2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Multi-view Graph Attention Contrastive Learning for Predicting miRNA-Disease Association

Li-Juan Qiao[1], Yu-Kai Ma[1], Yu-Tian Wang[1], Shuang Liu[1], Cun-Mei Ji[1] (✉), and Chun-Hou Zheng[2]

[1] School of Cyber Science and Engineering, Qufu Normal University, Qufu, China
cunmeiji@126.com
[2] School of Artificial Intelligence, Anhui University, Hefei, China

Abstract. MicroRNAs (miRNAs) represent a vital category of endogenous non-coding RNAs that influence essential biological functions, including cell differentiation, proliferation, and apoptosis, by means of post-transcriptional processes. Recent studies have shown that atypical levels of miRNA expressions are strongly linked to the progression of intricate conditions, including cancers and neurodegenerative disorders. This research presents an innovative prediction model, MGACMDA, which combines multi-view contrastive learning and residual graph attention to overcome several limitations of existing miRNA-disease association prediction methods, such as limited robustness to data sparsity, high sensitivity to network noise, and insufficient extraction of deep topological information. We propose three data enhancement methods to construct global, local and topological views, and design a graph attention encoder with residual connection to fuse shallow topological features with deep representations through residual mechanism. Finally, a momentum-driven multi-view contrastive learning module is designed, and momentum encoder is used to maintain the global negative sample queue, which significantly improves the discrimination ability of sparse association. MGACMDA was evaluated on HMDD v2.0 and HMDD v3.2 datasets with 5 - fold cross - validation. Using F1 score, AUC, and AUPR as metrics, along with case studies, results showed that MGACMDA is an efficient and robust approach for predicting miRNA - disease associations.

Keywords: miRNA-disease association prediction, Multi-view similarity network, Contrastive learning, Graph attention network, Similarity kernel fusion, Residual attention networks

1 INTRODUCTION

MicroRNAs (miRNAs) are endogenous, non-coding RNA molecules that consist of roughly 22 nucleotides and are encoded by genes within the organism. These molecules are essential for the regulation of gene expression following transcription in both plants and animals. In humans and mammals, miRNAs play a crucial role in shaping RNA expression. Recent research has revealed the presence of miRNAs across diverse organisms, where they play roles in a variety of biological functions[1].

Following the identification of Caenorhabditis elegans lin-4[2], there has been a significant upsurge in the number of scientists delving into the functions of microRNAs (miRNAs). Through their investigations, they have uncovered that the aberrant expression of these molecules is associated with the initiation and advancement of numerous intricate human diseases. According to Lin et al. [3] a higher concentration of miR-1179 significantly impaired the ability of pancreatic cancer cells to migrate and invade. Furthermore, San et al.[4] revealed how miR-155 plays a critical role in influencing drug resistance and migration in sensitized cells via exosomal transfer. Hence, it is of utmost importance to provide support to researchers in the domain of disease studies to enable them to identify potential associations between miRNAs and diseases.

Over the past decade, the investigation into computational methods aimed at forecasting relationships between miRNAs and various diseases has become a notable field of research. These methodologies are typically classified into two distinct categories: those that rely on similarity measures and those that are founded on machine learning principles. Similarity-based approaches utilize biological networks and association graphs, indicating that miRNAs sharing analogous functions are frequently linked to diseases that display similar phenotypic characteristics. As the volume of available experimental data has grown, machine learning-based methodologies have increasingly gained prominence. These approaches combine classification algorithms with feature extraction techniques to predict associations, setting themselves apart from conventional similarity-based methods. Zheng et al.[5] introduced the MLMDA model, which integrates deep self-encoder downscaling along with a random forest classifier. Meanwhile, Chen et al.[6] addressed the challenge of only predicting binary associations by creating the RBMMMDA model, which leverages a Restricted Boltzmann Machine (RBM).

The increasing prevalence of graph neural networks has spurred researchers to explore their effectiveness in predicting the associations between microRNAs (miRNAs) and diseases[7]. As proposed by Li et al. [8] , the GAEMDA model leverages a graph neural network encoder in conjunction with a bilinear decoder to reconstitute the associations between miRNAs and diseases. Ji et al. [9] , in their research, integrated a deep autoencoder with a graph neural network encoder and also utilized a bilinear decoder for the reconstruction of miRNA-disease relationships. Additionally, Li et al. [10] employed a fully connected graph in combination with a two-layer graph convolutional network to forecast potential associations between miRNAs and diseases.

Most models extract feature embeddings for miRNA-disease associations but face issues such as network noise, loss of local structural information, and difficulty preserving deeper associations. To address these, we propose MGACMDA, a new computational model using multi-view generation and contrast learning. We fuse various similarity kernels via SKF and construct graph attributes with three views: global, local, and topology-enhanced. Feature fusion is then performed using a graph attention encoder with residual concatenation to preserve critical biosignatures. Next, dynamic multi-view contrastive learning is employed with a momentum encoder to maintain a negative sample queue, ensuring consistency across views and improving the discrimination of sparse associations. Finally, the optimized graph structure is input into a deep neural network classifier for association prediction. Specifically, the primary contributions of our study include the following components:

1. Different data enhancement methods are employed to design the global-local-noise three-view generation strategy, where the global view provides a macro-association framework, the local view complements fine-grained biomodule signals, and the noise view enhances generalization capability. This strategy improves the model's robustness to sparse data and noise, fine-tunes the capture of local functional module information, and ensures global topological integrity.
2. The design of a graph attention encoder with residual linking is employed to avoid information loss by fusing shallow topological features with deep semantic representations through a residual mechanism.
3. A multi-view contrastive Learning module is developed to construct positive-negative sample pairs, ensuring that the representation of the same node under different augmented views is as consistent as possible, while simultaneously differentiating the representations of different nodes to obtain feature embeddings that capture both local structural information and global discriminative ability.

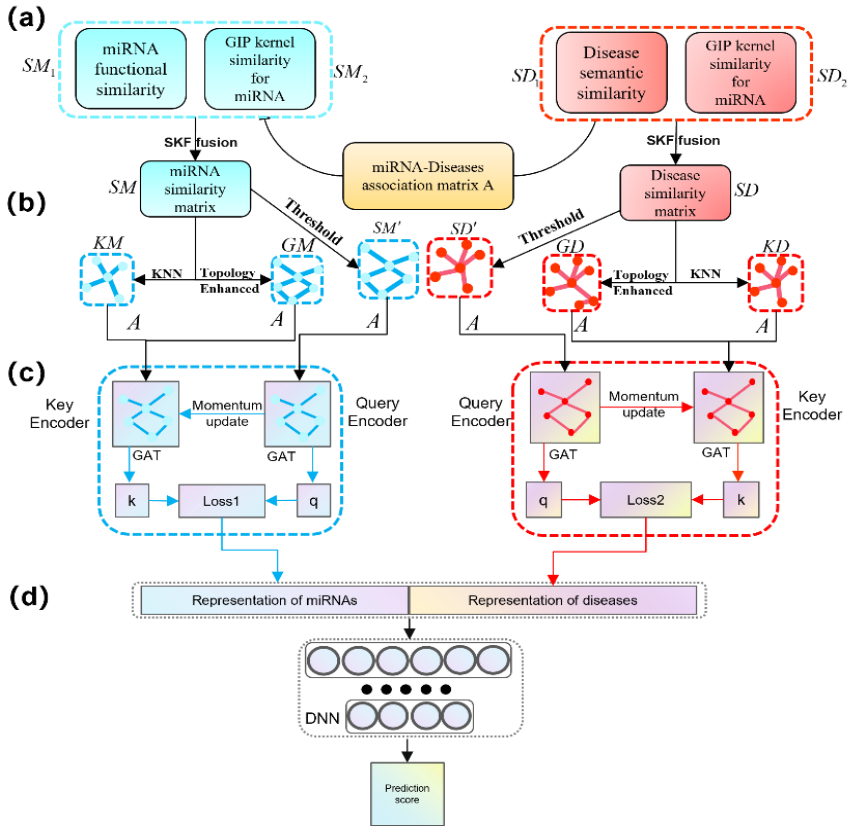

Fig. 1. Workflow of MGACMDA. **(a)** Construct the miRNA and disease similarity kernel module. **(b)** Generate three views through data augmentation. **(c)** Multi-view contrastive learning module, where the encoder is a residual graph attention encoder. **(d)** Deep neural network prediction module.

2 MATERIALS AND METHODS

2.1 An Overview of MGACMDA

In this study, we propose MGACMDA, a predictive model using graph-attention contrast learning. The model integrates multi-source similarity data through similarity kernel fusion and constructs global, local, and topological views. A graph attention encoder with residual connections extracts deep features and improves the distinction of sparse associations. A classifier is used to output prediction probabilities. The workflow of MGACMDA is illustrated in **Fig. 1**.

2.2 Disease semantic similarity

To measure the semantic similarity between two diseases, we employ their Directed Acyclic Graphs (DAGs), which are created by associating disease names with MeSH descriptors.

Let the DAG of disease d_i and disease d_j be denoted as $DAG\left(d_i\right)$ and $DAG\left(d_j\right)$, respectively, disease semantic similarity $DSS\left(d_i, d_j\right)$ can be calculated by their DAGs:

$$DSS\left(d_i, d_j\right) = \frac{\sum_{t \in \left(T_{d_i} \cap T_{d_j}\right)} \left(SV'_{d_i}\left(t\right) + SV'_{d_j}\left(t\right)\right)}{SV\left(d_i\right) + SV\left(d_j\right)} \tag{1}$$

Where $T_{d_i} \cap T_{d_j}$ denotes the collection of common ancestor nodes for d_i and d_j. $SV'_d\left(t\right)$ indicates the semantic contribution of node $t \in T_d$ to node d, while $SV\left(d\right) = \sum_{t \in T_d} SV'_d\left(t\right)$ signifies the semantic value associated with the disease. This leads to a similarity matrix SD_1 calculated pairwise for every disease, where each element, denoted as $SD_1\left(d_i, d_j\right)$, signifies the functional similarity existing between the i-th and j-th diseases.

2.3 Disease similarity measured by the GIP kernel

Since the disease semantic similarity matrix (SD) is sparse, several studies [9] have shown that reducing its sparsity improves prediction accuracy. To enrich the disease similarity information, this paper introduces the use of the Gaussian kernel function to calculate the Gaussian kernel interaction spectrum between diseases.

Should the matrix SD_2 be employed to illustrate the GIP similarity information among diseases, its dimensions will also be $nd \times nd$. The GIP similarity $SD_2\left(d_i, d_j\right)$ between disease d_i and disease d_j can be computed as indicated in (2).

$$SD_2\left(d_i, d_j\right) = \exp\left(-\rho_d \left\|K\left(d_i\right) - K\left(d_j\right)\right\|^2\right) \tag{2}$$

$$\rho_d = \rho_d' \Big/ \left(\frac{1}{nd}\sum_{i=1}^{nd}\left\|K\left(d_i\right)\right\|^2\right) \tag{3}$$

In this context, ρ_d denotes the parameter used to regulate the width of the Gaussian kernel, as outlined in equation (3), The variable nd signifies the count of diseases, and the value assigned to ρ_d' is established at 1, in alignment with prior research[11].

2.4 miRNA functional similarity

Let the set of diseases corresponding to miRNA u be $DG(u)$ and the set of diseases corresponding to miRNA v be $DG(v)$; if a disease $d_u \in DG(u)$, its similarity to $DG(v)$ can be taken to be the maximum of semantic similarity between this disease and all diseases in the set:

$$S\left(d_u, DG(v)\right) = \max_{d \in DG(v)} DSS\left(d_u, d\right) \tag{4}$$

$$S_{u \to v} = \sum_{d \in DG(u)} S\left(d, DG(v)\right) \tag{5}$$

Similarly, for $d_v \in DG(v)$, $S\left(d_v, DG(u)\right)$ can be calculated. Next, the similarity between two disease groups is defined as (5).Finally, the miRNA functional similarity between u and v is defined as (6).

$$FS(u,v) = \frac{S_{u \to v} + S_{v \to u}}{\left|DG(u)\right| + \left|DG(v)\right|} \tag{6}$$

After calculating all miRNAs two by two, a similarity matrix SM_1 of $n_m \times n_m$ is summarized, and each element $SM_1\left(m_i, m_j\right)$ in this matrix means the functional similarity of the i-th miRNA to the j-th miRNA.

2.5 GIP kernel similarity for miRNA

To enhance the richness of miRNA similarity information, this paper also introduces Gaussian nuclear interaction similarity data for miRNAs. Specifically, the Gaussian nuclear function is employed to compute the Gaussian nuclear interaction spectrum among miRNAs. In a manner analogous to the computation of disease GIP similarity outlined in (2), miRNA GIP similarity $SM_2\left(m_i, m_j\right)$ is defined in this way:

$$SM_2\left(m_i, m_j\right) = \exp\left(-\rho_m \left\| K\left(m_i\right) - K\left(m_j\right) \right\|^2\right) \tag{7}$$

$$\rho_m = \rho_m' \Big/ \left(\frac{1}{nm} \sum_{i=1}^{nm} \left\| K\left(m_i\right) \right\|^2\right) \tag{8}$$

Where ρ_m is the parameter that can control the Gaussian kernel width, which is defined by (8), nm represents the number of miRNA, with the value of ρ_m' established at 1 , following the findings of earlier research [11].

2.6 Integrating similarity for miRNA and disease

The algorithm for fusion of similarity kernels[12] was utilized to combine two different miRNA similarity kernels into a unified and all-encompassing miRNA similarity kernel. Analogously, two disease similarity kernels were amalgamated to form a single, comprehensive disease similarity kernel. To provide clarity, we will use miRNA similarity kernels as a case study to illustrate this fusion methodology, using the HMDD

v3.2 dataset as a reference to explain our approach while maintaining its general applicability.

First, we normalized the above two miRNA similarity kernels ($M_{x,y}$, $y = 1, 2$), and the normalized matrix is denoted as $\Theta_{x,y}$, the formula is as follows:

$$\Theta_{x,y}(i,j) = \frac{M_{x,y}(i,j)}{\sum_{v=1}^{893} M_{x,y}(v,j)} \tag{9}$$

To further suppress noise, we introduce neighborhood constraints for each similarity kernel. For each miRNA x_i, we select the previous most similar k miRNA to form the neighborhood set $N_{x,y}(i,k)$, and apply the neighborhood constraint to the similarity kernel:

$$\Phi_{x,y}(i,j) = M_{x,y}(i,j) I_{x,y,k}(i,j) \tag{10}$$

Where $I_{x,y,k}(i,j)$ is the indicator function, indicating whether miRNA x_j belongs to the miRNA x_i neighborhood.

Finally, through the following iterative process, multiple similarity kernels are fused:

$$\Theta_{x,y}(t+1) = \frac{\alpha}{2}\left(\Phi_{x,y}\sum_{k\neq y}\Theta_{x,k}(t)\Theta_{x,y}^{T}\right) + \frac{1-\alpha}{2}\sum_{k\neq y}\Theta_{x,k}(0) \tag{11}$$

In this context, t denotes the number of iterations, a signifies the fusion weight, and $\Theta_{x,y}(0)$ indicates the initialized similarity kernel. This method can efficiently combine the information from various similarity kernels. Ultimately, we derive the miRNA similarity matrix, denoted as SM, along with the disease similarity matrix, referred to as SD.

2.7 Data Augmentations

Thresholding

In order to retain the stronger association information while initially filtering the noise, we perform a thresholding operation on the similarity matrix. Next, using the miRNA similarity matrix SM as an example, specifically, given a threshold mth, if the similarity $SM_{ij} \geq mth$ is considered to be a potentially strong association between miRNA i and miRNA j, otherwise it is considered to be a weak association and is set to 0. This can be formally written as:

$$SM_{ij}' = \begin{cases} 1 & if\ SM_{ij} \geq mth, \\ 0 & otherwise. \end{cases} \tag{12}$$

Where SM' denotes the thresholded adjacency matrix. A selection grid search for the threshold mth determines the optimal value.

KNN-based Local Enhancement

Although thresholding removes some noise, it fails to capture local structure. Therefore, we use a KNN strategy to construct a locally enhanced view that preserves relevant neighbors. Formally it can be expressed as:

$$KM_{ij} = \begin{cases} 1 & \text{if } j \in \mathrm{KNN}_k(i), \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

The KNN method retains the most similar neighbors for each node, emphasizing local structure and reducing global noise. We combine this with the miRNA-disease association matrix to obtain a locally enhanced view, which better focuses on the "local tightness" relationship for improved learning of local node representations.

Gaussian Noise Based Topology Enhanced View Generation

In order to further simulate the random perturbations that may exist in real biological networks, as well as to improve the robustness of comparative learning, we add Gaussian noise to the similarity matrix SM to construct topology-enhanced views. The specific treatment is as follows:

$$GM = SM + \epsilon,\ \epsilon \sim N(0, \sigma^2 I), \tag{14}$$

where σ is the standard deviation hyperparameter that controls the noise intensity and I is the unit array. For the GM after the addition of noise is combined with the correlation matrix to obtain the final topologically enhanced view.

2.8 Graphical Attention network

Graph attention networks aggregate information more effectively by learning the importance of neighboring nodes. A multi-head graph attention layer, incorporating residual connections, was utilized by us to extract features associated with miRNAs and diseases. Below, we outline the detailed principles and derivations of the formulas.

First, for each node in the graph, let its input feature be $x_i \in R^d$, which we transform to a higher dimensional representation space using linear mapping:

$$h_i = Wx_i,\ \ \mathrm{i}=1,...,\mathrm{M}, \tag{15}$$

where $W \in R^{d' \times d}$ represents the weight matrix that can be learned, and d' indicates the dimension of the mapped features. In the single-head graph attention layer, for each neighboring node pair i and j (where $j \in \mathrm{N}(i)$), $\mathrm{N}(i)$ refers to the collection of neighbors associated with node i , meaning all nodes that are directly linked to node i (where edges are present). We splice their transformed features and compute the attention score using the parameter vector $a \in R^{2d'}$:

$$e_{ij} = Leaky\,\mathrm{Re}\,LU\left(a^T\left[h_i \,\|\, h_j\right]\right) \tag{16}$$

The notation " $\|$ " represents the vector splicing operation, while the LeakyReLU activation function, with a negative slope of 0.2, is employed to improve nonlinear representation[13] . Subsequently, the scores of all neighboring nodes are softmax normalized to obtain the attention coefficients:

$$\alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k \in \mathrm{N}(i)} \exp\left(e_{ik}\right)} \tag{17}$$

Leveraging attention coefficients, the features adjacent to the node i are weighted and aggregated, subsequently generating a new representation via the ELU activation function. To thoroughly capture the varied local structural information present in the

graph, this research incorporates a multi-head attention mechanism that builds upon single-head attention. As illustrated in **Fig. 2**, by executing M single-head attention layers concurrently, the individual node representations are calculated and subsequently merged along the feature dimension:

$$h_i'^{concat} = \Big\|_{m=1}^{M} ELU\left(\sum_{j\in \mathrm{N}(i)} \alpha_{ij} h_j\right)^{(m)}, \tag{18}$$

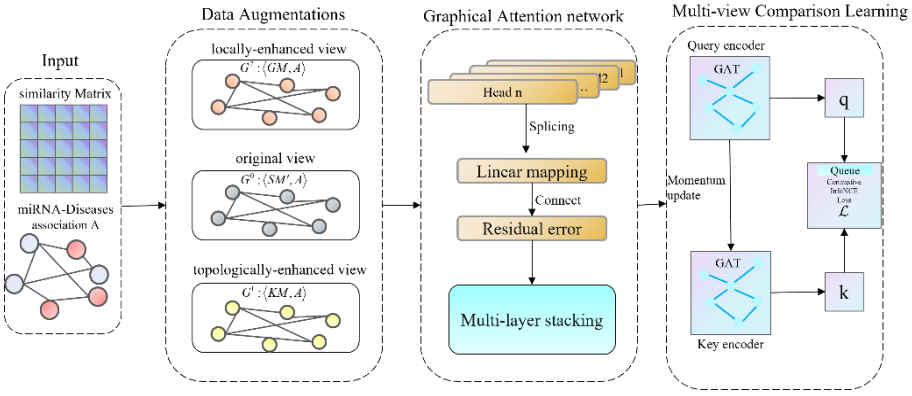

Fig. 2. The feature extraction architecture of MGACMDA

This is followed by fusion of the information from each head by linear mapping and ELU activation:

$$h_i'' = ELU\left(W_c h_i'^{concat}\right), \tag{19}$$

where W_c represents the weight matrix allocated to the target dimension following the splicing process.

To address the issue of information loss and gradient vanishing potentially arising from deep stacking, while maintaining the original feature information, we integrate residual connections across several head-attention layers[14]. The final output consists of the fused attention result summed with the residuals and activated by ReLU:

$$z_i = \mathrm{Re}\,LU\left(h_i'' + W_r h_i\right) \tag{20}$$

The entire graph attention encoder consists of multiple layers stacked with residuals from multiple graph attention layers. Let there be a total of L layers in the encoder, and the input of the l th layer is $H^{(l-1)}$, which is obtained after processing in this layer:

$$H^{(l)} = Dropout(ReLU(GATLayer(l)(H^{(l-1)}, A))), \tag{21}$$

In this context, A represents the graph's adjacency matrix, and the Dropout technique aids in enhancing the model's generalization capabilities.

2.9 Multi-view Contrastive Learning Model

Utilizing the MoCo framework, we present a module for self-supervised contrastive learning. The fundamental concept is to create pairs of positive and negative samples, maintaining consistency for identical nodes across different views, while differentiating between distinct nodes. This enables the extraction of feature embeddings that capture both local structure and global discrimination[15]. We construct a dynamically updated queue at $C \times MN$ for negative samples in contrast learning, enabling the model to leverage historical negative samples and enhance feature discrimination in miRNA-disease association prediction.

The query encoder and key encoder are represented as $f_q(\bullet)$ and $f_k(\bullet)$ respectively, both of which utilize a GAT-based approach. The ReLU activation function is widely employed in this context. With the notation established, where Z indicates the feature representations of the original view and Z' represents the feature representations of the augmented views (which include locally and topologically augmented views), and the graph attention encoder, which is denoted by GATE^K and incorporates K attention heads, can be articulated in our representation learning approach as follows:

$$Z = \mathrm{GATE}^{\mathrm{K}}(Y, A), \quad \mathrm{Z'}=\mathrm{GATE}^{\mathrm{K}}(Y', A) \tag{22}$$

After obtaining the latent representations of the query and the key, we first normalize them to ensure that all vectors are distributed on the unit sphere so that the dot product operation can better reflect the cosine similarity. For each sample i :

The positive sample score is calculated as:

$$l_i^{(+)} = \langle Z_i, Z_i' \rangle \tag{23}$$

The negative sample score is calculated as

$$l_{ij}^{(-)} = \langle Z_i, q_j \rangle, \ j = 1, ..., MN, \tag{24}$$

Where q_j is from the dynamic negative sample queue.

The scores derived from both positive and negative samples were aggregated into a logit vector. Subsequently, this vector was adjusted by applying the temperature parameter T :

$$I_i = \left[l_i^{(+)}, l_{i1}^{(-)}, ..., l_{iMN}^{(-)} \right] \ \text{且} I_i \leftarrow \frac{I_i}{T} \tag{25}$$

Finally, we optimized the model using InfoNCE loss:

$$\mathrm{L}_i = -\log \frac{\exp\left(l_i^{(+)} / T\right)}{\exp\left(l_i^{(+)} / T\right) + \sum_{j=1}^{MN} \exp\left(l_{ij}^{(-)} / T\right)} \tag{26}$$

The key encoder inputs are considered to come from different views (local augmented view and topological augmented view) so as to obtain feature embeddings with both local structural information and global discriminative power. We construct two

independent contrast learning branches to compute the losses L_1' and L_2', respectively, and weight the fusion with the weighting parameter λ:

$$L = \lambda L_1' + (1-\lambda) L_2'. \tag{27}$$

The swift evolution of encoder networks may diminish the consistency of embeddings, resulting in subpar performance. To address this issue, we implement a momentum update approach. Denote the parameters of f_q as θ_q, and those of f_k as θ_k. The parameters of the query encoder are optimized through the application of backpropagation. In contrast, the parameters of the key encoder are updated using a momentum-based approach. This momentum-based update mechanism effectively manages and rectifies the model's discrepancies in an appropriate manner. Subsequently, a linear combination operation is carried out on the parameters obtained from both the key encoder and the query encoder:

$$\theta_k \leftarrow m \cdot \theta_k + (1-m) \cdot \theta_q, \tag{28}$$

The momentum factor $m \in [0,1)$ controls the update speed. A larger m value makes θ_k update more slowly. The slow - growing property of the key encoder is crucial. It stabilizes the model's feature extraction and enables more effective use of historical negative samples in contrast learning. As the key encoder's inputs come from different views, it can better capture cross - view consistent features, enhancing the model's node - discrimination ability. A small m may cause rapid key encoder updates, disrupting the utilization of historical negative samples and view - consistency. In brief, a slow - growing key encoder is essential for fully leveraging the queue.

2.10 miRNA-Disease Prediction Module

After obtaining node representations with strong discriminative power after the comparative learning module, it is necessary for us to further develop classifiers that can be utilized to predict the existence of associations between microRNAs (miRNAs) and diseases. For each miRNA-disease pair to be predicted, we extract its final embedding from the vector representations of miRNA and disease obtained by comparative learning, denoted as $h^{(m)}$ and $h^{(d)}$, respectively. These two vectors are then spliced to obtain a composite representation z:

$$z = Concat\left(h^{(m)}, h^{(d)}\right), \tag{29}$$

This comprehensive representation contains both miRNA and disease information, providing sufficient contextual features for the classifier.

The representation is subsequently input into a neural network composed of four fully connected layers, allowing it to learn the inherent nonlinear mapping and produce the final predicted value. During the training process, we make use of the binary cross-entropy loss function to reduce the difference between the predicted value, denoted as $\hat{y}_i$ and the actual label, represented by y_i:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \log\left(\hat{y}_i\right) + \left(1-y_i\right) \cdot \log\left(1-\hat{y}_i\right)\right] \tag{30}$$

Where N the number of samples involved in the loss calculation.

3 EXPERIMENTAL

3.1 Human miRNA-disease associations

In this research, we combined multiple types of data—associations between miRNAs and diseases, sequences of miRNAs, and vocabulary related to diseases—to train and assess our approach. Data was obtained from the HMDDv2.0 and HMDDv3.2 databases, which include 5,430 and 35,547 validated associations of miRNA with diseases, respectively. We gathered miRNA sequence features from miRBase and excluded entries that had missing or misaligned sequences. The disease information was retrieved from MeSH, where each disease is designated with a unique name and organized in a hierarchical structure as a directed acyclic graph (DAG). After eliminating duplicates, we preserved 495 miRNAs in HMDDv2.0 and 1,126 miRNAs in HMDDv3.2, leading to a total of 5,430 and 18,629 associations between miRNAs and diseases, respectively.

Table 1. Benchmark Databases Description

Dataset	Disease	miRNA	associations
HMDD v2	383	495	5430
HMDD v3.2	893	1126	18629

3.2 Evaluation indicators

In order to assess the effectiveness of the suggested miRNA-disease association prediction model (MGACMDA) utilizing graph-attention contrastive learning, we performed an experimental analysis through five-fold cross-validation and chose various well-established evaluation metrics, such as AUC, AUPR, and F1-score.

In order to calculate the Area Under the Curve (AUC), multiple threshold values are applied to compute both the True Positive Rate (TPR) and the False Positive Rate (FPR). Subsequently, the ROC curve is generated, and the area beneath this curve is computed. It is important to note that both AUPR and AUC values span from 0 to 1, with higher values signifying improved predictive effectiveness of the method.

3.3 Parametric Analysis

The MGACMDA model was developed and executed in Python. It made use of PyTorch 1.13 and DGL libraries, and the implementation took place on an Ubuntu 20.04 operating system. We optimized the model by employing the Adam optimizer with a learning rate of 10^{-3}. The batch size was configured as 64, and the training process consisted of a total of 200 iterations. For the classifier, training was conducted over 50 iterations, incorporating 8 attention heads within the encoder.

In the construction stages of the original and augmented views, key parameters such as miRNA and disease similarity thresholds mth, dth, and the number of neighbors k_1,

k_2 in the K-nearest neighbor algorithm affect model performance. We employed the control variable method to identify optimal settings. First, we fix k_1 and k_2 as 10, and varied *mth* and *dth* as 0.1 to 0.9, with an interval of 0.1 to train the model, and the outcomes are presented in **Fig. 3.** Finally, we found that the model has the best prediction performance when *mth* , *dth* are set to 0.6 and 0.7 respectively, in order to find out the optimal value of k_1 , k_2 , then we fixed *mth* , *dth* to 0.6 and 0.7, and set k_1 , k_2 to 0 to 100, with an interval of 10 to train the model, and the model has the optimal prediction performance when k_1 , k_2 are set to 30, 50, and the predictive performance of the model reaches to 0.9448.and the outcomes are presented in **Fig. 4.**

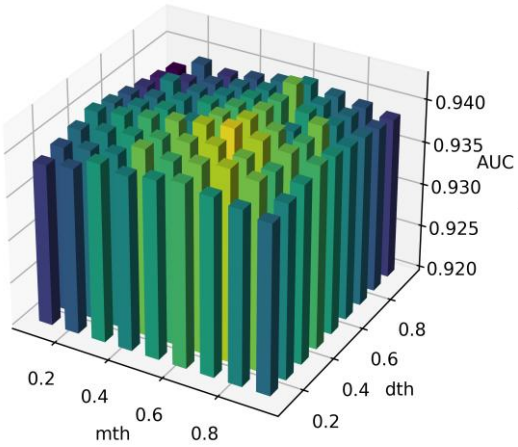

Fig. 4. AUC with Different Thresholds

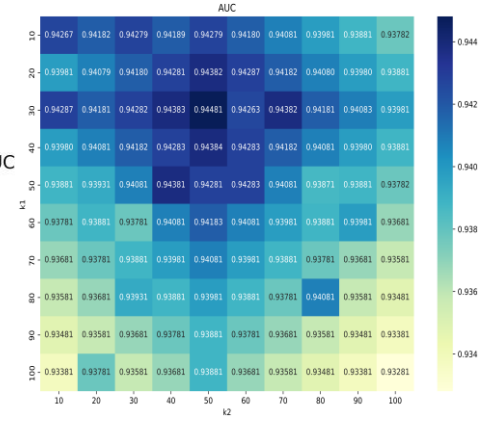

Fig. 3. AUC Heatmap with different k1 and k2 values

3.4 Comparison with other methods

The evaluation of MGACMDA involved a comparison with four sophisticated methods: NIMCGCN, DBMDA, MuCoMiD, and PMDAGS, utilizing AUC values for this analysis. We conducted 5-fold cross-validation on the HMDD v2.0 and HMDD v3.2 datasets. To assess the performance of the models employed during this process, we utilized the Area Under the Curve (AUC) metric.

NIMCGCN employs a graph convolutional network (GCN) along with the NIMC model to forecast miRNA-disease associations[16]. DBMDA transforms sequence similarity into Euclidean features and applies a rotated forest classifier[17]. MuCoMiD is a multi-task GCN model that consolidates various biological data sources to predict associations between miRNAs and diseases[18]. PMDAGS, on the other hand, derives feature embedding vectors by leveraging a nonlinear diffusion graph convolutional network, which are then used as input for a multi-layer perceptron to facilitate prediction[19].

Under identical testing conditions, the outcomes of the comparisons for the balanced test set are presented in **Table 2**. The MGACMDA approach exhibits better performance than alternative methods, attaining AUC values of 0.9223 for the HMDD v2.0 dataset and 0.9436 for the v3.2 dataset. The model's predictive effectiveness is illustrated by the ROC and Precision-Recall (P-R) curves shown **Fig. 5**. An increased area beneath the ROC curve signifies enhanced performance.

Table 2. Performance Comparison With Other Methods Using AUC Based On 5-CV On HMDD V2.0 And HMDD V3.2 Datasets

METHOD	HMDD V2.0	HMDD V3.2
	AUC	AUC
NIMCGCN	0.8851	0.9387
DBMDA	0.8438	0.8743
MUCOMID	0.8374	0.9168
PMDAGS	0.9222	0.9366
MGACMDA	0.9223	0.9436

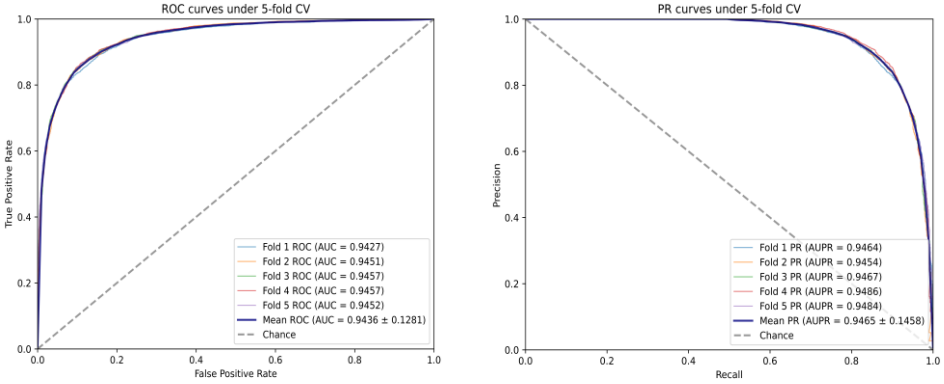

Fig. 5. ROC curves and P-R curves of MGACMDA on HMDD v3.2.

3.5 Ablation Study

Graph Attention Encoder Performance Analysis Using Residual Connections

Deep graph neural networks encounter challenges such as feature fading and gradient vanishing, which result in the loss of shallow topological information. To tackle this issue, we incorporate residual connectivity in the encoder to facilitate cross-layer feature fusion. To assess its effectiveness, we create MGACMDA-NR by eliminating all residual connections. Subsequently, we conduct five-fold cross-validation on the HMDD v3.2 dataset.

Table 3. Performance Comparison Using the Residual Graph Attention Encoder under 5-CV

	5-CV		
Method	**AUC**	**AUPR**	**F1-score**
MGACMDA	0.9448	0.9471	0.8740
MGACMDA-NR	0.8323	0.8207	0.7452

As shown in **Table 3**, the full model enhances AUC and AUPR by 13.51% and 15.40%, respectively, while the F1-score rises from 0.7452 to 0.8740. These findings suggest that the residual mechanism significantly alleviates the problem of information degradation in deep networks.

Synergistic Enhancement of Multi-View Comparative Learning

Single-view learning methods have limitations in miRNA-disease prediction, as they struggle to capture fine-grained features and are prone to overfitting in sparse scenarios. To address this, we propose a global-local-topological three-view-based comparative learning framework that enhances the model's ability to mine complementary features through cross-view consistency. To verify this, we set up three variant models by disabling specific views: MGACMDA-S1 (global view only), MGACMDA-S2 (local view only), and MGACMDA-S3 (topological view only). The HMDD v3.2 dataset underwent five-fold cross-validation.

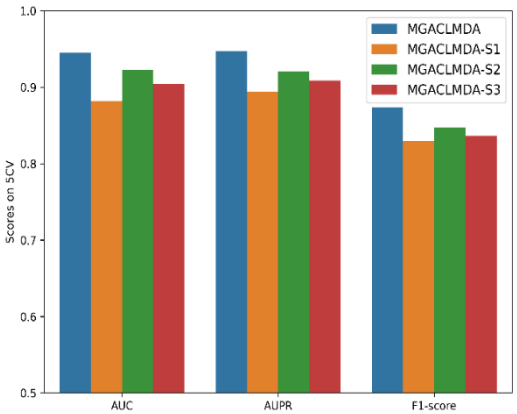

Fig. 6. Performance Comparison of MGACMDA under Different Views

As shown in **Fig. 6,** the multi-view comparison model shows significant improvements in AUC, AUPR, and F1-score, indicating that multi-view learning enhances the model's ability to mine heterogeneous features.

3.6 Predictive performance analysis using different classifiers

In the suggested MGACMDA framework, we employed a DNN classifier to forecast miRNA-disease relationships, and we assessed its effectiveness against MLP and LightGBM classifiers to determine its advantages.

MLP is a type of DNN with input, hidden, and output layers, where neurons perform nonlinear transformations using activation functions. LightGBM is an efficient Gradient Boosting Tree (GBT) framework that optimizes traditional gradient boosting for high training efficiency and low memory usage.

We named the MGACMDA models with the MLP and LightGBM classifiers as MGACMDA-M and MGACMDA-L, respectively. A five-fold cross-validation was used to evaluate each classifier's performance in miRNA-disease prediction. The results in **Fig. 7** show that the DNN classifier significantly outperforms MLP and LightGBM in AUC, AUPR, and F1-score.

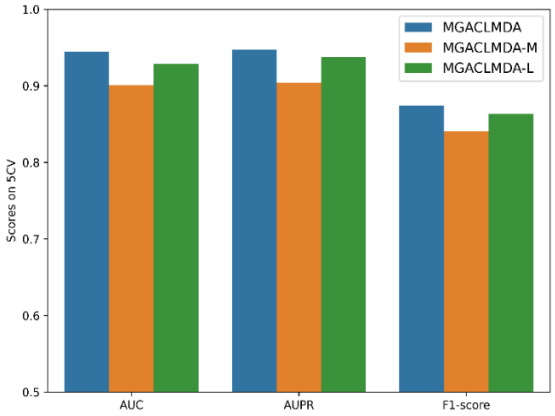

Fig. 7. Comparison of evaluation indexes of different classifiers

3.7 Case Studies

In order to thoroughly assess the true predictive capacity of MGACMDA, we performed case studies focusing on three significant human diseases: colon cancer and triple-negative breast cancer. Initially, we employed the established connections identified in the HMDD v3.2 database to serve as the training dataset for MGACMDA. Subsequently, we evaluated the leading 30 candidate miRNAs associated with these diseases, referencing both the HMDD v4.0 and dbDEMC databases.

Table 4. The MGACMDA predicted the top 30 miRNAs associated with colon neoplasms

miRNA(1-15)	Evidence	miRNA(16-30)	Evidence
h-m-21	V4.0 dEC	h-m-125b	V4.0 dEC
h-m-155	V4.0 dEC	h-m-210	V4.0 dEC
h-m-146a	V4.0 dEC	h-m-31	V4.0 dEC
h-m-34a	V4.0 dEC	h-m-19a	V4.0 dEC
h-m-145	V4.0 dEC	h-m-15a	dEC
h-m-223	V4.0 dEC	h-m-142	V4.0
h-m-126	V4.0 dEC	h-m-203	V4.0 dEC
h-m-17	V4.0 dEC	h-m-182	V4.0 dEC
h-m-221	V4.0 dEC	h-m-214	V4.0 dEC
h-m-150	dEC	h-m-30a	dEC
h-m-20a	V4.0 dEC	h-m-27a	V4.0 dEC
h-m-29a	V4.0 dEC	h-m-29b	V4.0 dEC
h-m-143	V4.0 dEC	h-m-26a	V4.0 dEC
h-m-16	dEC	hsa-let-7a	V4.0 dEC
h-m-222	V4.0 dEC	h-m-200b	V4.0 dEC

V4.0 denotes HMDD v4.0,dEC denotes dbDEMC,h-m denotes has-mir.

Colon cancer ranks as the third most prevalent cancer across the globe, exhibiting comparable incidence rates among both genders. According to the HMDD v3.2 database, **Table 4** presents the 30 most notable miRNAs linked to colon tumors. All identified molecules received validation in the dataset used for confirmation, with 29

being corroborated by HMDD v4.0 and 29 by dbDEMC v3.0. This indicates that the model demonstrates significant generalization capability.

Table 5. The MGACMDA predicted the top 30 miRNAs associated with breast cancer

miRNA(1-15)	Evidence	miRNA(16-30)	Evidence
h-m-21	V4.0 dEC	h-m-16	V4.0 dEC
h-m-155	V4.0 dEC	h-m-19a	V4.0 dEC
h-m-146a	V4.0 dEC	h-m-210	V4.0 dEC
h-m-34a	V4.0 dEC	h-m-31	V4.0 dEC
h-m-145	V4.0 dEC	h-m-15a	V4.0 dEC
h-m-17	V4.0 dEC	h-m-142	V4.0
h-m-223	V4.0 dEC	h-m-182	V4.0 dEC
h-m-126	V4.0 dEC	h-m-203	V4.0 dEC
h-m-221	V4.0 dEC	h-m-200b	V4.0 dEC
h-m-150	V4.0 dEC	h-m-29b	V4.0 dEC
h-m-20a	V4.0 dEC	h-m-30a	dEC
h-m-143	V4.0 dEC	h-m-27a	V4.0 dEC
h-m-29a	V4.0 dEC	h-m-26a	V4.0 dEC
h-m-222	V4.0 dEC	hsa-let-7a	V4.0 dEC
h-m-125b	V4.0 dEC	h-m-122	V4.0 dEC

V4.0 denotes HMDD v4.0,dEC denotes dbDEMC,h-m denotes has-mir.

Triple negative breast cancer (TNBC) represents 15% of breast cancer instances and is associated with a dismal prognosis stemming from the absence of targeted treatments. in **Table** 5, we conduct and present a Triple Negative Breast Cancer (TNBC) case study utilizing the HMDD v3.2 dataset. The findings of this case study reveal that among the top 30 candidate microRNAs (miRNAs), 26 were verified by the HMDD v4.0 dataset, and 29 were validated through the dbDEMC v3.0 database.

4 RESULTS AND DISCUSSION

Computational approaches aimed at detecting microRNA (miRNA) - disease associations play a pivotal role in disease-related research endeavors. We propose MGACMDA, a multi - view graph attention contrast - learning - based model for predicting such associations. Our model's multi - view contrastive learning outperforms single - view methods. Single - view methods capture limited information, while our multi - view framework combines global, local, and topological views. The global view gives an overall understanding, the local view catches fine - grained signals for accurate prediction, and the topological view with added Gaussian noise improves generalization. This helps the model extract comprehensive features and reduce overfitting in sparse data. Integrating these views with a residual attention encoder and dynamic contrast learning, MGACMDA becomes more robust to data sparsity, suppresses network noise, and extracts deep - topological information. Experiments on HMDD v2.0 and v3.2 datasets show AUCs of 0.9223 and 0.9448. Ablation experiments prove the significance of multi - view contrast learning and residual connectivity. Case studies on triple - negative breast cancer and colon cancer show that over 92% of the

top 30 predicted miRNAs were validated by independent databases. This study provides a new way to understand the miRNA regulatory network, and future applications may involve integrating single - cell sequencing and multi - omics data.

Funding. This work was supported by the National Natural Science Foundation of China under Grant 61873001 and the Natural Science Foundation of Shandong Province under Grant ZR2020KC022.

References

1. Sheng, Y., P.G. Engström, and B. Lenhard, *Mammalian MicroRNA Prediction through a Support Vector Machine Model of Sequence and Structure.* Plos One, 2007. **2**(9).
2. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. **75**(5): p. 843-854.
3. Lin, C., et al., *MicroRNA-1179 inhibits the proliferation, migration and invasion of human pancreatic cancer cells by targeting E2F5.* 2018. **291**: p. 65-71.
4. Santos, J.C., et al., *Exosome-mediated breast cancer chemoresistance via miR-155 transfer*. Scientific Reports, 2018. **8**.
5. Zheng, K., et al., *MLMDA: a machine learning approach to predict and validate MicroRNA-disease associations by integrating of heterogenous information sources.* J Transl Med, 2019. **17**(1): p. 260.
6. Chen, X., et al., *RBMMMDA: predicting multiple types of disease-microRNA associations.* Sci Rep, 2015. **5**: p. 13877.
7. Min, S., B. Lee, and S. Yoon, *Deep learning in bioinformatics.* Brief Bioinform, 2017. **18**(5): p. 851-869.
8. Li, Z., et al., *A graph auto-encoder model for miRNA-disease associations prediction.* Brief Bioinform, 2021. **22**(4).
9. Ji, C., et al., *AEMDA: inferring miRNA-disease associations based on deep autoencoder.* Bioinformatics, 2021. **37**(1): p. 66-72.
10. Li, J., et al., *FCGCNMDA: predicting miRNA-disease associations by applying fully connected graph convolutional networks.* Molecular Genetics and Genomics, 2020. **295**(5): p. 1197-1209.
11. Chen, X., et al., *WBSMDA: Within and Between Score for MiRNA-Disease Association prediction.* Scientific Reports, 2016. **6**(1): p. 21106.
12. Xie, G., et al., *SKF-LDA: Similarity Kernel Fusion for Predicting lncRNA-Disease Association.* Mol Ther Nucleic Acids, 2019. **18**: p. 45-55.
13. Maas, A.L. *Rectifier Nonlinearities Improve Neural Network Acoustic Models*. 2013.
14. Ray, D., O. Pinti, and A.A. Oberai, *Residual Neural Networks*, in *Deep Learning and Computational Physics*, D. Ray, O. Pinti, and A.A. Oberai, Editors. 2024, Springer Nature Switzerland: Cham. p. 41-48.

15. Tian, S.W., et al., *scGCC: Graph Contrastive Clustering With Neighborhood Augmentations for scRNA-Seq Data Analysis.* IEEE Journal of Biomedical and Health Informatics, 2023. **27**(12): p. 6133-6143.
16. Li, J., et al., *Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction.* Bioinformatics, 2020. **36**(8): p. 2538-2546.
17. Zheng, K., et al., *DBMDA: A Unified Embedding for Sequence-Based miRNA Similarity Measure with Applications to Predict and Validate miRNA-Disease Associations.* Molecular Therapy - Nucleic Acids, 2020. **19**: p. 602-611.
18. Dong, N., S. Mücke, and M. Khosla, *MuCoMiD: A Multitask Graph Convolutional Learning Framework for miRNA-Disease Association Prediction.* IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2022. **19**(6): p. 3081-3092.
19. Yan, C. and G. Duan, *PMDAGS: Predicting miRNA-Disease Associations With Graph Nonlinear Diffusion Convolution Network and Similarities.* IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2024. **21**(3): p. 394-404.