I^{l} represent the similarity matrices for diseases, miRNAs, and lncRNAs, respectively. The values in the similarity matrix range from 0 to 1, reflecting the similarity between two nodes of the same type. A higher score indicates greater similarity. The semantic similarity of diseases is derived from the directed acyclic graph of diseases introduced by Wang et al. [36]. The functional similarity between miRNAs and lncRNAs is calculated using the method proposed by Wang et al. [36]. and Chen et al. [37], which are based on the biological premise that if miRNAs (lncRNAs) are associated with more diseases, their similarity is higher.

Given that the similarity matrix I and the correlation interaction matrix O, the attribute matrix Z and the adjacency matrix A of the graph G are equivalent, $A = Z \in \mathbb{R}^{(N_m+N_d+N_l)\times(N_m+N_d+N_l)}$ is defined as,

$$A = Z = \begin{bmatrix} I^m & O^{md} & O^{ml} \\ O^{md^T} & I^d & O^{dl} \\ O^{ml^T} & O^{dl^T} & I^l \end{bmatrix} \tag{3}$$

where ^{T} represents the transpose operation. The i-th row of the matrix Z, denote as Z_i, contains the association, interaction, and similarity between the node $v_i \in V$ and all diseases, miRNAs, and lncRNAs, which could be regarded as the attribute embedding of v_i. $\{Z_i | 0 \leqslant i < N_m\}$ represents the attribute embedding set of all miRNAs, $\{Z_i | N_m \leqslant i < N_m + N_d\}$ and $\{Z_i | N_m + N_d \leqslant i < N_m + N_d + N_l\}$ contain all the attribute embeddings for all diseases and lncRNAs, respectively.

Strongly correlated node sequence based on random walk

To obtain the surrounding key nodes related to the target disease node d_j and the target miRNA node m_i as features of contrastive learning, a random walk [38] is performed on the graph *G* starting from the target node. The random walk process is as follows: assume that the current random walk starts from node v_i and crosses edge (v_i, v_j) to reach node v_j. Now the walk needs to jump from node v_j to the next node v_k. The jump probability $\pi_{v_j v_k}$ is expressed as,

$$\pi_{v_j v_k} = \begin{cases} \frac{w_{v_j v_k}}{\mathrm{p}} & \text{if } d_{v_i v_k} = 0 \\ w_{v_j v_k} & \text{if } d_{v_i v_k} = 1 \\ \frac{w_{v_j v_k}}{\mathrm{q}} & \text{if } d_{v_i v_k} = 2 \end{cases} \tag{4}$$

where $w_{v_j v_k}$ is the edge weight between the vertices v_j and v_k. The parameters *p* and *q* jointly control the tendency of the random walk. *p* controls the probability of revisiting the vertices visited in the previous step, and *q* controls the direction of exploration, whether it is an inward or an outward search. $d_{v_i v_k}$ represents the shortest path distance between nodes v_i and v_k. When $d_{v_i v_k}$=0, it means that returning to node v_i; when $d_{v_i v_k} = 1$, it means that v_i and v_k are directly connected; $d_{v_i v_k} = 2$ means that the shortest path between v_i and v_k is 2.

When obtaining the sampling features for the target node pair, the features are divided into positive and negative samples according to the requirements of contrastive learning. The positive sampling features are obtained through random walks on the graph *G*. Taking the extraction of the target miRNA node m_i features as an example: starting from m_i, *n* random walks are performed, each of length *h* steps, on the graph *G*, resulting in *n* random walk sequences. Among them the *t*-th random walk sequence is denoted as $L_t = \{v_t^1, v_t^2, \ldots, v_t^h\}$. Next, the frequency of occurrence of the *i*-th node across all sequences is counted. The node that occurs mostly frequently is considered the strongly correlated node $v_{max}^i = max\{v_1^i, v_2^i, ..., v_n^i\}$ for the target node. Finally, according to the attribute matrix *Z*, the sampling features of the strongly related node sequence $L_{m_i} = \{v_{max}^1, v_{max}^2, \ldots, v_{max}^h\}$ are concatenated,

$$Z_{m_i}^+ = [Z_{v_{max}^1} || Z_{v_{max}^2} || \ldots || Z_{v_{max}^h}] \tag{5}$$

where || represents the concatenation operation, $Z_{m_i}^+$ represents the strongly correlated positive sampling feature of node m_i, and $Z_{m_i,1}^+$ means the features of node m_i in the first row of the positive sampling features $Z_{m_i}^+$. The same method can be applied to obtain the strongly correlated positive sampling feature $Z_{d_j}^+$ for the disease node d_j.

Negative sampling features are constructed to further enhance the features of target nodes and their strongly related nodes through contrastive learning. The negative node sequences are obtained by performing random walks on the target node within the negative graph G^-. For the graph G^-, we consider edges that are not present in the original graph *G* as it's edges, and then exclude the edges that are part of the positive sampling sequences to get the edge set of graph G^-. The adjacency matrix A^- corresponding to the negative graph G^- is calculated as follows,

$$A^- = (1 - A) \odot (1 - \overline{A}^+) \tag{6}$$

where $\overline{A}^{+}$ is the adjacency matrix composed of the edges in the positive sampling sequences, and $\odot$ represents the dot product operation. Similar to the way of obtaining positive sampling features, the negative sampling features $Z_{m_i}^{-}$ of the miRNA node m_i and the negative sampling features $Z_{d_j}^{-}$ of the disease node d_j are derived.

Multi-granularity transformer

Since miRNAs with similar functions have a higher likelihood of interacting with similar lncRNAs and participate in the same disease processes, there is a mutual dependence between miRNA, disease, and lncRNA nodes. Because transformer [39] can capture the mutual dependence between nodes effectively, we designed a multi-granularity transformer module. This module refines and fuses miRNA, disease, and lncRNA node features from both fine-grained and coarse-grained perspectives. The process of multi-granularity transformer contrastive learning for m_i node is provided in Algorithm 1.

Input: Grpah G, attribute matrix Z, node m_i, number of random walks n, walk length h.
Output: The fine-grained positive (negative) sampling feature of node m_i, $\hat{Z}'^{+}_{m_i}$ ($\hat{Z}'^{-}_{m_i}$), and the coarse-grained positive (negative) sampling feature $\widetilde{Z}'^{+}_{m_i}$ ($\widetilde{Z}'^{-}_{m_i}$).

Calculate the negative graph G^{-} according to formula (6).
for g$\in \{G, G^{-}\}$ **do**
  $L \Leftarrow RandomWalk(n, h, g)$ by formula (4)
  $L_{max} \Leftarrow MaxCountNode(L, n)$
  $Z^{g}_{m_i} \Leftarrow Embedding(L_{max}, Z)$
  $\overline{Z}^{g}_{m_i} \Leftarrow Proj(Z^{g}_{m_i})$ by formula (7)
  $\hat{Z}^{g}_{m_i} \Leftarrow Transformer(Z^{g}_{m_i})$ by formula (8)
  $\widetilde{Z}^{g}_{m_i} \Leftarrow Transformer(\overline{Z}^{g}_{m_i})$ by formula (8)
  refine fine-grained positive (negative) sampling feature $\hat{Z}^{+}_{m_i}, \hat{Z}^{-}_{m_i}$ by formula (9)
  refine coarse-grained positive (negative) sampling feature $\widetilde{Z}^{+}_{m_i}, \widetilde{Z}^{-}_{m_i}$ by formula (10)
end for
return refined $\hat{Z}'^{+}_{m_i}, \hat{Z}'^{-}_{m_i}, \widetilde{Z}'^{+}_{m_i}, \widetilde{Z}'^{-}_{m_i}$

Algorithm 1 The computational flow of m_i node in multi-granularity transformer contrastive learning

Fine-grained features are analyzed from node perspectives in the walk sequence, for instance, the positive fine-grained feature of node m_i is represented as $Z^{+}_{m_i}$. Coarse-grained features are considered from the perspective of the entire walk sequence. The positive coarse-grained feature $\overline{Z}^{+}_{m_i}$ is obtained by fusing the node features of L_{m_i},

$$\overline{Z}^{+}_{m_i} = W_c Z^{+}_{m_i} + b_c \tag{7}$$

Similarly, the negative fine-grained feature matrix $Z^{-}_{m_i}$ and the negative coarse-grained feature matrix $\overline{Z}^{-}_{m_i}$ for miRNA node m_i are obtained; likewise, the positive fine-grained feature matrix $Z^{+}_{d_j}$ and the positive coarse-grained feature matrix $\overline{Z}^{+}_{d_j}$ for disease node d_j; and the negative fine-grained feature matrix $Z^{-}_{d_j}$ and the negative coarse-grained feature matrix $\overline{Z}^{-}_{d_j}$ for disease node d_j are also derived.

The transformer encoder is employed to encode the positive sampling features of miRNA at different granularities,

$$\begin{aligned} \hat{Z}^{+}_{m_i} &= Transformer(Z^{+}_{m_i}) \\ \widetilde{Z}^{+}_{m_i} &= Transformer(\overline{Z}^{+}_{m_i}) \end{aligned} \tag{8}$$

After transformer enhancement, the fine-grained positive (negative) sampling feature of node m_i is $\hat{Z}^+_{m_i}(\hat{Z}^-_{m_i})$, the coarse-grained positive (negative) sampling feature $\widetilde{Z}^+_{m_i}(\widetilde{Z}^-_{m_i})$, the fine-grained positive (negative) sampling feature of node d_j $\hat{Z}^+_{d_i}(\hat{Z}^-_{d_i})$ and with it's coarse-grained positive (negative) feature $\widetilde{Z}^+_{d_i}(\widetilde{Z}^-_{d_i})$.

Contrastive learning to enhance node features

The task is to perform contrastive learning [30, 31, 40] between the positive and negative sampling features between disease node d_j and miRNA node m_i. The InfoNCE loss function [41] is used for fine-grained features. The neural network is trained using InfoNCE to maximize the similarity of node m_i between it's positive sampling features. In the meantime, the similarity between the positive sampling feature of node m_i and the negative sampling features is minimized. This process aims to produce a more informative and discriminative representation. The formula is as follows,

$$\mathcal{L}^{m_i}_{\text{InfoNCE}} = -\log \frac{\exp(sim(\hat{Z}^+_{m_i,1}, \hat{Z}^-_{m_i,1})/\tau)}{\sum_{k=1}^{h} \exp(sim(\hat{Z}^+_{m_i,1}, \hat{Z}^-_{m_i,k})/\tau)} \tag{9}$$

where *sim*(,) represents the cosine similarity function, $\hat{Z}^+_{m_i,k}$ represents the *k*-th row of $\hat{Z}^+_{m_i}$, and τ represents the temperature parameter.

Minimizing coarse-grained similarity between positive and negative sampling features is optimal. To achieve this, cosine similarity loss is used as the coarse-grained contrastive loss,

$$\mathcal{L}^{m_i}_{c} = \frac{sim(\widetilde{Z}^+_{m_i}, \widetilde{Z}^-_{m_i}) + 1}{2} \tag{10}$$

Similarly, we can obtain the fine-grained and coarse-grained losses $\mathcal{L}^{d_j}_{\text{InfoNCE}}$ and $\mathcal{L}^{d_j}_{c}$ for the target disease d_j. Then, the fine-grained and coarse-grained losses are combined to form the final loss for the multi-granularity transformer contrastive learning module,

$$\mathcal{L}_{CL} = \sum_{m_i \in V} \sum_{d_j \in V} (\mathcal{L}^{m_i}_{\text{InfoNCE}} + \mathcal{L}^{d_j}_{\text{InfoNCE}} + \mathcal{L}^{m_i}_{c} + \mathcal{L}^{d_j}_{c}) \tag{11}$$

Feature reconstruction convolution and association prediction

Feature reconstruction convolution

The features of the node pair m_i-d_j are enhanced through multi-granularity transformer contrastive learning to obtain $\hat{Z}'^+_{m_i}$, $\hat{Z}'^+_{d_j}$, $\widetilde{Z}'^+_{m_i}$, $\widetilde{Z}'^+_{d_j}$. In order to supplement the original feature information, Z_{m_i}, and Z_{d_j} are concatenated to form the feature Z^{pc} of the node pair m_i-d_j,

$$Z^{pc} = [Z_{m_i} || Z_{d_j} || \hat{Z}'^+_{m_i} || \hat{Z}'^+_{d_j} || \widetilde{Z}'^+_{m_i} || \widetilde{Z}'^+_{d_j}] \tag{12}$$

The feature reconstruction convolutional neural network consists of a 1×1 (pointwise) convolution, a feature reconstruction unit, and a group convolution. First, in order to capture richer semantic relationships, a 1×1 (pointwise) convolution is employed to increase the number of channels to C_o, where different channels represent semantic features from different angles,

$$\hat{Z}^{pc} = Conv_{1\times1}(Z^{pc}) \tag{13}$$

where $Conv_{1\times1}(\cdot)$ represents the convolution operation. To improve the prediction performance of the model, we designed a feature reconstruction unit. Suppose the channels of $\hat{Z}^{pc}$ are divided into G_r groups, each group having $\frac{C_o}{G_r}$ channels. For each input group $\hat{Z}_k^{pc}, k = 1, \ldots, G_r$, $\widetilde{Z}_k^{pc}$ is obtained after group normalization,

$$\widetilde{Z}_k^{pc} = \gamma_k \frac{\hat{Z}_k^{pc} - \mu_k}{\sqrt{\sigma_k^2 + \epsilon}} + \beta_k \tag{14}$$

where μ_k and σ_k represent the mean and standard deviation within the *k*-th group, and ϵ is a small constant used to prevent division by zero. γ_k and β_k represent the learnable weights corresponding to the *k*-th group of channels. The value of γ_k can dynamically adjust the relative importance of the channels by scaling. As the richness of semantics increases, so does the noise, and therefore, the corresponding noise needs to be suppressed. To address this we introduce a threshold operation to filter features,

$$\begin{aligned} Z_k^s &= T_s\left(Sigmoid\left(\frac{\gamma_k}{\sum_{s=1}^{C_r}\gamma_s} \odot \widetilde{Z}_k^{pc}\right)\right) \odot \hat{Z}_k^{pc} \\ Z_k^r &= T_r\left(Sigmoid\left(\frac{\gamma_k}{\sum_{s=1}^{C_r}\gamma_s} \odot \widetilde{Z}_k^{pc}\right)\right) \odot \hat{Z}_k^{pc} \end{aligned} \tag{15}$$

where *Sigmoid* is an activation function that maps weights between 0 and 1. $T_s(\cdot)$ represents a threshold operation that sets all values greater than 0.5 to 1, retaining the important information of $\hat{Z}_k^{pc}$, while $T_r(\cdot)$ represents an operation that sets values less than 0.5 to 0, removing the unimportant information of $\hat{Z}^{pc}$. This operation creates sparse features by filtering certain regions, potentially leading to the loss of valuable information and generating noise. To address this, cross-reconstruction operations are employed to fuse features from different regions,

$$Z^{sr} = (Z_{h1}^s + Z_{h2}^r)||(Z_{h2}^s + Z_{h1}^r) \tag{16}$$

where $Z^s = Z_{h1}^s||Z_{h2}^s$, $Z^r = Z_{h1}^r||Z_{h2}^r$. Finally, group convolution $Conv_{G_r}$ is applied to further extract features and obtain the final fusion feature $\hat{Z}^{sr}$,

$$\hat{Z}^{sr} = Conv_{G_r}(Z^{sr}) \tag{17}$$

Integration and optimization

We utilize the matrix $\hat{Z}^{sr}$ to predict the association between node pair $m_i - d_j$. First, the features are processed through Adaptive Average Pooling (AAP) and Flattening operations,

$$H = F(AAP(\hat{Z}^{sr})) \tag{18}$$

The flattened features are then passed through a fully connected layer to learn the score of the association between the node pair $m_i - d_j$. This association score *y* represents the probability distribution of whether or not there is an association. The score is defined as follows,

$$y = softmax(W_p H + b_p) \tag{19}$$

where b_p and W_p represent the bias vector and weight matrix respectively. The output $y = [y^1, y^0]$, contains two elements, where y^1 records the probability of the probability of node pair $m_i - d_j$ being associated, and y^0 the probability of not being associated. The cross entropy loss function between the true label and the miRNA-disease association prediction score *y* is expressed as follows,

$$\mathcal{L} = -\sum_{i=1}^{N_t} \left[y_l \times \log\left(\mathrm{y}^1\right) + (1 - y_l) \times \log\left(1 - \left(\mathrm{y}^0\right)\right) \right] \tag{20}$$

where N_t represents the size of a group of training samples, and y_l denotes the actual label. When miRNA is associated with the disease, $y_l = 1$, and when it is not associated, $y_l = 0$. During the training process, gradient descent algorithms and backpropagation are used to optimize the model.

Experimental evaluation and analysis

Parameter settings

APMF uses a 24GB GTX 3090 for training. The hyperparameters for the multi-granularity transformer contrastive learning module are as follows: learning rate = 0.00009, weight decay rate = 0.0005, batch size = 2048, number of heads = 2, and number of layers = 1. For the feature reconstruction convolution module, the hyperparameters are learning rate = 0.00007, weight decay rate = 0.0005, batch size = 2048, and threshold = 0.5. An early stopping strategy is employed during model training, with a patience value of 3 for both the contrastive learning module (unsupervised learning) and the feature reconstruction convolution module (supervised learning). For random walks, the sequence length is set to 5, the number of random walks is set to 20, and both the tendency parameters *p* and *q* are set to 1. The 1×1 convolution increases the number of channels to 8, and the number of groups in the group convolution is set to 2 with a kernel size of 2×7.

Evaluation metrics

The area under the Receiver Operating Characteristic (ROC) curve (AUC) [42] and the area under the precision-recall (PR) curve (AUPR) [43] are used as metrics to evaluate the predictive performance of the model. Five-fold cross-validation is applied to train and validate APMF and all comparison methods. All known miRNA-disease associations are treated as positive samples, while all unobserved disease-miRNA associations are treated as negative samples. 20% of the positive samples and 20% of the negative samples are randomly chosen for the test set and the rest are subjected to a five-fold cross-validation. Four copies of each fold are utilized as the training set, and the remaining are used for validation.

Ablation experiments

Ablation experiments are conducted to assess the contributions of the key components in the model (Table 1), including multi-granularity transformer contrastive learning (MGTCL), feature reconstruction convolution (FRC), random walk module (RWM), fine-grained transformer contrastive learning (FGTCL) and coarse-grained transformer

Table 1 Ablation experimental results of our method

	MGTCL	FRC	RWM	FGTCL	CGTCL	APMF
Average AUC	0.951	0.943	0.962	0.962	0960	0.964
Average AUPR	0.469	0.462	0.476	0.477	0.470	0.491

Table 2 Prediction results for the different combination of contrastive loss function

Combination of loss function	AUC	AUPR
InfoNCE + InfoNCE	0.957	0.474
InfoNCE + KL divergence	0.963	0.485
InfoNCE loss + Squared difference	0.961	0.482
InfoNCE loss + Cosine similarity	0.964	0.491

contrastive learning (CGTCL). MGTCL means deleting the features of contrastive learning and directly using the original features as input to feature reconstruction convolution. For FRC, a multi-layer perceptron (MLP) is used in place of feature reconstruction convolution. For CGTCL and FGTCL, the corresponding features are learned through MLP. In the RWM experiment, a sequence of randomly selected neighboring nodes of the target node is used instead of the node sequence obtained by the random walk module.

The experimental results show that FRC is the most important among all modules, remove this module with both AUC and AUPR decreasing most compared to the other components, decreasing by 2. 1% and 2. 9%, respectively. This drop is attributed to the feature reconstruction convolution(FRC), which enhances features across multiple channels, excludes redundant information and cross-reconstruction operation enriches semantic information. The AUPR and AUC of removing MGTCL decreased by 2.2% and 1.3%, respectively, from the perspective of AUC and AUPR decrease, it is the second important component only to FRC, suggesting that MGTCL has effectively learned rich and discriminative representations through the way of contrastive learning. RWM ranked third in importance, for removing this part the AUPR and AUC decreased by 1.5% and 0.2%, respectively, demonstrating that RWM helps the model select strongly related nodes, thus providing more accurate graph nodes for contrastive learning. The performance decline of removing FGTCL and CGTCL are comparable in both AUC and AUPR and have a minimum drop compared with components mentioned above, which also indicates that the FGTCL and CGTCL complement the effect of the model.

In order to explore the impact of using different contrastive loss functions on the experimental results, an experiment is conducted where the cosine similarity loss is replaced with InfoNCE loss, KL divergence loss, and Squared difference loss. From the experimental results in Table 2, it is observed that when a combination of InfoNCE loss and cosine similarity loss is selected, the AUC and AUPR are 0.7% and 1.7% higher than those obtained using a single InfoNCE loss function. The reason for the experimental results may be that using InfoNCE and Cosine similarity contrast losses can better enhance the differences in contrast graphs, thereby helping the model learn more discriminative and robust feature representations.

Comparison with other methods

Seven advanced methods for predicting miRNA-disease associations are compared with APMF (Table 3), including AMHMDA [27], MDformer [32], MSGCL [30], DNI-MDCAP

Table 3 Comparison with other methods

Methods	AUC	AUPR
AMHMDA	0.940	0.421
MDformer	0.927	0.443
MSGCL	0.951	0.338
DNI-MDCAP	0.922	0.181
GCLMTP-MLP	0.952	0.412
MAGCN	0.950	0.433
MGCNSS	0.962	0.424
APMF(Ours)	0.964	0.491

[17], GCLMTP-MLP [31], MAGCN [28], and MGCNSS [29]. All the methods compared are executed using their optimal parameters reported in their respective papers, and they all used the same dataset partitioning as APMF in cross-validation. These comparison methods are briefly described below.

- AMHMDA: It uses graph-based methods and similarity networks to predict miRNA-disease associations. It combines information from different sources, applies a fusion mechanism to extract useful features, and introduces extra nodes to improve connections. Finally, it integrates these features for prediction.
- MDformer: A transformer-based model for miRNA-disease association prediction. It integrates multi-source biological features, encodes them using a transformer-based feature encoder, and employs a MLP for prediction.
- MSGCL: Contrastive learning model for miRNA-disease association prediction. It optimizes graph structure by contrastive loss between a learner view with an anchor view to enhance prediction accuracy.
- DNI-MDCAP: It add additional miRNA similarity metrics, using embedding learning-based techniques, and applying a semi-supervised learning strategy to improve prediction accuracy.
- GCLMTP-MLP: A graph contrastive learning model based on GCN to extracts embeddings, and employs MLP for prediction.
- MAGCN: It uses graph convolution networks with an attention mechanism and CNN to predicting miRNA-disease associations.
- MGCNSS: It constructs a graph integrating miRNA and disease similarities, uses graph convolution to capture meta-path relations, and applies a negative sample selection strategy to improve prediction accuracy.

The average AUC and AUPR of all seven methods are shown in the Table 3 above. APMF achieved the best average AUC of 0.964, which is 2.4% higher than AMHMDA, 3.7% higher than MDformer, 1.3% higher than MSGCL, 4.2% higher than DNI-MDCAP, 1.2% higher than GCLMTP-MLP, 1.4% higher than MAGCN, and 0.2% higher than MGCNSS. The average AUPR of APMF is 0.491, which is 7.0%, 4.8%, 15.3%, 31.0%, 7.9%, 5.8%, and 6.7% higher than AMHMDA, MDformer, MSGCL, DNI-MDCAP, GCLMTP-MLP, MAGCN and MGCNSS respectively. For each prediction approach, the average five-fold AUC (AUPR) results across 2077 diseases are assessed. To compare any two methods, a paired Wilcoxon test is performed using 2077 sets of average five-fold AUC (AUPR) values. The statistical analysis present in Table 4 demonstrates that our model consistently outperforms all comparison methods, with a statistically significant advantage (p-value < 0.05), confirming robustness and reliability of APMF.

Table 4 Paired Wilcoxon test results for the comparative analysis between APMF and all other models

Methods	AUC	AUPR
AMHMDA	4.2779e−04	1.7539e−03
MDformer	6.4442e−06	2.0576e−04
MSGCL	1.6153e−07	2.1321e−04
DNI-MDCAP	9.5266e−09	2.8991e−08
GCLMTP-MLP	5.5370e−09	3.4276e−07
MAGCN	2.5593e−06	3.4072e−04
MGCNSS	2.1695e−07	2.7873e−04

DNI-MDCAP is a machine learning model demonstrated poor adaptability to dataset changes and consequently achieved the worst results. In comparison, AMHMDA and MDformer performed better due to their ability to mine potential unknown relationships through hypergraph and transformer, respectively. However, the features they learned were limited by the original input features. MSGCL and GCLMTP-MLP improved the results based on graph contrastive learning, but the graph contrastive learning based on GCN encoder is insufficient for the deep information mining of miRNA-disease associations. MAGCN combined GCN with multichannel attention mechanism and a CNN encoder for feature learning and achieved relatively good prediction results. MGCNSS used multi-layer graph convolution to autonomously identify meta-path relationships of varying lengths within heterogeneous networks and adopted a long-distance-based sample selection strategy, which also achieved relatively good results. The key improvement of our method over these approaches lies in the use of multi-granularity contrastive learning, which enables the model to learn rich and discriminative representations, while feature reconstruction convolution enhances important features and eliminates redundant ones.

Parameter analysis

An analysis of six key hyperparameters is conducted: the learning rates for unsupervised and supervised learning, batch size, sequence length and number of repetitions for random walks, the number of transformer heads, and the threshold for feature reconstruction convolution. For each hyperparameter, different values were explored, and a 5-fold cross-validation experiment was performed.

Learning rate: The learning rates of the multi-granularity transformer contrastive learning (MGTCL) and the feature reconstruction convolution (FRC) are analyzed respectively. For MGTCL, values of 0.00001, 0.00005, 0.00009, 0.0001, 0.0005, 0.001, and 0.01 are selected for analysis. As shown in Fig. 3a and b, when the learning rate is 0.00009, APMF attains the best performance across all metrics. When the learning rate increases from 0.00001 to 0.00009, the performance of APMF improves. When the learning rate ranged between 0.00009 and 0.001, these indicators decrease. For FRC, we tested learning rates of 0.00001, 0.00005, 0.00007, 0.0001, 0.0005, 0.001 and 0.01. The best performance for FRC is observed with a learning rate of 0.00007.

Batch size: To investigate the effect of batch size, values of 512, 1,024, 2,048, and 4,096 are tested. Figure 3c and d shows that the best performance is achieved with a batch size of 2,048, which is selected as the optimal batch size for APMF.

Random walk sequence length: Random walk sequence lengths of 1, 3, 5, 7, and 9 are evaluated, and it is observed that a length of 5 yields the optimal prediction performance,

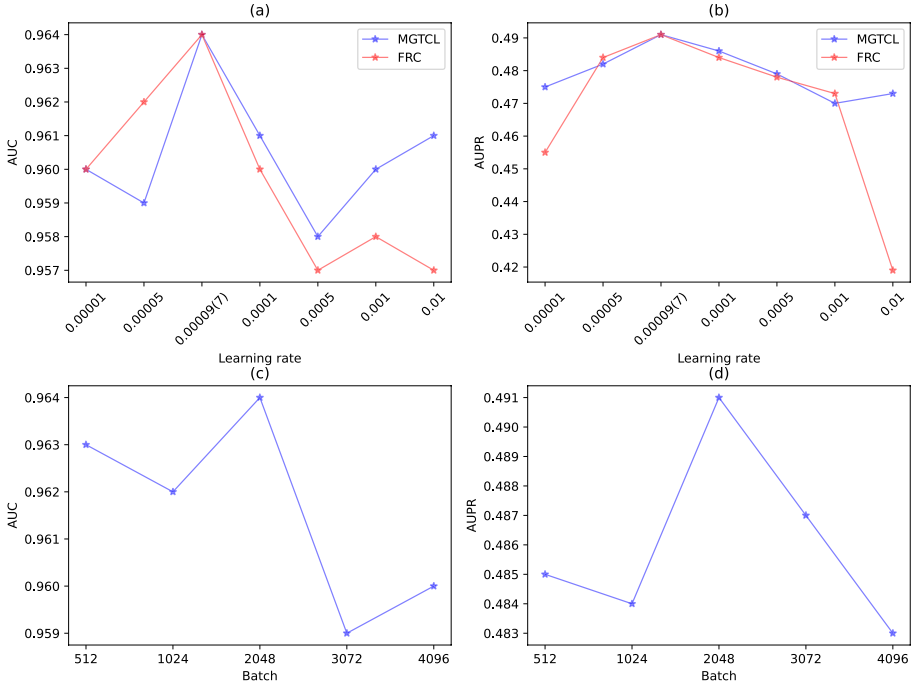

Fig. 3 The parameter sensitivity analysis with different learning rates and batchsize

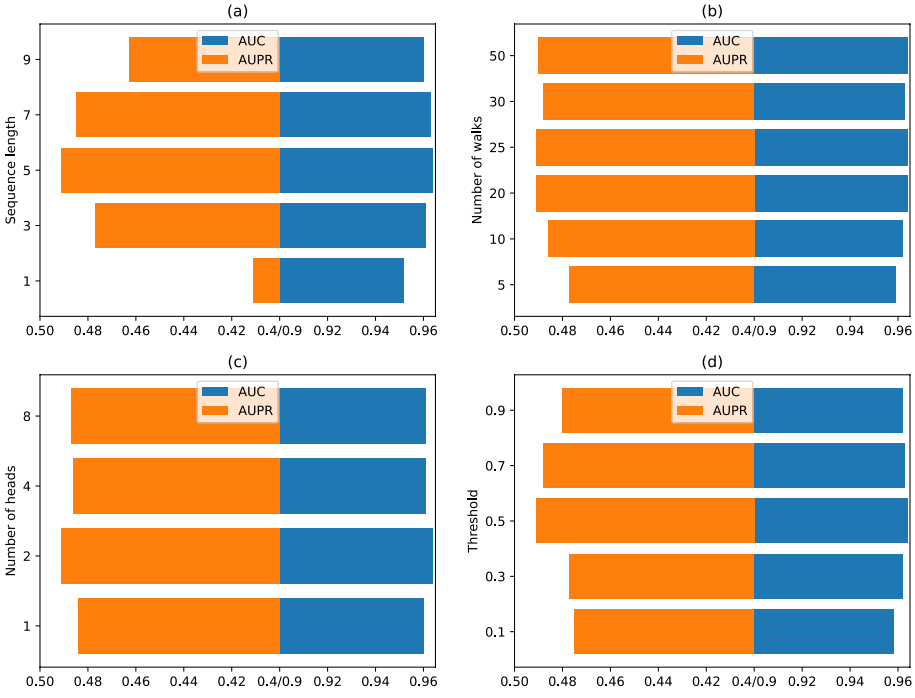

Fig. 4 The parameter sensitivity analysis with random walk sequence length, number of walks, number of transformer heads and threshold of feature reconstruction convolution

achieving an AUC of 0.964 and an AUPR of 0.491 (Fig. 4a). In terms of AUC, this configuration outperforms lengths of 1, 3, 7, and 9 by 1.2%, 0.3%, 0.1%, and 0.4%, respectively. Similarly, AUPR improve by 8.0%, 1.4%, 0.6%, and 2.8% compared to the same lengths. Sequence lengths that are too short fail to capture sufficient neighborhood information,

whereas excessively long sequences introduce low-relevance nodes, thereby reducing representation quality.

Number of random walk repetitions: Repetition counts of 5, 10, 20, 25, 30, and 50 are tested (Fig. 4b). The best performance is observed at 20 or 25 repetitions, both yielding an AUC of 0.964, which exceeds that of 5, 10, and 30 repetitions by 0.5%, 0.2%, and 0.1%, respectively, and is comparable to that of 50 repetitions. The corresponding AUPR of 0.491 represents gains of 1.4%, 0.5%, and 0.3% over 5, 10, and 30 repetitions, and a slight improvement of 0.1% over 50. As model performance plateaus beyond 20 repetitions, 20 is adopted as the optimal setting to balance accuracy and computational efficiency.

Number of transformer heads: We investigate the effect of varying the number of transformer heads, setting it to 1, 2, 4, and 8. As shown in Fig. 4c, the model achieves its best performance when the number of heads is set to 2, yielding the highest AUC and AUPR. Specifically, the AUC increases by 0.4% and 0.3% compared to settings of 1 and 4/8 heads, respectively, while the AUPR improves by 0.7%, 0.5%, and 0.4% over the same configurations. These results demonstrate that selecting an appropriate number of heads enables the model to better capture the complex relationships between miRNAs and diseases from multiple perspectives, thereby enhancing its representation and predictive capabilities.

Threshold of feature reconstruction convolution: We further analyze the impact of the threshold setting in the feature reconstruction convolution module by evaluating threshold values of 0.1, 0.3, 0.5, 0.7, and 0.9. As shown in Fig. 4d, the model achieves the best performance when the threshold is set to 0.5, with an AUC of 0.964 and an AUPR of 0.491. Compared with the other threshold settings, this configuration improves the AUC by 0.6%, 0.2%, 0.1%, and 0.2%, and the AUPR by 1.6%, 1.4%, 0.3%, and 1.1%, respectively. Lower thresholds may fail to filter out sufficient redundant information–sometimes even amplifying it–while higher thresholds may remove important features, both of which negatively impact model performance. We further explore an adaptive threshold strategy, which dynamically adjusts the threshold during training based on data characteristics and model state. This approach achieves an AUC of 0.965 and an AUPR of 0.498, representing improvements of 0.1% and 0.7% over the fixed threshold of 0.5. The adaptive mechanism continuously adjusts the threshold during the training process, allowing the model to better adapt to the complexity and diversity of the data, thereby improving the accuracy and robustness of the prediction.

Case studies

To further showcase APMF's capability in identifying potential miRNA-disease associations, a case analysis is conducted on the model. The model was trained using all positive samples along with an equal number of negative samples. The miRNAs for the disease are then ranked in descending order according to the association score. Case analyses of prostate neoplasms, esophageal neoplasms, and breast neoplasms are performed, and the top 15 candidate miRNAs are listed in Tables 5, 6, and 7, respectively. To verify the prediction results, the HMDD v4.0 [33] and dbDEMC [44] databases were queried. The HMDD database provides experimentally supported human miRNA and disease associations from published papers, while the dbDEMC database includes differentially expressed miRNAs in human cancers, comprising 36 types of cancer and 2,224 miRNAs.

Table 5 The top 15 Prostatic neoplasm-related miRNA candidates

Rank	miRNA name	Description	Rank	miRNA name	Description
1	hsa-mir-214	L^a, L^b	9	hsa-mir-28	L^a, L^b
2	hsa-mir-122	L^a, L^b	10	hsa-mir-223	L^a, L^b
3	hsa-mir-20a	L^a, L^b	11	hsa-mir-143	L^a, L^b
4	hsa-mir-15a	L^a, L^b	12	hsa-mir-146a	L^a, L^b
5	hsa-mir-21	L^a, L^b	13	hsa-mir-23a	L^a, L^b
6	hsa-mir-182	L^a, L^b	14	hsa-mir-142	L^a, L^b
7	hsa-mir-145	L^a, L^b	15	hsa-mir-18a	L^a, L^b
8	hsa-mir-17	L^a, L^b			

L^a:HMDD v4.0; L^b: dbDEMC

Table 6 The top 15 Esophageal neoplasm-related miRNA candidates

Rank	miRNA name	Description	Rank	miRNA name	Description
1	hsa-mir-21	L^a, L^b	9	hsa-mir-23a	L^b
2	hsa-mir-20a	L^a, L^b	10	hsa-mir-18a	L^a, L^b
3	hsa-mir-122	L^a, L^b	11	hsa-mir-146a	L^a, L^b
4	hsa-mir-214	L^a, L^b	12	hsa-mir-142	L^a, L^b
5	hsa-mir-182	L^b	13	hsa-mir-223	L^a, L^b
6	hsa-mir-15a	L^a, L^b	14	hsa-mir-143	L^a, L^b
7	hsa-mir-17	L^a, L^b	15	hsa-mir-200c	L^a, L^b
8	hsa-mir-145	L^a, L^b			

L^a:HMDD v4.0; L^b: dbDEMC

Table 7 The top 15 Breast neoplasm-related miRNA candidates

Rank	miRNA name	Description	Rank	miRNA name	Description
1	hsa-mir-20a	L^a, L^b	9	hsa-mir-18a	L^a, L^b
2	hsa-mir-15a	L^a, L^b	10	hsa-mir-143	L^a, L^b
3	hsa-mir-214	L^a, L^b	11	hsa-mir-145	L^a, L^b
4	hsa-mir-122	L^a, L^b	12	hsa-mir-155	L^a, L^b
5	hsa-mir-223	L^a, L^b	13	hsa-mir-28	L^a, L^b
6	hsa-mir-21	L^a, L^b	14	hsa-mir-146a	L^a, L^b
7	hsa-mir-182	L^a, L^b	15	hsa-mir-23a	L^a, L^b
8	hsa-mir-17	L^a, L^b			

L^a:HMDD v4.0; L^b: dbDEMC

Among the top 15 candidate miRNAs associated with Prostatic Neoplasms listed in Table 5, all 15 are included in both HMDD v4.0 and dbDEMC, confirming their relevance to Prostatic Neoplasms. For Esophageal Neoplasms, 13 of the 15 candidate miRNAs listed in Table 6 are included in HMDD v4.0, and all 15 are included in dbDEMC. Table 7 lists the top 15 candidate miRNAs associated with Breast Neoplasms, with each miRNA confirmed by both HMDD v4.0 and dbDEMC. To demonstrate the model's predictive ability for unknown associations, we further analyzed experiment data in Table 6, the association between esophageal neoplasms and hsa-mir-182 (hsa-mir-23a) does not exist in the HMDD v4.0 database but is confirmed by dbDEMC database. Since the training model data does not use the dbDEMC database information, the results are confirmed by the dbDEMC database, which further demonstrates the predictive ability of the model for unknown associations. And numerous studies have shown that changes in miRNA expression levels are closely related to the occurrence of many diseases. By modulating multiple oncogenes, miR-145 regulates various cellular processes in Prostatic neoplasm, which play a role in the transition from localized to metastatic disease

[45]. Studies have found that in the lesion tissues of patients with esophageal neoplasm, the expression changes of hsa-mir-145 and hsa-mir-143 are the most significant. And Hsa-mir-21 also has been shown to have excellent sensitivity and specificity for Esophageal neoplasm. By targeting synuclein-γ, miRNA-15a causes cell cycle arrest and promotes apoptosis in breast cancer cells [46].

Predicting novel miRNA-disease associations

After evaluating APMF's prediction performance through cross-validation and case analysis, we used the model to predict candidate miRNAs associated with each disease. Supplementary Table ST1 lists the top 30 candidate miRNAs predicted by APMF for all diseases.

Conclusion

We proposed a disease-related miRNA prediction method. A random walk strategy is used to capture the strongly related nodes for the target node within a graph constructed from multi-source data. The multi-granularity transformer module refines and fuses the positive and negative sampling features of miRNA, disease and lncRNA nodes through fine-grained refinement and coarse-grained fusion, respectively. Discriminative features are extracted by training the optimized model with contrastive learning loss. The feature reconstruction convolution obtained semantic features from various perspectives by upscaling convolution channels, adjusting prediction features based on channel importance weights, setting thresholds to filter out noise, and performing cross-reconstruction to fuse different channel features. Five-fold cross-validation results demonstrated that APMF outperforms the seven state-of-the-art in terms of AUPR and AUC. Ablation experiments further validated the effectiveness of APMF's key innovations. The prediction ability of APMF for miRNAs and diseases is further confirmed through case studies on three diseases.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-026-06396-1.

Supplementary file 1.

Supplementary file 2.

Author contributions

Ping Xuan: Designed the method and participated in manuscript writing. Zhicheng Guo: Designed the experiments and participated in manuscript writing. Siyuan Lu: Participated in experiment design and manuscript writing. Hui Cui: Participated in experiment design. Jian Ding: Participated in manuscript writing.Tiangang Zhang: Participated in method design and manuscript writing.

Funding

This work is supported by Natural Science Foundation of Heilongjiang Province (LH2023F044). Natural Science Foundation of China (62372282, 62562023, 62172143). Guangdong Basic and Applied Basic Research Foundation (2024A1515010176). STU Scientific Research Initiation Grant (NTF22032).

Data availability

The datasets are extracted from the HMDD32 and lncRNASNPdata34 databases, and they are provided in https://github.com/AntarcticLu/MicRNA_Disease_Dataset. The source code for training and testing our prediction model are freely available at https://github.com/pingxuan-hlju/APMF.

Declarations

Ethics approval and consent to participate

Not Applicable.

Consent for publication
Not Applicable.

Competing interests
The authors declare no Conflict of interest.

Received: 6 February 2025 / Accepted: 3 February 2026
Published online: 29 April 2026

References

1. Ambros V. The functions of animal microRNAs. Nature. 2004;431:350–5.
2. Huang H-Y, Lin Y-C-D, Cui S, Huang Y, Tang Y, Xu J, et al. miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. 2022;50:D222–30.
3. Van Meter EN, Onyango JA, Teske KA. A review of currently identified small molecule modulators of microRNA function. Eur J Med Chem. 2020;188:112008.
4. Chen X, Xie D, Zhao Q, You Z-H. MicroRNAs and complex diseases: from experimental results to computational models. Brief Bioinform. 2019;20:515–39.
5. Zhao D, Wu K, Sharma S, Xing F, Wu S-Y, Tyagi A, et al. Exosomal miR-1304-3p promotes breast cancer progression in African Americans by activating cancer-associated adipocytes. Nat Commun. 2022;13:7734.
6. Metzinger-Le Meuth V, Metzinger L. miR-223 and other miRNA's evaluation in chronic kidney disease: innovative biomarkers and therapeutic tools. Non-coding RNA Res. 2019;4:30–5.
7. Zhang Y, Cai Y, Zhang H, Zhang J, Zeng Y, Fan C, et al. Brown adipose tissue transplantation ameliorates diabetic nephropathy through the miR-30b pathway by targeting Runx1. Metabolism. 2021;125:154916.
8. Wang W-X, Rajeev BW, Stromberg AJ, Ren N, Tang G, Huang Q, et al. The expression of microRNA miR-107 decreases early in Alzheimer's disease and may accelerate disease progression through regulation of β-site amyloid precursor protein-cleaving enzyme 1. J Neurosci. 2008;28:1213–23.
9. Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, et al. An analysis of human MicroRNA and disease associations. PLoS ONE. 2008;3(10):e3420.
10. Gu C, Liao B, Li X, Li K. Network consistency projection for human miRNA-disease associations inference. Sci Rep. 2016;6:36054.
11. Jiang Q, Hao Y, Wang G, Juan L, Zhang T, Teng M, et al. Prioritization of disease microRNAs through a human phenome-microRNAome network. BMC Syst Biol. 2010;4:1–9.
12. Zhang Y, Shen J, Zhang R, Zhao Z. Network representation learning via improved random walk with restart. Knowl-Based Syst. 2023;263:110255.
13. Baptista A, Brière G, Baudot A. Random walk with restart on multilayer networks: from node prioritisation to supervised link prediction and beyond. BMC Bioinform. 2024;25:70.
14. Li A, Deng Y, Tan Y, Chen M. A novel miRNA-disease association prediction model using dual random walk with restart and space projection federated method. PLoS ONE. 2021;16:e0252971.
15. Xie X, Wang Y, He K, Sheng N. Predicting miRNA-disease associations based on PPMI and attention network. BMC Bioinform. 2023;24:113.
16. Xuan P, Han K, Guo M, Guo Y, Li J, Ding J, et al. Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. PLoS ONE. 2013;8:e70204.
17. Han Y, Zhou Q, Liu L, Li J, Zhou Y. DNI-MDCAP: improvement of causal MiRNA-disease association prediction based on deep network imputation. BMC Bioinform. 2024;25:22.
18. Liu W, Lin H, Huang L, Peng L, Tang T, Zhao Q, Yang L. Identification of miRNA–disease associations via deep forest ensemble learning based on autoencoder. Briefings in Bioinform. 2022;23:bbac104
19. Zhang Z-C, Zhang X-F, Wu M, Ou-Yang L, Zhao X-M, Li X-L. A graph regularized generalized matrix factorization model for predicting links in biomedical bipartite networks. Bioinformatics. 2020;36:3474–81.
20. Pian C, Mao S, Zhang G, Du J, Li F, Leung SY, et al. Discovering cancer-related miRNAs from miRNA-target interactions by support vector machines. Mol Ther-Nucl Acids. 2020;19:1423–33.
21. Xuan P, Han K, Guo Y, Li J, Li X, Zhong Y, et al. Prediction of potential disease-associated microRNAs based on random walk. Bioinformatics. 2015;31(11):1805–15.
22. Lu P, Jiang J. AE-RW: predicting miRNA-disease associations by using autoencoder and random walk on miRNA-gene-disease heterogeneous network. Comput Biol Chem. 2024;110:108085.
23. Yao H-B, Hou Z-J, Zhang W-G, Li H, Chen Y. Prediction of MicroRNA-disease potential association based on sparse learning and multilayer random walks. J Comput Biol. 2024;31:241–56.
24. Xuan P, Shen T, Wang X, Zhang T, Zhang W. Inferring disease-associated microRNAs in heterogeneous networks with node attributes. IEEE/ACM Trans Comput Biol Bioinform. 2020;17:1019–31.
25. Chen X, Sun L-G, Zhao Y. NCMCMDA: miRNA-disease association prediction through neighborhood constraint matrix completion. Brief Bioinform. 2021;22:485–96.
26. Peng L, Yang C, Huang L, Chen X, Fu X, Liu W. RNMFLP: predicting circRNA–disease associations based on robust nonnegative matrix factorization and label propagation. Brief Bioinform. 2022;23:bbac155.
27. Ning Q, Zhao Y, Gao J, Chen C, Li X, Li T, et al. AMHMDA: attention aware multi-view similarity networks and hypergraph learning for miRNA-disease associations identification. Brief Bioinform. 2023;24:bbad094.
28. Wang W, Chen H. Predicting miRNA-disease associations based on lncRNA-miRNA interactions and graph convolution networks. Brief Bioinform. 2023;24:bbac495.
29. Tian Z, Han C, Xu L, Teng Z, Song W. MGCNSS: miRNA-disease association prediction with multi-layer graph convolution and distance-based negative sample selection strategy. Brief Bioinform. 2024;25:bbae168.
30. Ruan X, Jiang C, Lin P, Lin Y, Liu J, Huang S, et al. MSGCL: inferring miRNA-disease associations based on multi-view self-supervised graph structure contrastive learning. Brief Bioinform. 2023;24:bbac623.

31. Sheng N, Wang Y, Huang L, Gao L, Cao Y, Xie X, et al. Multi-task prediction-based graph contrastive learning for inferring the relationship among lncRNAs, miRNAs and diseases. Brief Bioinform. 2023;24:bbad276.
32. Dong B, Sun W, Xu D, Wang G, Zhang T. MDformer: a transformer-based method for predicting miRNA-Disease associations using multi-source feature fusion and maximal meta-path instances encoding. Comput Biol Med. 2023;167:107585.
33. Cui C, Zhong B, Fan R, Cui Q. HMDD v4. 0: a database for experimentally supported human microRNA-disease associations. Nucl Acids Res. 2024;52:1327–32.
34. Lipscomb CE. Medical subject headings (MeSH). Bull Med Libr Assoc. 2000;88:265.
35. Yang Y, Wang D, Miao Y-R, Wu X, Luo H, Cao W, et al. lncRNASNP v3: an updated database for functional variants in long non-coding RNAs. Nucl Acids Res. 2023;51:D192–8.
36. Wang D, Wang J, Lu M, Song F, Cui Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. Bioinformatics. 2010;26:1644–50.
37. Chen X, Clarence Yan C, Luo C, Ji W, Zhang Y, Dai Q. Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity. Sci Rep. 2015;5:11338.
38. Grover A, Leskovec J. node2vec: scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. 2016;855–864
39. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst. 2017;30
40. Yu J, Xia X, Chen T, Cui L, Hung NQV, Yin H. XSimGCL: towards extremely simple graph contrastive learning for recommendation. IEEE Trans Knowl Data Eng. 2023;36:913–26.
41. Oord Avd, Li Y, Vinyals O. Representation learning with contrastive predictive coding. Preprint at arXiv:1807.03748 (2018)
42. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med. 2013;4:627.
43. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE. 2015;10:e0118432.
44. Xu F, Wang Y, Ling Y, Zhou C, Wang H, Teschendorff AE, et al. dbDEMC 3.0: functional exploration of differentially expressed miRNAs in cancers of human and model organisms. Genom Proteom Bioinform. 2022;20:446–54.
45. Coradduzza D, Cruciani S, Arru C, Garroni G, Pashchenko A, Jedea M, et al. Role of miRNA-145, 148, and 185 and stem cells in prostate cancer. Int J Mol Sci. 2022;23:1626.
46. Li P, Xie X-B, Chen Q, Pang G-L, Luo W, Tu J-C, et al. MiRNA-15a mediates cell cycle arrest and potentiates apoptosis in breast cancer cells by targeting synuclein-γ. Asian Pac J Cancer Prev. 2014;15:6949–54.

Publisher's Note