

# Predicting miRNA-disease associations based on spectral graph transformer with dynamic attention and regularization

Zhengwei Li, Xu Bai, Ru Nie, Yanyan Liu, Lei Zhang, and Zhuhong You, *Member,*

IEEE

**Abstract**— Extensive research indicates that microRNAs (miRNAs) play a crucial role in the analysis of complex human diseases. Recently, numerous methods utilizing graph neural networks have been developed to investigate the complex relationships between miRNAs and diseases. However, these methods often face challenges in terms of overall effectiveness and are sensitive to node positioning. To address these issues, the researchers introduce DARSFormer, an advanced deep learning model that integrates dynamic attention mechanisms with a spectral graph Transformer effectively. In the DARSFormer model, a miRNA-disease heterogeneous network is constructed initially. This network undergoes spectral decomposition into eigenvalues and eigenvectors, with the eigenvalue scalars being mapped into a vector space subsequently. An orthogonal graph neural network is employed to refine the parameter matrix. The enhanced features are then input into a graph Transformer, which utilizes a dynamic attention mechanism to amalgamate features by aggregating the enhanced neighbor features of miRNA and disease nodes. A projection layer is subsequently utilized to derive the association scores between miRNAs and diseases. The performance of DARSFormer in predicting miRNA-disease associations is exemplary. It achieves an AUC of 94.18% in a five-fold cross-validation on the HMDD v2.0 database. Similarly, on HMDD v3.2, it records an AUC of 95.27%. Case studies involving colorectal, esophageal, and prostate tumors confirm 27, 28, and 26 of the top 30 associated miRNAs against the dbDEMC and miR2Disease databases, respectively. The code and data for DARSFormer are accessible at <https://github.com/baibaibaialone/DARSFormer>.

## Index Terms— Prediction of miRNA-disease associations

This work was supported by the Guangxi Natural Science Foundation under Grant 2023JJA170019, in part by the Natural Science Foundation of Shandong Province under Grant ZR2022LZL003. We would like to thank the anonymous referees for their encouraging remarks, valuable feedback, and suggestions for improving the paper. (Corresponding authors: Ru Nie; Lei Zhang; Zhuhong You.)

Zhengwei Li is with the School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China, and also with the Big Data and Intelligent Computing Research Center, Guangxi Academy of Science, Nanning 530007, China (email: zwli@cumt.edu.cn).

Ru Nie is with the School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China, and also with the Mine Digitization Engineering Research Center of the Ministry of Education, Xuzhou 221116, China (email: nr@cumt.edu.cn).

Xu Bai, Yanyan Liu, Lei Zhang are with the School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China (emails: 1307981929@qq.com; 1554949563@qq.com; zhanglei@cumt.edu.cn).

Zhuhong You is with the School of Computer Sciences, Northwestern Polytechnical University, Xi'an 710129, China (e-mail: zhuhongyou@nwpu.edu.cn).

involves the utilization of graph transformers, dynamic attention mechanisms, and orthogonal graph neural networks.

## I. INTRODUCTION

MiRNAs have been a central subject in the investigation of disease pathophysiology. These non-coding RNAs, consisting of approximately 22 nucleotides, are capable of precisely regulating the expression of the 3' untranslated regions (3' UTR) of target mRNAs [1], [2]. The growing body of experimental data suggests that the abnormal expression and functional dysregulation of miRNAs are intimately linked to a variety of diseases. For instance, evidence indicates that the suppression of miR-92a may facilitate the acceleration of the wound healing process [3], [4], while miR-17 has been associated with Autosomal Dominant Polycystic Kidney Disease (ADPKD) [5], [6]. Consequently, the prediction of potential miRNA-disease associations holds significant implications for understanding disease pathogenesis and advancing new therapeutic strategies, drug discovery, and gene-oriented treatments.

Historically, the prediction of potential miRNA-disease associations heavily relied on biological experiments, including Reverse Transcription Polymerase Chain Reaction (PCR) [7], Northern Blotting [8], and Microarray Analysis [9]. Despite the exemplary performance of these methods, their substantial time requirements cannot be overlooked. Additionally, computational methods have gained widespread application across various fields due to their efficiency and cost-effectiveness. Consequently, an increasing number of researchers have committed to developing various computational methods for predicting miRNA-disease associations. These methods generally fall into two primary categories: similarity-based approaches and machine learning-based approaches.

In similarity-based approaches, it is postulated that microRNAs (miRNAs) with analogous functions are likely to be linked with similar diseases, and the converse is also true. Initially, the methods for predicting the relevance of miRNAs to diseases based on similarity were confined to associations between miRNAs and diseases that had been experimentally validated. For instance, Jiang *et al.* developed a method that prioritizes diseases across the entire human microRNAome by integrating functionally related microRNA networks with human phenotypic networks [10]. A noted limitation is the

scarcity of experimentally validated miRNA-disease associations, which could impact the model's performance and generalizability. Zhao *et al.* introduced DCSMDA, which incorporates known lncRNA-disease and miRNA-lncRNA associations, along with measurements of disease and lncRNA similarities [11]. This approach established a miRNA-lncRNA-disease network and demonstrated high accuracy in multiple case studies, effectively addressing the challenge of sparse information regarding miRNA-disease associations. With ongoing research, subsequent methods have been developed to predict unknown links between miRNAs and diseases. For example, Chen *et al.* proposed SDMMDA, which combines various similarities and known associations and introduced the concepts of "super miRNAs" and "super diseases" to enhance accuracy, without depending on previously confirmed miRNA-disease associations [12].

In recent years, machine learning technology has emerged as a significant area of research and has yielded notable results across various domains. Consequently, the scientific community has explored the integration of similarity-based approaches with machine learning techniques. For example, the study by Chen *et al.* introduced RLSMDA, which employs regularized least squares within a semi-supervised learning framework. This approach combines existing microRNA-disease associations, disease similarity, and miRNA functional similarity to forecast new miRNA-disease connections [13]. This method proves particularly effective for diseases with insufficient confirmed miRNA association data. Similarly, HDMP adopts the weighted k-nearest neighbors concept [14], improving the prediction of miRNA-disease associations by utilizing miRNA functional similarity and assigning increased weights to miRNAs within the same family or cluster. The effectiveness of HDMP has been corroborated through cross-validation and case studies. Additionally, NCCMMDA merges matrix completion with neighborhood constraints and leverages known miRNA-disease associations, miRNA functional similarity, disease semantic similarity, and Gaussian interaction profile kernel similarity [15]. This technique redefines the prediction challenge as an optimization problem, which is addressed using the fast iterative shrinkage-thresholding algorithm. Nonetheless, these methodologies face challenges in managing high-dimensional data and in the extraction and integration of features, which leads to a limited understanding of the complex underlying associations.

Graph Neural Networks (GNNs) play a pivotal role in the field of machine learning, with various researchers employing these networks to explore miRNA-disease associations. For example, GRPAMDA utilizes a graph random propagation network enhanced by DropFeature, alongside an attention network. This model constructs a heterogeneous graph comprising miRNA and diseases, enhances node features via a random propagation mechanism, and aggregates these features through an attention mechanism. Subsequently, miRNA-disease association scores are generated by a projection layer [16]. Building upon this framework, Zhong *et al.* introduced ADPMDA, which features an adaptive depth propagation graph neural network that incorporates attention mechanisms to aggregate node features and adaptively adjust local and global informa-

tion through propagation layers. This approach also involves projecting miRNA and disease features into a low-dimensional space using an attention mechanism, while the network's propagation layer is designed to learn node embeddings that balance local and global information [17]. Additionally, Qing *et al.* developed GCNA-MDA, which employs dual autoencoders to extract robust node representations and capture the topological information of the miRNA-disease network using GCNs. This method integrates association and feature similarity data, substantially enhancing the performance of miRNA-disease associations prediction [18]. Despite the successes of these methods, they are not without limitations. On the one hand, the focus of existing methods is primarily on aggregating local information from neighboring nodes via spatial graph neural networks [21], [22], [23]. To fully leverage global information, multiple convolutional layers are often required to be stacked, potentially impeding propagation efficiency. On the other hand, the reliance on static attention mechanisms may oversimplify the models into linear layers, leading to overfitting [24], [25].

To address the challenges outlined, a novel deep learning model named DARSFormer has been proposed. This model integrates dynamic attention mechanisms with spectral domain graph Transformers. The integration of miRNA-disease associations information facilitates the construction of a heterogeneous miRNA-disease graph. Eigenvalue scalars, derived from the decomposition of the adjacency matrix, are encoded into vector space, capturing the magnitude and variation of all eigenvalues effectively. A regularized graph neural network is then introduced to enhance the structural features of nodes and to ensure stable forward and backward propagation. Subsequently, a dynamic attention mechanism is employed to aggregate features, capturing the rich structure and node information of the miRNA disease heterogeneous network for node embedding. Within the standard Transformer framework, each feature dimension is assigned a unique learnable Laplacian matrix, ultimately forming a node-specific feature representation. The final embeddings of miRNA and disease nodes are input into a projection layer to determine association scores. The entire model undergoes end-to-end training through cross-entropy loss and the backpropagation algorithm. DARSFormer is compared with six other state-of-the-art methods, and several case studies are conducted. The results from these studies confirm the effectiveness of DARSFormer in predicting potential miRNA-disease associations.

## II. MATERIALS AND METHODS

This section offers a detailed exposition of the essential preparatory work required for the construction of the DARSFormer model, as well as an in-depth discussion on the comprehensive architectural framework of the model. To enhance the clarity of the ensuing discussion, Table I summarizes the key symbols utilized in the study.

### A. Human miRNA–disease associations datasets

In this study, experimentally validated human miRNA-disease associations were obtained from HMDD v2.0 [17],

**TABLE I: Summary of the main notations used in the DARSFormer study.**

Variable	Description
$DM$	Adjacency matrix of diseases to miRNAs
$MFS$	Sparse matrix storing miRNA functional similarity
$MGS$	Gaussian interaction profile kernel similarity for miRNAs
$\alpha_r$	The kernel bandwidth parameter
$\alpha_w$	The initial kernel bandwidth
$IM$	MiRNA feature matrix
$ID$	Disease feature matrix
$F_{r(i)}$	The $i$ -th miRNA node feature representation
$F_{d(i)}$	The $i$ -th disease node feature representation
$U$	Eigenvectors of adjacency matrix decomposition
$G$	MiRNA-disease network graph
$V$	Edge Set of miRNA-disease Graph
$A$	MiRNA-disease adjacency matrix
$\rho(\lambda)$	Eigenvalue mapping function
$\xi$	Hyperparameter
$H$	Eigenvalue residual concatenation matrix
$Ortho(\cdot)$	Regularization process
$W^Q$	Query Matrix in Transformer
$W^K$	Key Matrix in Transformer
$W^V$	Value Matrix in Transformer
$\alpha_T$	Feed-forward neural network
$S_m$	The $m$ -th new basis
$\hat{S}$	The composite structure
$X_{:,i}^{(l-1)}$	The $i$ -th column feature of layer $l-1$
$f(\cdot)$	A projection layer
$sigmoid(\cdot)$	A nonlinear activation function
$\lambda_{reg}$	The L2 regularization term
$\mathcal{L}_{L2}$	L2 regularization loss function
$\mathcal{L}_{total}$	The loss function of the entire model

including 5,430 known associations involving 495 miRNAs and 383 diseases. Additionally, the latest version, HMDD v3.2 [18], was utilized to further validate the performance of the model, extracting 8,968 confirmed miRNA-disease associations, involving 788 miRNAs and 374 diseases.

**1) MiRNA and disease similarity information:** The similarity between miRNAs in this study is calculated through a two-step process. Initially, functional similarity scores of miRNAs are obtained, based on the assumption that miRNAs with analogous functions are likely associated with similar diseases. These data are used to construct an adjacency matrix  $MFS$ , which stores miRNA functional similarity information [26]. However, the sparsity of  $MFS$  necessitates a more comprehensive similarity measure. This is achieved by considering the Gaussian interaction profile kernel similarity for miRNAs, involving the Gaussian interaction profile ( $GI$ ) of each miRNA, represented by its corresponding column in the adjacency matrix  $DM$ , and a kernel bandwidth parameter  $\alpha_r$  [27], [28]. The integrated miRNA similarity ( $IM$ ) is then calculated by combining the miRNA functional similarity and the Gaussian interaction profile kernel similarity ( $MGS$ ) [29].

The calculation of disease similarity involves two main aspects. First, the Medical Subject Headings (MeSH) database is utilized to determine the semantic similarity between diseases [14]. This is accomplished by analyzing the relationships and connections between diseases using a directed acyclic graph (DAG) framework. The second aspect considers the frequency and contribution of the same disease across different DAGs [30]. This approach aids in understanding the various semantic

roles a disease may play in different contexts. By combining these two perspectives, a comprehensive semantic similarity measure for diseases is achieved.

## B. DARSFormer

The approach takes the previously constructed heterogeneous graph as its input. Initially, two distinct linear layers are utilized to transform the two types of nodes into a common feature space, forming the combined feature representation. Spectral decomposition and feature augmentation are then performed on the heterogeneous graphs. In the final phase, a dynamic graph Transformer is employed to predict the probabilities of miRNA-disease associations, as depicted in Figure 1. The entire DARSFormer model can be described by the following six steps.

In DARSFormer (refer to Figure 1), **Step I** involves the global representation of features and the acquisition of eigenvalues and eigenvectors through the decomposition of the adjacency matrix; **Step II** facilitates the encoding of extracted eigenvalues through a set-to-set filter, which maps these values from scalar to vector space; **Step III** applies an orthogonal transformation to the parameter matrix of graph convolution, enhancing the structural features of nodes; **Step IV** incorporates Transformer architecture, substituting the self-attention mechanism with a dynamic attention mechanism; **Step V** assigns an independent and learnable Laplacian matrix to each feature dimension, augmenting the learning capability of the model; **Step VI** concentrates on optimizing the model in an end-to-end manner to improve performance and accuracy. The subsequent sections will provide a detailed discussion of these six steps.

**Step I:** Based on confirmed miRNA-disease associations, an association matrix  $DM \in \mathbb{R}^{nd \times nm}$  has been derived between miRNAs and diseases, where  $nm$  and  $nd$  denote the number of miRNAs and diseases, respectively. In this matrix,  $DM_{ij}$  is set to 1 when a verified association exists between miRNA  $i$  and disease  $j$ ; otherwise, it is set to 0. Additionally, a heterogeneous association network comprising miRNA and disease nodes is constructed based on this relationship matrix. The adjacency matrix  $A$  is defined accordingly.

$$A = \begin{bmatrix} 0 & DM^\top \\ DM & 0 \end{bmatrix} \quad (1)$$

The integrated similarity for miRNAs and diseases functions as node features. For instance, the attribute feature of node  $r(i)$  is defined by its integrated similarity  $r(i)$ . This relationship is illustrated as follows:

$$F_{r(i)} = (v_1, v_2, \dots, v_{495}) \quad (2)$$

$F_{r(i)}$ , the  $i$ -th column of matrix  $IM$ , and  $v_j$ , represent the integrated similarity between miRNAs  $r(i)$  and  $r(j)$ . Similarly, the integrated similarity for disease  $d(i)$  is assigned as its node attribute, as detailed below:

$$F_{d(i)} = (w_1, w_2, \dots, w_{383}) \quad (3)$$

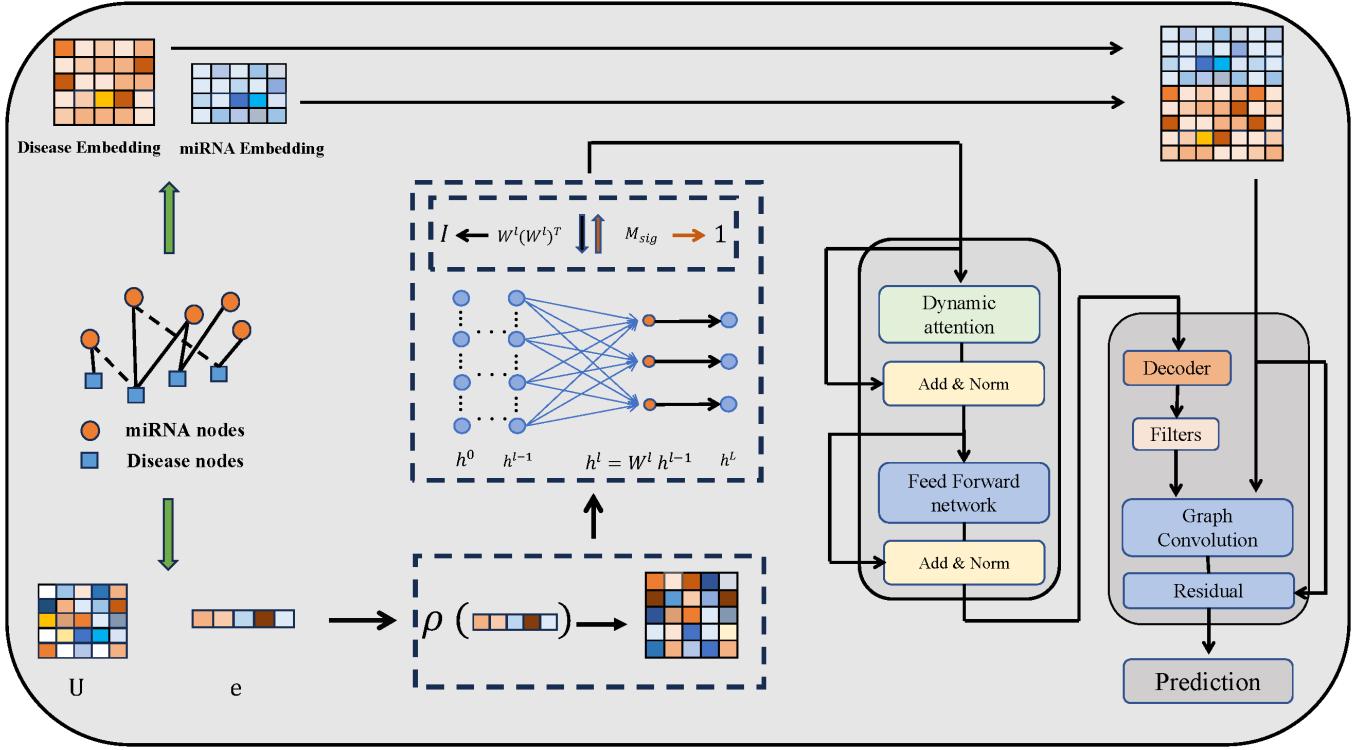


Fig. 1: The overall framework of the DARSFormer model.

The  $i$ -th column of matrix  $ID$ , denoted as  $F_d(i)$ , and  $w_j$ , indicate the integrated similarity between diseases  $d(i)$  and  $d(j)$ .

In consideration of the inherent heterogeneity, miRNA and disease nodes occupy distinct feature spaces. To bridge these disparities, node-type transformation matrices have been employed to amalgamate these elements into a unified feature space. The methods for projecting each node type are detailed as follows:

$$H_r = F_r * W_r \quad (4)$$

$$H_d = F_d * W_d \quad (5)$$

The linear transformation matrices  $W_r$  and  $W_d$  are employed to project the 495-dimensional miRNA nodes and the 383-dimensional disease nodes into a 64-dimensional space, respectively. For miRNA nodes,  $F_r$  and  $H_r$  denote their original and projected features, while  $F_d$  and  $H_d$  similarly represent those for disease nodes. The features  $H_r$  and  $H_d$  are concatenated to form  $X$ , which serves as the feature representation for the heterogeneous graph.

Within the framework of miRNA-disease networks, a graph  $G(\mathcal{V}, \mathcal{E})$  is introduced, with  $\mathcal{V}$  representing a set of nodes totaling  $|\mathcal{V}| = n$  and  $\mathcal{E}$  indicating the set of edges. The adjacency matrix  $A$ , with dimensions  $n \times n$ , is defined such that  $A_{ij} = 1$  if there is an edge between nodes  $i$  and  $j$ , and  $A_{ij} = 0$  otherwise. This configuration aids in the computation of the normalized graph Laplacian matrix  $L = I_n - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ , where  $I_n$  is an identity matrix of size  $n \times n$ , and  $D$  is the diagonal degree matrix with  $D_{ii} = \sum_j A_{ij}$  and  $D_{ij} = 0$  for  $i \neq j$ . Given the undirected nature of  $G$ , the symmetric

matrix  $L$  is subject to spectral decomposition, resulting in  $L = U\Lambda U^\top$ , where  $U$  contains the eigenvectors and  $\Lambda$  is a diagonal matrix comprising the eigenvalues  $[\lambda_1, \lambda_2, \dots, \lambda_n]$ . The visual distributions of the eigenvalues and eigenvectors (PCA) are illustrated in Figure 2 and Figure 3, labeled as 'e' and 'U', respectively.

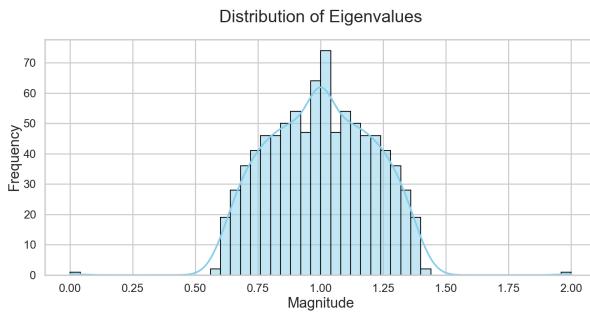
**Step II.** The architecture of the transformer, crucial in spectral filtering, is employed to effectively leverage both the magnitudes and the relative differences of all eigenvalues. However, the direct use of scalar eigenvalues in attention maps constrains the expressiveness of the self-attention mechanism. To overcome this limitation, a function,  $\rho(\lambda) : \mathbb{R}^1 \rightarrow \mathbb{R}^d$ , is introduced to vectorize each scalar eigenvalue, thereby enhancing the attention mechanism [31].

$$\begin{aligned} \rho(\lambda, 2i) &= \sin\left(\frac{\xi\lambda}{10000^{2i/d}}\right), \\ \rho(\lambda, 2i+1) &= \cos\left(\frac{\xi\lambda}{10000^{2i/d}}\right). \end{aligned} \quad (6)$$

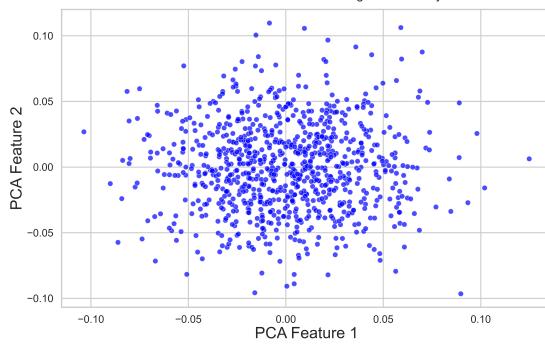
Within the framework of eigenvalue encoding, where  $i$  denotes the dimension of representation and  $\xi$  represents a hyperparameter, the advantages of  $\rho(\lambda)$  are manifold:

- It captures the relative frequency shifts of eigenvalues and provides high-dimensional vector representations,
- It offers a multi-scale representation with wavelengths ranging from  $2\pi$  to  $10000 \cdot 2\pi$ ,
- It allows for the modulation of  $\lambda$  through the hyperparameter  $\xi$ .

It is important to note that in addressing the challenge of feature mapping, a residual methodology is utilized. This



**Fig. 2: Distribution of eigenvalues, denoted as 'e', in th**



**Fig. 3: Distribution of eigenvalues, denoted as 'U', in the DARSFormer model.**

strategy involves the use of residual configurations for the concatenation of eigenvalues and eigenvectors, which are represented as:

$$Z = [\lambda_1 \|\rho(\lambda_1), \dots, \lambda_n \|\rho(\lambda_n)]^\top \in \mathbb{R}^{n \times (d+1)} \quad (7)$$

Subsequently,  $Z$  is transformed linearly into a  $\mathbb{R}^{n \times d}$  vector space for further analysis.

**Step III.** Upon acquiring the mapped structural features  $Z$ , an orthogonal graph neural network is utilized to enhance these features. The parameter matrix  $W^{(l)}$  for each layer undergoes orthogonalization, based on the GCN2 framework. It is observed that orthogonal weights  $W^{(l)}$  contribute to the stability of both forward embeddings and backward gradients within the graph neural network process, thereby enhancing the network's overall stability [32], [33]. The orthogonal graph convolution is specifically designed to maintain orthogonality through the incorporation of mixed weight initialization and orthogonal transformations.

While orthogonal initialization has proven effective for training neural networks, its tendency for rapid convergence may result in local optima, potentially exacerbating overfitting issues. To strike a balance between orthogonality and the learning capabilities of the model, mixed weights are initially set for the layer  $l$  weights as  $M^{(l)} = \beta Q^{(l)} + (1-\beta)I \in \mathbb{R}^{d \times d}$ , where  $Q^{(l)}$  is initialized using conventional random methods and  $\beta$  functions as a hyperparameter.

Following the establishment of the initial weights, an additional orthogonal transformation layer is applied to further refine orthogonality before its application in feature transformation. This process, known as Newton's iteration method, involves two phases: spectral bounding and orthogonal projection. Initially,  $M^{(l)}$  is derived as  $\hat{M}^{(l)} = \frac{M^{(l)}}{\|M^{(l)}\|_F}$ , followed by its orthogonal projection onto the matrix  $\hat{M}^{(l)}$ , resulting in the orthogonal weights  $W^{(l)}$ . The orthogonal projection is mathematically defined as  $W^{(l)} = P^{-\frac{1}{2}} \hat{M}^{(l)}$ , where  $P = \hat{M}^{(l)} \hat{M}^{(l)\top}$  represents the covariance matrix. Due to the computational complexity involved in calculating the square root of the covariance matrix,  $P^{-\frac{1}{2}}$  is computed through Newton's iteration, following the iterative formula:

$$\begin{cases} B_0 = I \\ B_t = \frac{1}{2}(3B_{t-1} - B_{t-1}^3 P), \quad t = 1, 2, \dots, T \end{cases} \quad (8)$$

Under the condition where  $\|I - M\|_2 < 1$ , convergence of  $B_T$  to  $P^{-\frac{1}{2}}$  is observed, resulting in the derivation of orthogonal weights expressed as  $W^{(l)} = B_T \hat{M}^{(l)}$ .

**Step IV.** The canonical Transformer architecture is characterized by two principal components: Multi-Head Attention (MHA) and Feed-Forward Neural Network (FFN). A sequence  $H$  is defined as  $H = \text{Ortho}\left(Z = [\lambda_1 \|\rho(\lambda_1), \dots, \lambda_n \|\rho(\lambda_n)]^\top\right) \in \mathbb{R}^{n \times d}$ , with  $d$  representing the hidden dimensionality. In the MHA component, the input  $H$  undergoes a mapping to query, key, and value spaces through the weight matrices  $W^Q$ ,  $W^K$ , and  $W^V$ , respectively, followed by a transformation via FFN. The model, utilizing a query dimensionality of  $d_q$ , is succinctly represented.

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_q}}\right)V, \quad (9)$$

where  $Q = HW^Q$ ,  $K = HW^K$ , and  $V = HW^V$ .

Although the standard Transformer architecture generally features a higher number of learnable parameters compared to GAT, it faces challenges in effective training under limited supervision, which may increase the risk of overfitting. To improve node-level representation learning adaptability in Transformers, a modification is suggested: the replacement of the conventional self-attention mechanism with GATv2. This innovative attention mechanism acts as a versatile approximator and is shown to be superior to the dot-product self-attention mechanism in the Transformer, with the benefit of reduced parameter count. A comparative analysis between GAT and GATv2 is provided [34].

$$\begin{aligned} \text{GAT: } \alpha(h_i, h_j) &= \text{LeakyReLU}(\alpha^\top \cdot [Wh_i || Wh_j]), \\ \text{GATv2: } \alpha(h_i, h_j) &= \alpha^\top \cdot \text{LeakyReLU}(W[h_i || h_j]). \end{aligned} \quad (10)$$

The symbol  $\alpha^\top$  denotes a feed-forward neural network, while  $W$  constitutes a basic linear operation. **Step V.** Upon completion of the preceding steps, a new feature matrix  $Z' = [\lambda'_1, \lambda'_2, \dots, \lambda'_n]^\top$  was generated, which encapsulates the structural characteristics. For each dimension vector of the eigenvalues in  $Z$ , an independent, learnable base is assigned.

These bases are then concatenated through a residual method, detailed as follows:

$$S_m = U \text{diag}(\lambda'_m) U^\top \quad (11)$$

$$\hat{S} = \text{FFN}([I_n || S_1 || \dots || S_m]) \quad (12)$$

In the presented framework, where  $S_m \in \mathbb{R}^{n \times n}$  signifies the  $m$ -th new basis and  $\hat{S} \in \mathbb{R}^{n \times n \times d}$  represents the composite structure, it is observed that these bases align with the documented objectives of polynomial bases in the literature. However, the approach utilized is adaptive and diverges from the recursive norms typically associated with Chebyshev polynomials.

In conclusion, a separate graph Laplacian matrix is assigned to each feature dimension, based on the learned basis  $\hat{S}$ . This assignment is articulated as follows:

$$\hat{X}_{\cdot,i}^{(l-1)} = \hat{S}_{\cdot,\cdot,i} X_{\cdot,i}^{(l-1)} \quad (13)$$

and

$$X^{(l)} = \sigma(\hat{X}^{(l-1)} W_x^{(l-1)}) + X^{(l-1)} \quad (14)$$

In the  $l$ -th layer, node representations are denoted as  $X^{(l)}$ , where  $\hat{X}_{\cdot,i}^{(l-1)}$  represents the  $i$ -th channel dimension, and  $W_x^{(l-1)}$  signifies the transformation process. The activation function is represented by  $\sigma$ . The inclusion of a residual connection is considered an optional component.

**Step VI.** Upon obtaining the final node representations for both miRNA, denoted as  $r(i)$ , and disease, denoted as  $d(j)$ , the probability of association between them, referred to as  $\hat{P}_{ij}$ , is determined. This is achieved by inputting their concatenated embeddings into a projection layer. Specifically,

$$\hat{P}_{ij} = \text{sigmoid}(f(X_{r(i)} \oplus X_{d(j)})) \quad (15)$$

In this context,  $f(\cdot)$  is defined as a projection layer with specified input dimensions and a single output dimension, while  $\text{sigmoid}(\cdot)$  is described as a nonlinear activation function, and  $\oplus$  represents the concatenation operation.

To enhance the model's performance, a combination of two loss functions with additional emphasis is employed. The primary function, cross-entropy loss, is utilized to measure the divergence between the predictions of the model and the actual labels. It is articulated as follows:

$$\mathcal{L} = -[y \log \hat{P} + (1 - y) \log(1 - \hat{P})], \quad (16)$$

where  $y$  is identified as the genuine association between miRNA and diseases. Additionally, the L2 regularization term is implemented to regulate parameters and improve generalization. This is specified as:

$$L2(W) = \sum_{j=1}^M \|W^j\|_2^2, \quad (17)$$

with  $M$  denoting the total count of model parameters and  $W^j$  indicating the  $j$ -th parameter.

The combined loss, employed for the comprehensive training of the framework, is succinctly formulated as:

$$\mathcal{L}_{total} = \mathcal{L}_{BCE}(y, \hat{y}) + \lambda_{reg} \mathcal{L}_{L2}(W) \quad (18)$$

where  $\lambda_{reg}$  modulates the L2 regularization term.

### III. RESULTS

This section presents the predictive performance of the model based on HMDD versions 2.0 and 3.2. Additionally, a variety of experimental results concerning case studies and parameter determination are introduced in detail.

#### A. Implementation settings

The experiments utilized the Deep Graph Library (DGL) and Pytorch Geometry (PyG). For the training of the DARSFormer model, a five-fold cross-validation method was employed, with the Adam optimizer used for optimizing parameters. The initial settings were established as follows: a batch size of 1000, a learning rate of 0.001, a feature dropout rate of 0.5, a transformer dropout rate of 0.1, eight dynamic attention heads, and two layers. Further adjustments of key parameters, including learning rate and dropout rates, are detailed in Table II.

#### B. Evaluation metrics

Five crucial performance metrics were used to assess the capability of the DARSFormer model to predict miRNA-disease associations:

- **Precision:** Measures the proportion of correctly identified positive cases among all cases classified as positive.
- **Accuracy:** Captures the overall correctness of the model, calculated as the ratio of true predictions (both true positives and true negatives) to the total number of cases.
- **Recall:** Also known as sensitivity, this metric quantifies the proportion of actual positives correctly identified by the model.
- **F1 Score:** Represents the harmonic mean of precision and recall, offering a balance between the two metrics.
- **Area Under the Curve (AUC):** Represents the area under the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate against the false positive rate. AUC is a comprehensive measure of model performance.

The remarkable experimental results, emphasizing the proficiency and reliability of the DARSFormer model in its respective tasks, have been observed and reported following extensive experiments and thorough analysis. The prediction results of DARSFormer on HMDD v2.0 under the setting of 5-fold cross-validation are presented in Table III. The model has exhibited a notable average accuracy of 87.09%, accompanied by a precision of 86.56%, a recall of 87.81%, and an F1 score of 87.17%. These metrics collectively substantiate the model's ability to produce accurate and relevant results. Furthermore, a graphical assessment through the Receiver Operating Characteristic (ROC) curve, as illustrated in Figure 4, demonstrates the model's performance under a five-fold cross-validation scheme, yielding AUC values of 94.25%, 93.77%, 93.91%, 94.08%, and 94.88%. The results generate a remarkable average AUC of 94.18% with a minimal standard deviation of 0.0039, highlighting the model's exceptional capacity to distinguish between different classes. Simultaneously, Figure 5 offers insight into the Precision-Recall (P-R)

TABLE II: Settings of hyperparameters used in the DARSFormer model.

Hyper-parameter	HMDD v2.0	HMDDv3.2	Parameter List	Description
$\lambda_{\text{reg}}$	$10^{-2}$	$10^{-3}$	$[10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}]$	L2 regularization coefficient
layers	2	2	$[2, 3, 4, 5, 6]$	Number of layers in GNNs
dropout_feat	0.5	0.5	$[0.1 - 0.9]$	Dropout of features
dropout_trans	0.1	0.1	$[0.1 - 0.9]$	Dropout of transformer
$l_r$	$10^{-2}$	$10^{-2}$	$[10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}]$	Learning rate
Embedding dimensions	64	128	$[16, 32, 64, 128, 256]$	Dimension of feature
attention_heads dimensions	2	4	$[1, 2, 4, 8, 16]$	Number of attention heads
$\xi$	100	100	$[1, 10, 100]$	Eigenvalue encoding coefficient

curve of DARSFormer, displaying its performance across five-fold cross-validation with AUC values of 94.25%, 92.99%, 93.04%, 93.60%, and 94.45%, and an average AUC of 93.67% with a standard deviation of 0.0060. This further emphasizes the model's balanced performance in terms of precision and recall. Collectively, the data and graphical analyses establish DARSFormer's position as a robust and reliable model.

TABLE III: Results of 5-fold cross-validation performed on the DARSFormer model.

Test set	Precision	Accuracy	Recall	F1-score
1	86.27%	87.02%	88.47%	87.35%
2	87.39%	86.53%	85.84%	86.61%
3	85.53%	86.58%	86.72%	86.12%
4	87.41%	87.38%	87.42%	87.42%
5	86.19%	87.94%	90.59%	88.33%
mean	<b>86.56%</b> $\pm 0.0178$	<b>87.09%</b> $\pm 0.0043$	<b>87.81%</b> $\pm 0.0032$	<b>87.17%</b> $\pm 0.0072$

### C. Performance comparison

To enhance the understanding of DARSFormer's performance, comparative experiments were conducted to assess the AUC values derived from the ROC curves against other leading models. It is important to recognize that the ROC curve offers an intuitive depiction of model efficacy based on training samples with balanced positive and negative distributions. The models involved are summarized below:

**GRPMDA** [16] : A deep learning model is introduced, utilizing an attention mechanism. This model integrates stochastic propagation and neighbor feature enhancement to predict miRNA-disease associations.

**NSAMDA** [35] : Neighbor Selection Graph Attention Networks (NSAMDA) are utilized within miRNA and disease similarity networks, combined with an inner product decoder approach, to forecast miRNA-disease associations. This method constructs a heterogeneous graph by integrating miRNA sequence similarity and comprehensive similarity information, and employs an attention mechanism to select key neighbors for feature aggregation, enhancing the accuracy and efficiency of predictions.

**GAEMDA** [36] : A graph autoencoder is developed to create low-dimensional embeddings of miRNA and disease nodes, employing a bilinear decoder to identify associations between them.

**QIMCMDA** [37] : A matrix completion strategy incorporating q-kernel information is employed to construct a human

miRNA-disease association matrix, which is validated through both five-fold and leave-one-out cross-validation.

**MFMDA** [38] : A matrix factorization approach is introduced to closely approximate an original matrix, with its effectiveness demonstrated via five-fold cross-validation.

**HGANMDA** [39] : A hierarchical graph attention network model is developed, aiming to predict miRNA-disease associations by integrating miRNA-disease, miRNA-lncRNA, and disease-lncRNA interactions.

To ensure fair comparison, five algorithms underwent five-fold cross-validation experiments using the HMDD v2.0 dataset. Figure 6 and Table IV illustrate the comparative analysis of ROC curves between DARSFormer and the six benchmark algorithms.

TABLE IV: Comparative results of the DARSFormer model and other models on the HMDD v2.0 dataset.

Model	AUC(%)	Acc.(%)	Rec.(%)	F1(%)	Pre.(%)
DARSFormer	<b>94.18</b>	<b>87.09</b>	87.81	<b>87.17</b>	<b>86.56</b>
GRPMDA	93.46	86.00	86.37	86.06	85.79
NSAMDA	93.69	85.85	87.28	86.05	85.93
GAEMDA	93.56	84.93	<b>90.70</b>	85.75	81.37
QIMCMDA	91.66	-	-	-	-
MFMDA	90.61	-	-	-	-
HGANMDA	93.74	86.28	86.87	86.36	85.88

### D. Ablation study

To assess the criticality of various components in the DARSFormer model, three variants were introduced: DARSFormer-noregular, DARSFormer-self\_attention mechanism, and DARSFormer-less\_Linner, which serve as baselines for comparison. The DARSFormer-noregular variant omits the use of a regularization parameter matrix, which is typically employed for data augmentation purposes. In the DARSFormer-self\_attention mechanism, the dynamic attention mechanism is replaced with the standard self-attention mechanism found in Transformers, facilitating an analysis of the performance discrepancies between these two attention strategies. Unlike employing a linear layer as the initial encoder, DARSFormer-less\_Linner directly subjects the features to convolution operations.

The comparative results between DARSFormer and its variant models, in terms of average evaluation metrics, are depicted in Figure 7. For DARSFormer and DARSFormer-noregular, the integration of eigenvalues through the regularization of the parameter matrix used in feature enhancement

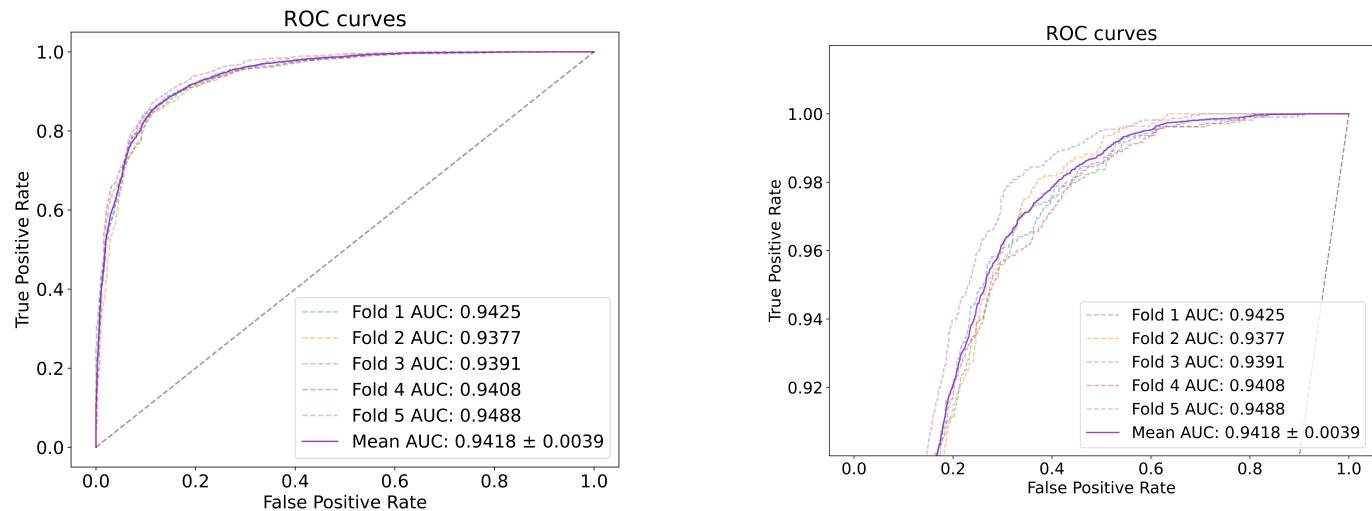


Fig. 4: ROC curves performed by DARSFormer model based on HMDD v.2.0

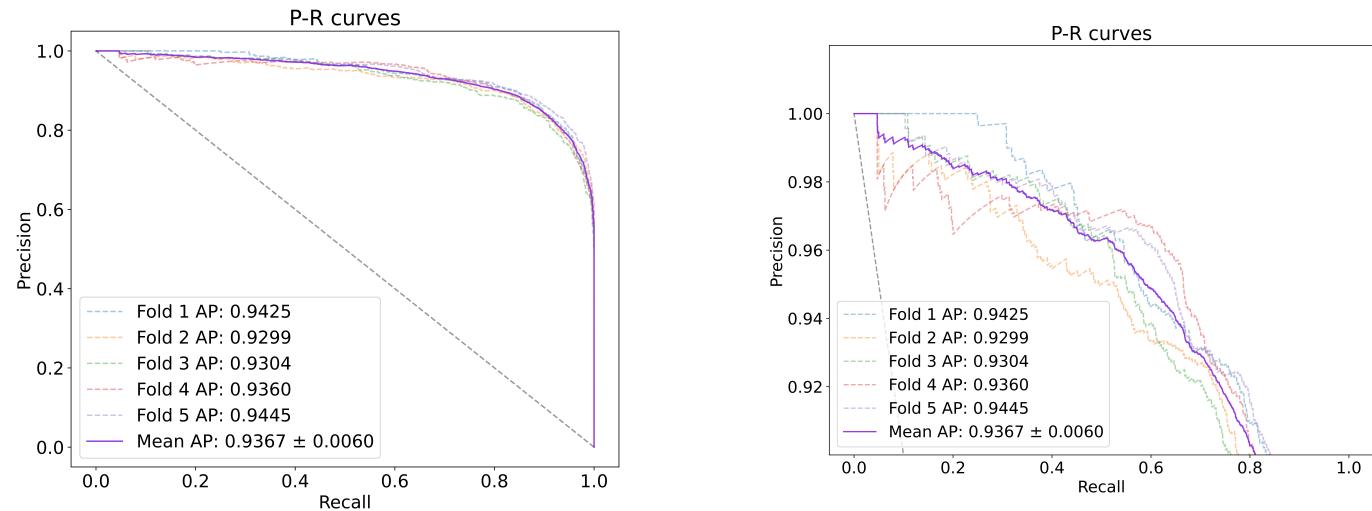


Fig. 5: P-R curves performed by DARSFormer model based on HMDD v.2.0

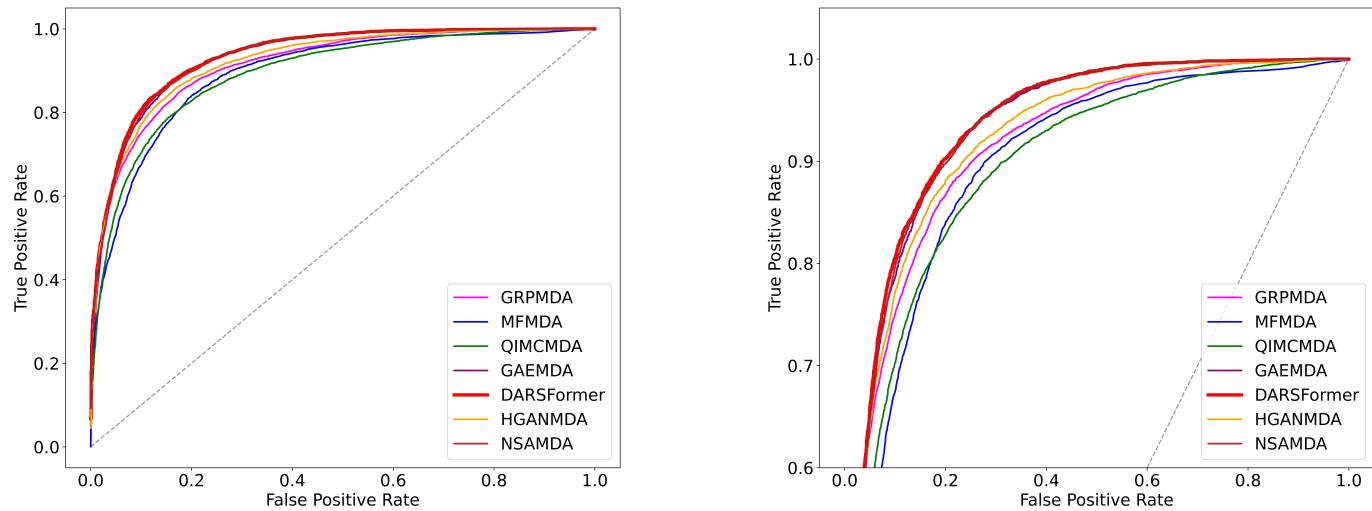


Fig. 6: Comparison of ROC curves.

demonstrates modest improvements. This implies that the regularization of the parameter matrix used in convolution could be advantageous for achieving enhanced convolutional effects. In the comparison between DARSFormer-self\_attention mechanism and DARSFormer, the dynamic attention mechanism exhibits superior representational capabilities relative to the standard self-attention mechanism in Transformers. This superiority may be linked to dynamic attention's improved capacity to assimilate information from neighboring nodes and discern relative information among nodes. Regarding DARSFormer-less\_Linner and DARSFormer, the findings indicate that the introduction of a linear layer prior to the convolutional operations on features promotes more effective learning of feature representations, ultimately resulting in improved embeddings of miRNA and diseases. The experimental results highlight the benefits of incorporating a linear layer before feature convolution for improved feature fusion, and suggest that substituting the self-attention mechanism with a dynamic attention mechanism is beneficial. Additionally, the regularization of the parameter matrix contributes to a slight performance enhancement.

### E. Parameter analysis

In the analysis presented, an examination of specific parameters within the model was undertaken to clarify their influence. Experiments were repeated ten times within a five-fold cross-validation framework on established miRNA-disease associations.

**1) The number of GNN layers:** Graph Neural Networks (GNNs) were utilized to extract features of miRNA and diseases from various angles. According to Figure 8, the number of layers in GNNs appears to have a marginal impact on the model's efficacy. Results show consistent stability with minimal fluctuations in performance as the number of layers varies from 1 to 8. A noticeable decline in performance is observed when the number of layers increases to 8, likely due to overfitting caused by an excessive number of layers. The addition of more GNN layers may result in a reduction in the diversity of node features, thus enhancing the prominence of less significant features. For this research, the GNN layer count was established at 2.

**2) The size of Embedding :** As depicted in the graphical representation (Figure 9), it is evident that the dimensionality of embedding representations in 'DARSFormer' plays a pivotal role in determining the model's effectiveness. Different sizes of node embeddings capture varying amounts of information for the same node. Throughout the experiments, node embedding dimensions were tested at 16, 32, 64, 128, 256, and 512. Empirical data reveal that the model achieves optimal performance when the node embedding dimension is set at 64. It is also observed that both smaller and larger node embedding sizes negatively influence the model's overall performance, a situation possibly linked to the dataset's size.

**3) The size of  $\xi$ :** The parameter  $\xi$  is pivotal in Equation (19), as it directly influences the outcomes of the encoder. By simplistically setting  $\xi$  to 1, as illustrated by the graph (Figure 10), the encoded eigenvalues are distinguishable only

within the initial few dimensions. Adjustment of  $\xi$  to 10 enables differentiation among the eigenvalues across the initial 50 dimensions. Moreover, an increase of  $\xi$  to 100 facilitates a clear distinction among the eigenvalues over the first 120 dimensions.

### F. Case study

To further substantiate the reliability of the DARSFormer model in practical applications, case studies were conducted across several diseases of significant concern. The role of predicting disease associations with microRNAs is crucial, given that approximately 200 miRNAs have been identified as significantly dysregulated in various malignancies. These miRNAs potentially impact cancer progression by targeting oncogenes or tumor suppressor genes, thereby making the accurate prediction of miRNA-disease associations a notable advancement in the field of human medicine and health. In this research, the focus was directed towards colorectal tumors, esophageal tumors, and prostate cancer to further demonstrate the predictive capabilities of the model. Specifically, edges in the miRNA disease heterogeneous graph that connected specific diseases to numerous miRNAs were designated as the test set. All other edges were categorized as the training set, which was used to train the DARSFormer model. The trained model was then employed to generate correlation scores between specific diseases and miRNAs, which were subsequently ranked in descending order. Validation of the associations with high predictive scores was conducted using the DBDEMC [49] and miR2Disease [50] datasets.

In the examination of colorectal tumors, often associated with colorectal cancer (CRC), miRNAs have been recognized as critical regulators in the processes of oncogenesis, progression, metastasis, and inflammation-related pathways. Specifically, miR-21, an oncogenic miRNA, is upregulated in various cancers, including CRC. This miRNA likely promotes tumor growth and dissemination by targeting tumor suppressor genes such as PTEN and PDCD4 [42]. On the other hand, miR-34a, linked to the p53 tumor suppressor pathway, shows diminished expression in several cancers, including CRC. This reduction affects tumor proliferation and apoptosis through its influence on downstream targets like NOTCH, BCL2, and SIRT1 [43]. In Table V, 27 out of the top 30 miRNAs related to colorectal tumors have been validated using the dbDEMC and miR2Disease databases.

In Table VI, the validation of 28 out of the top 30 Esophageal-related miRNAs was verified using the dbDEMC and miR2Disease databases. Esophageal carcinoma, a common malignancy within the digestive tract, primarily comprises esophageal squamous cell carcinoma and adenocarcinoma. Recent studies have focused on the molecular foundations and potential therapeutic targets in these cancers, particularly the role of microRNAs (miRNAs) in tumorigenesis. For instance, miR-21, frequently overexpressed in various cancers and functioning as an oncomir, has been found to promote proliferation and survival in esophageal cancer cells by inhibiting tumor suppressor genes such as PTEN [44]. Conversely, miR-375, typically underexpressed in esophageal

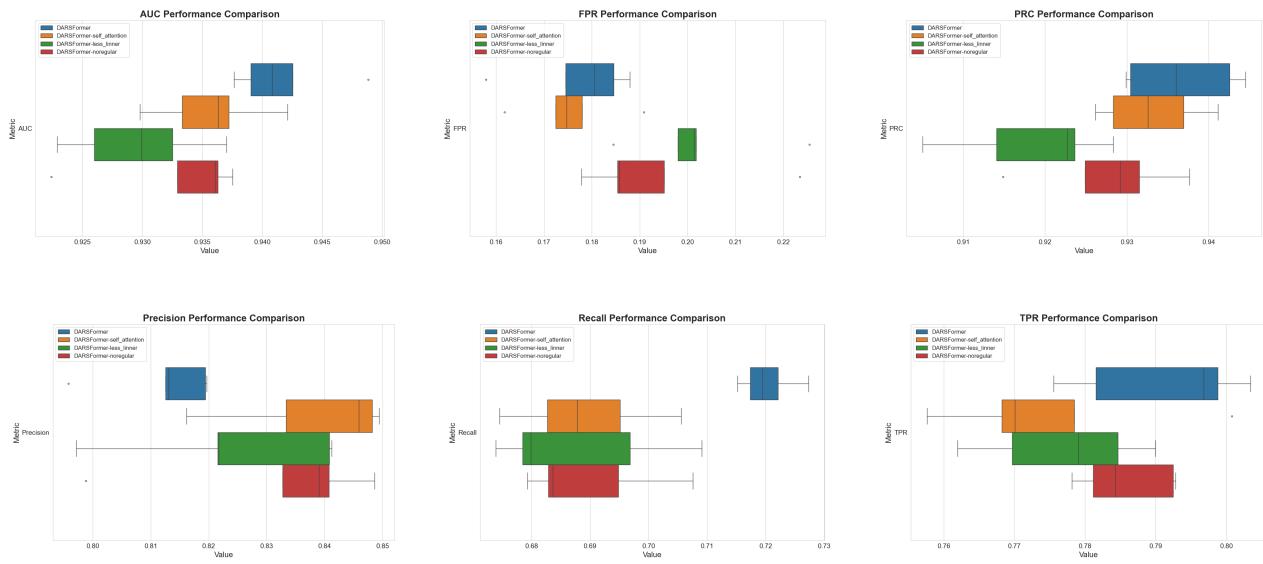


Fig. 7: Box plot with different indicators.

TABLE V: Top 30 colon cancer-related miRNAs predicted

Rank	miRNA	Evidence	Rank	miRNA	Evidence
1	hsa-mir-125b	dbDEMC	16	hsa-mir-196a	dbDEMC
2	hsa-mir-29a	dbDEMC and miR2Disease	17	hsa-mir-214	dbDEMC
3	hsa-mir-34a	dbDEMC and miR2Disease	18	hsa-mir-195	dbDEMC
4	hsa-mir-221	dbDEMC and miR2Disease	19	hsa-mir-30a	dbDEMC
5	hsa-mir-222	dbDEMC	20	hsa-mir-9	dbDEMC
6	hsa-mir-29b	dbDEMC and miR2Disease	21	hsa-mir-17	dbDEMC
7	hsa-mir-133a	dbDEMC and miR2Disease	22	hsa-mir-181b	dbDEMC
8	hsa-mir-199a	Unconfirmed	23	hsa-let-7c	dbDEMC
9	hsa-mir-1	dbDEMC and miR2Disease	24	hsa-let-7a	dbDEMC and miR2Disease
10	hsa-mir-223	dbDEMC and miR2Disease	25	hsa-mir-15b	dbDEMC
11	hsa-mir-145	dbDEMC and miR2Disease	26	hsa-mir-23a	dbDEMC
12	hsa-mir-106b	Unconfirmed	27	hsa-mir-146b	dbDEMC
13	hsa-mir-142	dbDEMC	28	hsa-let-7b	dbDEMC
14	hsa-mir-206	dbDEMC and miR2Disease	29	hsa-mir-34c	miR2Disease
15	hsa-mir-31	dbDEMC	30	hsa-mir-106a	Unconfirmed

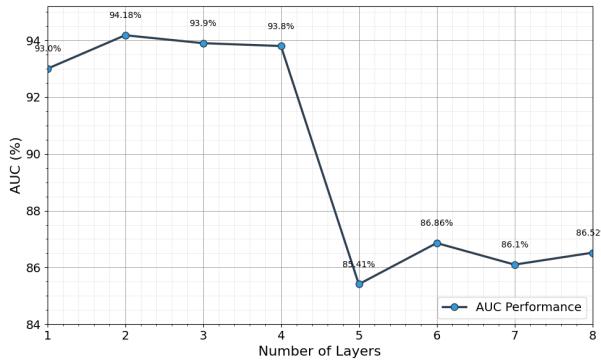


Fig. 8: The value of DARSFormer under different numbers of layers.

cancers and regarded as a potential tumor-suppressive miRNA, especially in esophageal squamous cell carcinoma, targets multiple pathways involved in tumor progression [45].

The prevalence of malignancy, noted as a significant cancer among males, has been extensively studied across various

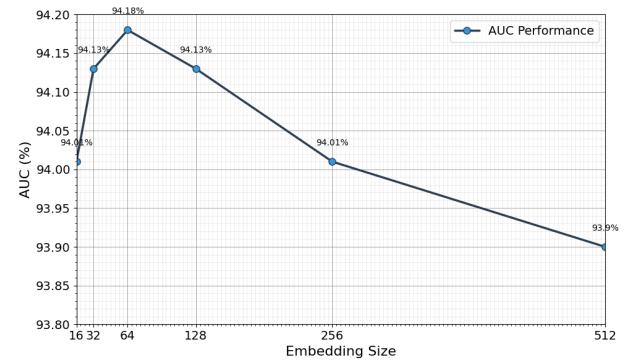


Fig. 9: The value of DARSFormer under different embedding size.

domains ranging from gene regulation and molecular biology to clinical treatments and epidemiological approaches. MicroRNAs (miRNAs) have garnered attention in molecular research due to their pivotal role in the progression of cancer. It has been observed that certain miRNAs are downregulated

TABLE VI: Top 30 Esophageal Cancer-related miRNAs

Rank	miRNA	Evidence	Rank	miRNA	Evidence
1	hsa-mir-16	dbDEMC	16	hsa-mir-142	dbDEMC
2	hsa-mir-221	dbDEMC	17	hsa-mir-9	dbDEMC
3	hsa-mir-17	dbDEMC	18	hsa-mir-24	dbDEMC
4	hsa-mir-125b	dbDEMC	19	hsa-mir-195	dbDEMC
5	hsa-mir-1	dbDEMC	20	hsa-let-7e	dbDEMC
6	hsa-mir-122	Unconfirmed	21	hsa-mir-146b	dbDEMC
7	hsa-mir-19b	dbDEMC	22	hsa-mir-30a	dbDEMC
8	hsa-mir-29b	dbDEMC	23	hsa-mir-181b	dbDEMC
9	hsa-mir-29a	dbDEMC	24	hsa-let-7i	dbDEMC
10	hsa-mir-18a	dbDEMC	25	hsa-let-7d	dbDEMC
11	hsa-mir-222	dbDEMC	26	hsa-mir-182	dbDEMC
12	hsa-mir-133b	dbDEMC	27	hsa-mir-15b	dbDEMC
13	hsa-mir-106b	dbDEMC	28	hsa-let-7g	dbDEMC
14	hsa-mir-200b	dbDEMC	29	hsa-let-7f	Unconfirmed
15	hsa-mir-181a	dbDEMC	30	hsa-mir-124	dbDEMC

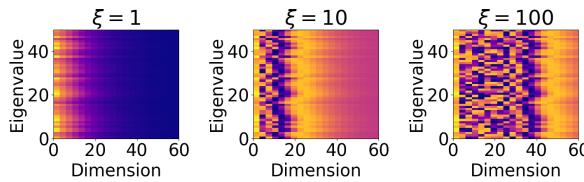


Fig. 10: Eigenvalue encoding with different values of  $\xi$ .

in prostate cancer, potentially serving as tumor suppressors by inhibiting the growth and dissemination of tumor cells. Notably, decreased concentrations of miR-143 and miR-145 are associated with the development of prostate cancer [46], [47]. Conversely, the overexpression of certain miRNAs may facilitate cancer progression by targeting genes that suppress tumors, exemplified by the increased levels of miR-21, which leads to the downregulation of genes such as PTEN [48]. Evidence from Table VII confirms the involvement of 26 out of the top 30 miRNAs with prostate tumors, as verified through the dbDEMC and miR2Disease databases.

#### IV. CONCLUSION

Many human malignancies originate from the dysregulation of gene expression governed by miRNAs. The aberrant expression of miRNAs plays a crucial role in the etiology of various diseases. Therefore, the accurate prediction of the links between diseases and miRNAs is instrumental in advancing human health. In the study presented, a spectral-domain graph Transformer with a dynamic attention mechanism and regularization, termed DARSFormer, is introduced for predicting miRNA-disease associations. By incorporating positional encoding for feature mapping and leveraging the Transformer's inference capabilities for feature representation, this model surpasses most recent models in terms of AUC performance. Case studies focusing on colorectal, esophageal, and prostate neoplasms reveal that 27, 28, and 26 of the top 30 miRNAs, respectively, are confirmed to be associated with these conditions. Future research aims to enhance the model's effectiveness by integrating miRNA sequence information and lncRNA data into the miRNA-disease heterogeneous graph. In

conclusion, the DARSFormer model provides a reliable and significant tool for predicting miRNA-disease associations.

#### V. REFERENCES

- [1] V. Ambros, "The functions of animal microRNAs," *Nature*, vol. 431, no. 7006, pp. 350–355, 2004.
- [2] R. C. Lee and V. Ambros, "An extensive class of small RNAs in *Caenorhabditis elegans*," *Science*, vol. 294, no. 5543, pp. 862–864, 2001.
- [3] W. T. Abplanalp, A. Fischer, D. John, A. M. Zeiher, W. Gosgnach, H. Darville, R. Montgomery, L. Pestano, G. Allée, I. Paty, F. Fougerousse, and S. Dimmeler, "Efficiency and Target Derepression of Anti-miR-92a: Results of a First in Human Study," *Nucleic Acid Ther*, vol. 30, no. 6, pp. 335–345, 2020.
- [4] Z. Shao, J. Xu, X. Wang, Y. Zhou, Y. Wang, Y. Li, J. Zhao, and K. Li, "Exosomes derived from adipose tissues accelerate fibroblasts and keratinocytes proliferation and cutaneous wound healing via miR-92a/Hippo-YAP axis," *J Physiol Biochem*, vol. 80, no. 1, pp. 189–204, 2024.
- [5] A. Kamel, T. Owen, I. Cole, T. Valencia, and E. C. Lee, "Pharmacokinetics and Absorption, Distribution, Metabolism and Excretion of RGLS4326 in Mouse and Monkey, an Anti-miR-17 Oligonucleotide for the Treatment of Polycystic Kidney Disease," *Drug Metab Dispos*, vol. 51, no. 11, pp. 1536–1546, 2023.
- [6] M. Yheskel, R. Lakhia, P. Cobo-Stark, A. Flaten, and V. Patel, "Anti-microRNA screen uncovers miR-17 family within miR-17 92 cluster as the primary driver of kidney cyst growth," *Sci Rep*, vol. 9, no. 1, pp. 1920, 2019.
- [7] W. M. Freeman, S. J. Walker, and K. E. Vrana, "Quantitative RT-PCR: pitfalls and potential," *Biotechniques*, vol. 26, no. 1, pp. 112–122, 124–125, 1999.
- [8] E. Várallyay, J. Burgýán, and Z. Havelda, "MicroRNA detection by northern blotting using locked nucleic acid probes," *Nat Protoc*, vol. 3, no. 2, pp. 190–196, 2008.
- [9] S. Baskerville and D. P. Bartel, "Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes," *RNA*, vol. 11, no. 3, pp. 241–247, 2005.
- [10] Q. Jiang, Y. Hao, G. Wang, L. Juan, T. Zhang, M. Teng, Y. Liu, and Y. Wang, "Prioritization of disease microRNAs through a human phenome-microRNAome network," *BMC Syst Biol*, vol. 4 Suppl 1, no. Suppl 1, pp. S2, 2010.
- [11] H. Zhao, L. Kuang, L. Wang, P. Ping, Z. Xuan, T. Pei, and Z. Wu, "Prediction of microRNA-disease associations based on distance correlation set," *BMC Bioinformatics*, vol. 19, no. 1, pp. 141, 2018.
- [12] X. Chen, Z. C. Jiang, D. Xie, D. S. Huang, Q. Zhao, G. Y. Yan, and Z. H. You, "A novel computational model based on super-disease and miRNA for potential miRNA-disease association prediction," *Mol Biosyst*, vol. 13, no. 6, pp. 1202–1212, 2017.
- [13] X. Chen and G. Y. Yan, "Semi-supervised learning for potential human microRNA-disease associations inference," *Sci Rep*, vol. 4, pp. 5501, 2014.

TABLE VII: Top 30 Prostate Cancer-related miRNAs Predicted

Rank	miRNA	Evidence	Rank	miRNA	Evidence
1	hsa-mir-21	dbDEMC and miR2Disease	16	hsa-mir-122	Unconfirmed
2	hsa-mir-155	dbDEMC	17	hsa-mir-210	miR2Disease
3	hsa-mir-34a	dbDEMC and miR2Disease	18	hsa-mir-15a	dbDEMC and miR2Disease
4	hsa-mir-146a	miR2Disease	19	hsa-mir-222	dbDEMC and miR2Disease
5	hsa-mir-16	dbDEMC and miR2Disease	20	hsa-mir-133b	dbDEMC
6	hsa-mir-20a	miR2Disease	21	hsa-mir-29a	dbDEMC and miR2Disease
7	hsa-mir-92a	Unconfirmed	22	hsa-let-7a	dbDEMC and miR2Disease
8	hsa-mir-200b	Unconfirmed	23	hsa-mir-126	dbDEMC and miR2Disease
9	hsa-mir-17	miR2Disease	24	hsa-mir-29c	dbDEMC
10	hsa-mir-1	dbDEMC	25	hsa-mir-18a	Unconfirmed
11	hsa-mir-221	dbDEMC and miR2Disease	26	hsa-mir-181a	dbDEMC and miR2Disease
12	hsa-mir-19b	dbDEMC and miR2Disease	27	hsa-mir-19a	dbDEMC
13	hsa-mir-29b	dbDEMC and miR2Disease	28	hsa-mir-199a	dbDEMC and miR2Disease
14	hsa-mir-31	dbDEMC and miR2Disease	29	hsa-mir-200a	dbDEMC
15	hsa-mir-133a	dbDEMC	30	hsa-mir-34c	dbDEMC

- [14] P. Xuan, K. Han, M. Guo, Y. Guo, J. Li, J. Ding, Y. Liu, Q. Dai, J. Li, Z. Teng, and Y. Huang, "Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors," *PLoS One*, vol. 8, no. 8, pp. e70204, 2013.
- [15] X. Chen, L. G. Sun, and Y. Zhao, "NCMCMDA: miRNA-disease association prediction through neighborhood constraint matrix completion," *Brief Bioinform*, vol. 22, no. 1, pp. 485–496, 2021.
- [16] T. Zhong, Z. Li, Z. H. You, R. Nie, and H. Zhao, "Predicting miRNA-disease associations based on graph random propagation network and attention network," *Brief Bioinform*, vol. 23, no. 2, pp. bbab589, 2022.
- [17] H. Hu, H. Zhao, T. Zhong, X. Dong, L. Wang, P. Han, and Z. Li, "Adaptive deep propagation graph neural network for predicting miRNA-disease associations," *Brief Funct Genomics*, vol. 22, no. 5, pp. 453–462, 2023.
- [18] Q. Liao, Y. Ye, Z. Li, H. Chen, and L. Zhuo, "Prediction of miRNA-disease associations in microbes based on graph convolutional networks and autoencoders," *Front Microbiol*, vol. 14, pp. 1170559, 2023.
- [19] Y. Li, C. Qiu, J. Tu, B. Geng, J. Yang, T. Jiang, and Q. Cui, "HMDD v2.0: a database for experimentally supported human microRNA and disease associations," *Nucleic Acids Res*, vol. 42, pp. D1070–4, 2014.
- [20] Z. Huang, J. Shi, Y. Gao, C. Cui, S. Zhang, J. Li, Y. Zhou, and Q. Cui, "HMDD v3.0: a database for experimentally supported human microRNA-disease associations," *Nucleic Acids Res*, vol. 47, no. D1, pp. D1013–D1017, 2019.
- [21] Zhao B W, He Y Z, Su X R, et al. Motif-Aware miRNA-Disease Association Prediction Via Hierarchical Attention Network[J]. IEEE Journal of Biomedical and Health Informatics, 2024.
- [22] Li G, Zhao B, Su X, et al. Discovering consensus regions for interpretable identification of rna n6-methyladenosine modification sites via graph contrastive clustering[J]. IEEE Journal of Biomedical and Health Informatics, 2024.
- [23] Yang Y, Su X, Zhao B, et al. Fuzzy-based deep attributed graph clustering[J]. IEEE Transactions on Fuzzy Systems, 2023.
- [24] Peng W, He Z, Dai W, Lan W. MHCLMDA: multihypergraph contrastive learning for miRNA-disease association prediction. *Brief Bioinform*. 2023 Nov 22;25(1):bbad524. doi: 10.1093/bib/bbad524. PMID: 38243694; PMCID: PMC10796254.
- [25] Yan C, Wang J, Ni P, Lan W, Wu FX, Pan Y. DNRLMF-MDA: Predicting microRNA-Disease Associations Based on Similarities of microRNAs and Diseases. *IEEE/ACM Trans Comput Biol Bioinform*. 2019 Jan-Feb;16(1):233-243. doi: 10.1109/TCBB.2017.2776101. Epub 2017 Nov 22. PMID: 29990253.
- [26] D. Wang, J. Wang, M. Lu, F. Song, and Q. Cui, "Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases," *Bioinformatics*, vol. 26, no. 13, pp. 1644–1650, 2010.
- [27] D. Wang, J. Wang, M. Lu, F. Song, and Q. Cui, "Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases," *Bioinformatics*, vol. 26, no. 13, pp. 1644–1650, 2010.
- [28] M. Lu, Q. Zhang, M. Deng, J. Miao, Y. Guo, W. Gao, and Q. Cui, "An analysis of human microRNA and disease associations," *PLoS One*, vol. 3, no. 10, pp. e3420, 2008.
- [29] X. Chen, C. C. Yan, X. Zhang, Z. H. You, L. Deng, Y. Liu, Y. Zhang, and Q. Dai, "WBSMDA: Within and Between Score for MiRNA-Disease Association prediction," *Sci Rep*, vol. 6, pp. 21106, 2016.
- [30] M. Chen, Y. Peng, A. Li, Z. Li, Y. Deng, W. Liu, B. Liao, and C. Dai, "A novel information diffusion method based on network consistency for identifying disease related microRNAs," *RSC Adv*, vol. 8, no. 64, pp. 36675–36690, 2018.
- [31] D. Bo, C. Shi, L. Wang, and R. Liao, "Specformer: Spectral graph neural networks meet transformers," *arXiv preprint*, arXiv:2303.01028, 2023.
- [32] K. Guo, K. Zhou, X. Hu, Y. Li, Y. Chang, and X. Wang, "Orthogonal graph neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 4, pp. 3996–4004, June 2022.
- [33] A. Trockman and J. Z. Kolter, "Orthogonalizing convolutional layers with the cayley transform," *arXiv preprint*, arXiv:2104.07167, 2021.
- [34] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?," *arXiv preprint*, arXiv:2105.14491, 2021.
- [35] Zhao, Huan, et al. "Predicting miRNA-disease associations based on neighbor selection graph attention networks." *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 20.2 (2022): 1298-1307.
- [36] Z. Li, J. Li, R. Nie, Z. H. You, and W. Bao, "A graph auto-encoder model for miRNA-disease associations prediction," *Brief Bioinform*, vol. 22, no. 4, pp. bbaa240, 2021.
- [37] L. Wang, Y. Chen, N. Zhang, W. Chen, Y. Zhang, and R. Gao, "QIMCMDA: MiRNA-Disease Association Prediction by q-Kernel Information and Matrix Completion," *Front Genet*, vol. 11, pp. 594796, 2020.
- [38] P. Sun, S. Yang, Y. Cao, R. Cheng, and S. Han, "Prediction of Potential Associations Between miRNAs and Diseases Based on Matrix Decomposition," *Front Genet*, vol. 11, pp. 598185, 2020.
- [39] Li, Zhengwei, et al. "Hierarchical graph attention network for miRNA-disease association prediction." *Molecular Therapy* 30.4 (2022): 1775-1786.
- [40] G. Li, J. Luo, Q. Xiao, C. Liang, and P. Ding, "Predicting microRNA-disease associations using label propagation based on linear neighborhood similarity," *J Biomed Inform*, vol. 82, pp. 169–177, 2018.
- [41] L. Yu, X. Shen, D. Zhong, and J. Yang, "Three-Layer Heterogeneous Network Combined With Unbalanced Random Walk for miRNA-Disease Association Prediction," *Front Genet*, vol. 10, pp. 1316, 2020.
- [42] A. Bahreyni, M. Rezaei, A. Bahrami, M. Khazaei, H. Fuji, M. Ryzhikov, G. A. Ferns, A. Avan, and S. M. Hassanian, "Diagnostic, prognostic, and therapeutic potency of microRNA 21 in the pathogenesis of colon cancer, current status and prospective," *J Cell Physiol*, vol. 234, no. 6, pp. 8075–8081, 2019.
- [43] H. Tazawa, N. Tsuchiya, M. Izumiya, and H. Nakagama, "Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells," *Proc Natl Acad Sci U S A*, vol. 104, no. 39, pp. 15472–15477, 2007.
- [44] V. Coppola, M. Musumeci, M. Patrizii, A. Cannistraci, A. Addario, M. Maugeri-Saccà, M. Biffoni, F. Francescangeli, M. Cordenonsi, S. Piccolo, L. Memeo, A. Pagliuca, G. Muto, A. Zeuner, R. De Maria, and D. Bonci, "BTG2 loss and miR-21 upregulation contribute to prostate cell transformation by inducing luminal markers expression and epithelial-mesenchymal transition," *Oncogene*, vol. 32, no. 14, pp. 1843–1853, 2013.
- [45] K. L. Kong, D. L. Kwong, T. H. Chan, S. Y. Law, L. Chen, Y. Li, Y. R. Qin, and X. Y. Guan, "MicroRNA-375 inhibits tumour growth and

- metastasis in oesophageal squamous cell carcinoma through repressing insulin-like growth factor 1 receptor," *Gut*, vol. 61, no. 1, pp. 33–42, 2012.
- [46] Y. Zhang, Z. Wang, M. Chen, L. Peng, X. Wang, Q. Ma, F. Ma, and B. Jiang, "MicroRNA-143 targets MACC1 to inhibit cell invasion and migration in colorectal cancer," *Mol Cancer*, vol. 11, pp. 23, 2012.
- [47] J. Qin, F. Wang, H. Jiang, J. Xu, Y. Jiang, and Z. Wang, "MicroRNA-145 suppresses cell migration and invasion by targeting paxillin in human colorectal cancer cells," *Int J Clin Exp Pathol*, vol. 8, no. 2, pp. 1328–1340, 2015.
- [48] T. Li, D. Li, J. Sha, P. Sun, and Y. Huang, "MicroRNA-21 directly targets MARCKS and promotes apoptosis resistance and invasion in prostate cancer cells," *Biochem Biophys Res Commun*, vol. 383, no. 3, pp. 280–285, 2009.
- [49] Yang Z, Wu L, Wang A, et al. dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. Nucleic Acids Res 2017;45:D812–8.
- [50] Jiang Q, Wang Y, Hao Y, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res 2009;37:D98–104

## VI. CONFLICT OF INTEREST

The author declares that there are no conflicts of interest regarding the publication of this paper.