Article

DeepmiRNATar: A deep learning-based model for miRNA targets prediction

Huimin Peng, Chenyu Li, Ying Lu, Dazhou Li*

School of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang 110142, China
* Corresponding author: Dazhou Li, lidazhou@syuct.edu.cn

Abstract: MicroRNAs (miRNAs) play a crucial role in regulating fundamental biological processes such as the cell cycle, differentiation, and apoptosis by directly interacting with multiple genes (mRNAs). This regulatory mechanism has a profound impact on cellular function and the overall physiological condition of an organism. However, the prediction of miRNA-mRNA interactions encounters computational challenges in the field of biology due to the diverse sequences and complex data patterns. To overcome these obstacles, this research effort introduced DeepmiRNATar, a tool designed to precisely pinpoint miRNA targets, offering essential assistance in the realm of disease management. DeepmiRNATar leverages the Word2vec-based DeepLncLoc approach for encoding miRNA sequence characteristics and utilizes the DNABERT pre-trained model for in-depth semantic comprehension of target sequences. Through the integration of TextCNN, BiLSTM, and SpatialConv Attention mechanisms, the model scrutinizes structural features, temporal relationships, and overall interactions within the sequences. Following a series of experimental assessments, DeepmiRNATar attained an impressive AUC of 99.15% on the evaluation dataset, on par with the current leading prediction methodologies. Notably, the precision-recall curve, sensitivity, and *F*-measure values reached 99.18%, 97.43%, and 95.47%, respectively. Compared to existing miRNA target prediction models, DeepmiRNATar demonstrates a notable enhancement in overall predictive accuracy. The successful creation and experimental validation of the DeepmiRNATar model signify a significant advancement in miRNA target identification technology.

Keywords: MiRNA-mRNAs prediction; deep learning; TextCNN; BiLSTM; SpatialConv attention
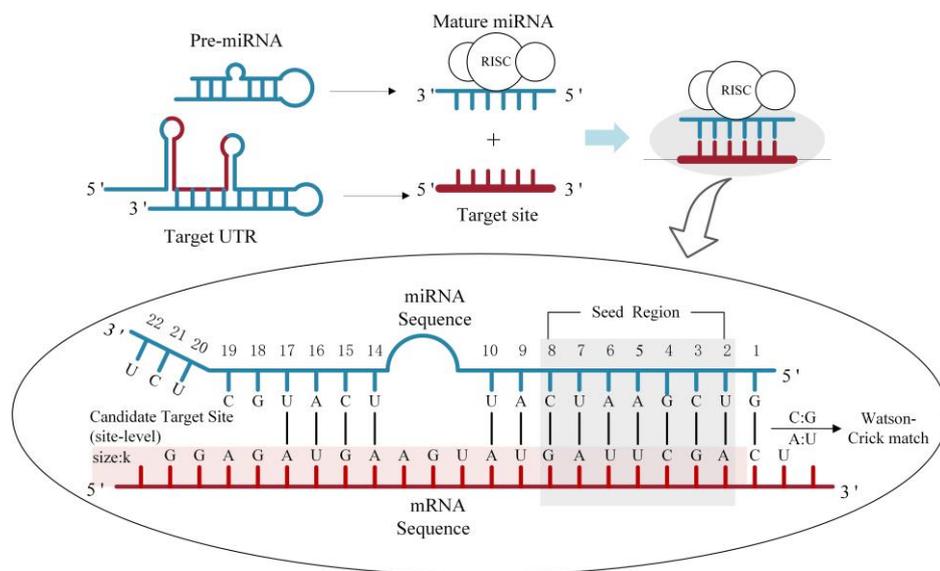
1. Introduction

MicroRNAs (miRNAs) are a group of non-coding single-stranded RNA molecules approximately 22 nucleotides long, found widely in plants and animals [1]. Studies suggest that miRNAs may control around one-third of human genes [2], and their abnormal levels are directly linked to the onset of various diseases [3], including cancer and cardiovascular conditions. For example, miR-21 is excessively expressed in multiple cancers, stimulating tumor growth and spread by suppressing tumor suppressor gene expression [4]. Similarly, irregular miR-155 expression is associated with different inflammatory diseases and specific cancer types [5]. miRNAs typically develop complementary binding sequences with targets (mRNAs), acting after transcription to inhibit gene expression or mRNA cleavage [6]. Understanding the interactions between miRNAs and their targets is crucial for improving the accuracy of mRNA vaccine design and the identification of drug targets in drug development [7].

Given the importance of miRNAs in cellular regulation, multiple computational techniques have been proposed to tackle the miRNA target prediction challenge [6–10]. Presently, miRNA target identification heavily relies on biological experiments and computational predictions. Traditional biological methods for miRNA target identification, while highly precise, are expensive and time-intensive. Early software simulations like miRanda [6], TargetScan [7], and Probability of Interaction by Target Accessibility (PITA) [8], primarily aimed to improve prediction performance by enhancing seed sequence matching and target accessibility. However, these methods face significant issues with high false positive rates. This problem arises because their algorithms rely on fixed seed sequence matching rules, which are inadequate for capturing the complex interactions between miRNAs and mRNAs. To boost prediction accuracy, machine learning methods were subsequently introduced. These methods used manually crafted feature descriptors like those in TargetSpy [9] and mirSVR [10]. While these methods offered some improvements, they were limited by their reliance on predefined features, which could not fully capture the complexity of miRNA-mRNA interactions. Manually defined feature descriptors could not entirely capture the intricate interactions between miRNAs and target sequences.

With advancements in technology, deep learning methods have gained widespread application in miRNA target prediction [11]. MiRTDL [12] utilizes the robust feature extraction capabilities of convolutional neural networks (CNNs) with 20 features to forecast miRNA targets. While this enhances automatic feature evaluation, the outcomes remain limited by these predetermined features. DeepTarget [13] employs deep recurrent neural networks (RNNs) to discern interactions from miRNA-mRNA sequence data, thereby circumventing the dependence on predefined features. However, RNNs might not efficiently capture the spatial features intrinsic to miRNA-mRNA interactions. DeepMirTar [14] adopts a strategy involving stacked denoising autoencoders to process 750 features, yet it still concentrates on traditional feature categories, failing to transcend their inherent constraints. miRAW [15] integrates autoencoders with feedforward networks to learn miRNA-mRNA interaction features from raw data directly, but feedforward networks struggle with the spatial and sequence complexity of this data. Additionally, miTAR [16] is a hybrid miRNA target prediction model that comprises six layers, combining convolutional neural networks (CNNs) with bidirectional RNNs (BiRNN). However, miTAR employs concatenated sequence inputs, which makes it challenging for the model to accurately differentiate and capture the specific interaction features between miRNAs and targets. TEC-mitar [17] merges CNNs with Transformer encoders to process sequence data. By generating a "contact map," this model adeptly manages the variations and products between miRNA and candidate target site sequences (CTS), thereby predicting interaction probabilities. However, this model relies on complex sequence transformations, which may result in high computational resource consumption when handling high-dimensional data.

The interactions between miRNAs and their targets hinge on a sequence known as the seed region within the miRNA molecule, encompassing bases 2–8 at the 5' end of the miRNA [18]. This seed region pairs with a complementary sequence in the 3' Untranslated Region (3' UTR) of the mRNA (**Figure 1**), leading to mRNA degradation or translational inhibition, thereby controlling gene expression levels.

Most predictive models take only the seed region sequence into account. However, recent research indicates that many targets involve nucleotides beyond the seed region, implying that target prediction algorithms should consider the entire miRNA sequence [19,20]. Consequently, this study employs deep learning methods to analyze the entire mature miRNA transcript, rather than focusing solely on the seed region. By accounting for the entire miRNA sequence, the model can identify key interaction sites beyond the seed region, enhancing the accuracy and scope of predictions. This approach allows identification of more potential miRNA-mRNA interactions, uncovering more complex regulatory mechanisms.

Figure 1. MiRNA biogenesis and functional mechanism.

This paper introduces a deep learning-based method named DeepmiRNATar, designed to enhance miRNA target prediction. It leverages DeeplocLnc and DNABERT to process miRNA and target sequences, respectively, and employs Text Convolutional Neural Network (TextCNN) and Bidirectional Long Short-Term Memory (BiLSTM) to extract local and global sequence features. An advanced multi-head attention mechanism is utilized for target prediction. Compared to a rule-based software simulation method, a machine learning-based approach, and three recently developed deep learning models, DeepmiRNATar showed superior performance across nearly all evaluation metrics. DeepmiRNATar not only significantly improves prediction accuracy but also eliminates the need for cumbersome manual feature extraction, highlighting the vast potential of deep learning in tackling complex bioinformatics challenges.

2. Materials and methods

2.1. Dataset

The dataset employed in this research has been sourced from DeepMirTar and miRAW studies. DeepMirTar offers miRNA-gene interaction data with thorough annotations at both the site and gene levels [14]. miRAW provides an extensive validated dataset, containing both canonical and non-canonical miRNA-mRNA

interaction pairs [15]. Positive pairs in these datasets are derived from reputable biological databases such as miRTarBase. For DeepMirTar, negative pairs are created by arbitrarily shuffling actual mature miRNA sequences, whereas for miRAW, negative pairs are chosen from mRNA 3' UTR regions that can form stable miRNA bindings. The miRNA duplex secondary structure is forecasted using RNACoFold, with negative binding energy serving as a criterion for identifying potential negative sites. For comprehensive details on the dataset generation process, please refer to DeepMirTar [14] and miRAW [15].

The DeepMirTar dataset consists of two categorizations of miRNA-mRNA information: site-level and gene-level datasets (**Figure 2**). In the site-level dataset, it is assumed that a gene sequence contains multiple candidate target sites (CTSs), which are potential miRNA binding sites, some of which are true miRNA targets. Each CTS is marked to show whether it is an authentic miRNA target. Conversely, in the gene-level dataset, CTSs are not individually marked; instead, the entire gene sequence is labeled. The site-level dataset is used to assess prediction model accuracy at the CTS level, whereas the gene-level dataset evaluates performance at the gene level. Given that a single miRNA can interact with multiple sites on the mRNA sequence of a target gene, this study opts to use site-level datasets to ensure a thorough identification of all potential target sites. This approach offers more detailed information on the target sites, preventing the omission of essential miRNA-mRNA interactions.
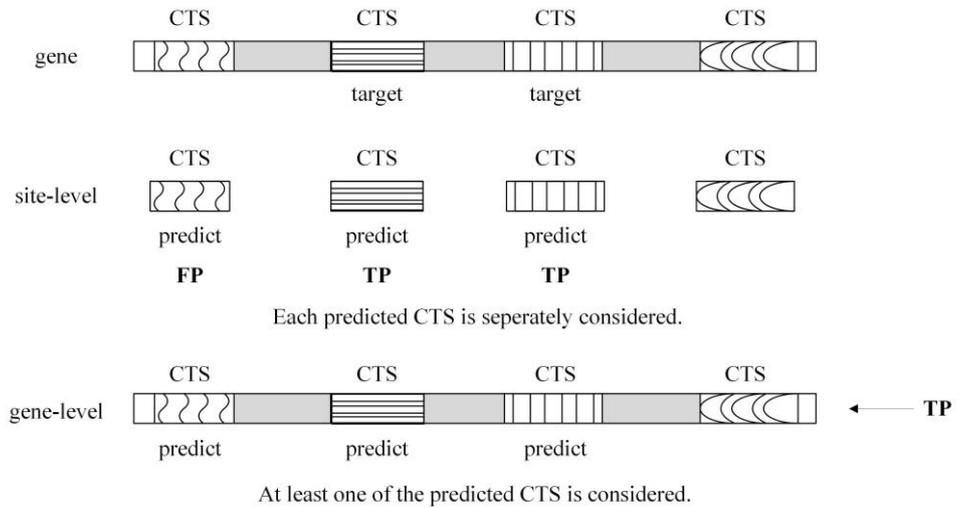
Figure 2. Definition of dataset for site-level and gene-level targeting.

Table 1. DeepmiRNATar dataset overview.

number	miRNA name	mRNA name	miRNA sequence	mRNA sequence	label
1	hsa-miR-520d-3p	CDH1	AAAGUGCUUC...	AGCTCCCCAA...	0
2	hsa-miR-19a-3p	ZNF217	UGUGCAAAUC...	TACAGTTGTG...	1
3	hsa-miR-301b-3p	BTBD3	CAGUGCAAUG...	AGTAAATGGT...	1
4	hsa-miR-145-5p	SPARC	GUCCAGUUUU...	AGTAATGACT...	0
5	hsa-miR-141-3p	EDEM1	UAACACUGUC...	TTTTCATGTT...	1

The miRAW dataset includes validated miRNA-mRNA pairs from Diana TarBase [21] and MirTarBase [22], offering a substantial number of positive and negative samples. Initially, the datasets from DeepMirTar and miRAW were combined, and duplicates were removed. Subsequently, the miRNA sequences from the miRAW database were cross-verified with the latest miRBase version, removing any unlisted miRNA entries and we ensure that there is no duplication of data between the training set and the test sets to ensure the integrity of the evaluation process. Following these processes, a final dataset of 55,722 miRNA-mRNA interaction pairs was compiled. **Table 1** provides a sample of this dataset.

In addition, this paper obtained 48 positive sample data from the PAR-CLIP experiment study by Hafner et al. [23] as an independent dataset, which was used during the testing phase to examine the model's ability to identify and recognize positive samples. This independent dataset has been used in the studies by Ding et al. [24] and Wang et al. [25]. The PAR-CLIP experiment offers significant advantages in microRNA target gene analysis. By introducing photo-crosslinking technology, it significantly reduces the occurrence of false-positive predictions and greatly narrows the search range for microRNA binding sites. Therefore, PAR-CLIP data can more precisely identify the binding sites of AGO2 protein and RNA across the genome.

2.2. Data processing

2.2.1. miRNA sequence data processing

For feature extraction of miRNA sequences, typical methods include one-hot encoding and *k*-mer feature representation. One-hot encoding translates each nucleotide into a four-dimensional vector but fails to capture nucleotide relationships, leading to sparse data. The *k*-mer feature accounts for the frequency of *k* nucleotides in the sequence but overlooks positional information.To address these limitations, this study uses the DeepLocLnc [26] method for miRNA sequence encoding.

In **Figure 3**, for a given miRNA nucleotide sequence S_{mi} and a positive integer k, S_{mi} can be segmented into several miRNA nucleotide subsequences of length k using a "sliding window" approach. As k increases, the k-mers can span over more short-length repeated sequences, but a larger k value results in fewer k-mers. The interconnections between k-mers mean that a reduction in the number of k-mers can hinder the extraction of features from nucleotides at different positions.
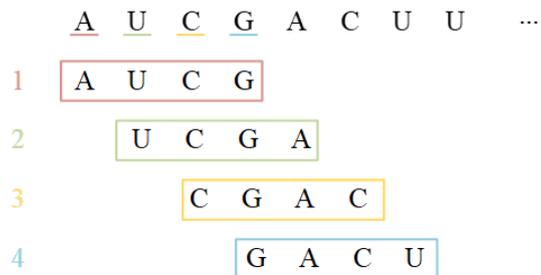
Figure 3. Schematic diagram of the "sliding window" word segmentation operation.

To address these issues, this study chooses a *k* value of 4. This choice is based on the *k*-value settings discussed in references [27,28] and takes into account the length

of microRNA sequences. Setting *k* to 4 allows for the generation of overlapping *k*-mers, which captures sufficient sequence information while avoiding the feature sparsity issues associated with larger *k* values. This *k*-mer selection method effectively maintains local detail information in the sequence while capturing long-range dependencies, balancing the trade-off between redundant features and information sparsity.

This approach utilizes the gensim library to learn *k*-mer representation of miRNA sequences, using the word2vec algorithm to generate vector representations for each *k*-mer. The sequence is divided into *k*-mers using a sliding window, and pre-trained vectors are assigned to them. An average pooling layer extracts features, and subsequence vectors are merged to create a concise sequence representation, which serves as input for the deep learning model. This method avoids the sparsity of one-hot encoding and preserves contextual information via word2vec. In this study, *k* is set to 4, and the miRNA sequence feature representation dimension is set to 34.

2.2.2. Target sequence data processing

This study employs the DNABERT model for encoding target sequences, which is proficient in directly visualizing nucleotide-level significance and semantic relationships within sequences [29]. The diversity of regulatory codes is pivotal to gene expression regulation, varying by cell type and organism. DNABERT aims to explore this diversity comprehensively, capturing the semantic information of sequences using a self-attention mechanism.

The BERT model comprises a stack of multi-layer Transformer encoders. The architecture used is identical to the original BERT, encompassing 12 Transformer encoding layers, each with 768 hidden units and 12 attention heads. DNABERT processes sequences represented as *k*-mer tokens. By embedding each token into numerical vectors, sequences are represented as a matrix *E*. DNABERT captures contextual information by employing multi-head self-attention on *E*:

$$head_i = \text{softmax}\left[\frac{EW_i^Q(EW_i^K)^T}{\sqrt{d_k}}\right]EW_i^V \tag{1}$$

$$MultiHead(E) = \text{Concat}(head_1, \dots, head_h)W^O \tag{2}$$

W^O and $W_i^Q, W_i^K, W_i^V, \{W_i^Q, W_i^K, W_i^V\}_{i=0}^{b}$ are the learned parameters for linear projections, and $head_i$ represents the output of a single attention head. Attention on each subspace calculates the similarity between two *k*-mers using scaled dot-product attention. The similarity coefficients are then employed to linearly combine corresponding word vectors, resulting in the output. Feature matrices from multiple subspaces are concatenated for a comprehensive feature representation.

During the pre-training phase, the objective task utilizes the traditional Masked Language Model (MLM). Given the specificity of gene sequences, if only a single *k*-mer fragment is masked, it can be readily inferred from adjacent *k*-mer fragments. This simplifies the pre-training task, hindering deep semantic relationship learning. Consequently, unlike the original MLM, consecutive *k* words (*k*-mer fragments) are masked during pre-training, conducted in a self-supervised manner without manual labels, facilitating the extraction of implicit general knowledge from unlabeled data.

In this study, *k* is set to 4, and the target sequence feature representation dimension is set to 96.

2.3. The architecture of DeepmiRNATar

The DeepmiRNATar model is based on an encoder-decoder framework, leveraging advanced deep learning methodologies to model and analyze the intricate interactions between miRNA and target sequences. Initially, miRNA and target sequences are regarded as natural language texts, and feature information is extracted from both using the DeepLncLoc technique and the DNABERT pre-trained model. The extracted feature representations are then input into the encoder, where they undergo interactions and weight calculations through the attention mechanism. These features are further processed by a multi-layer feedforward network and layer normalization in the decoder. Finally, the Softmax function is employed to predict the interaction probability between miRNA and their targets. The architecture of the DeepmiRNATar model is depicted in **Figure 4**.
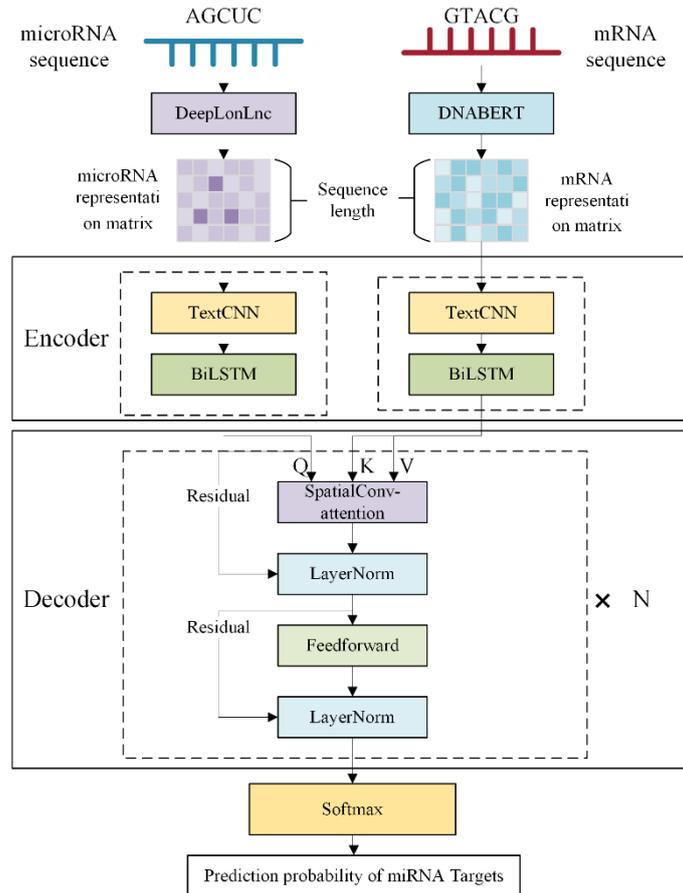
Figure 4. DeepmiRNATar model architecture diagram.

Within the encoder phase, TextCNN is employed to extract local features from sequences through one-dimensional convolutional kernels, effectively encoding information in miRNA and gene sequences. Subsequently, the BiLSTM layer processes these sequences in parallel to capture long-term dependencies, ensuring comprehensive consideration of the contextual information. This tandem architecture

leverages convolutional neural networks' feature extraction capabilities while utilizing BiLSTM to capture positional and directional information in the sequences.

The decoder consists of four modules, each comprising a multi-head self-attention layer, augmented by a convolution operation and a feedforward network layer. The application of layer normalization and residual connections markedly enhances the model's ability to handle gradient instability and maintain stability during training. This architectural design improves the model's accuracy in identifying complex biological sequence patterns and accelerates convergence.

The DeepmiRNATar algorithm is elaborated as follows:

**Algorithm 1** Pseudo-code of DeepmiRNATar

---

1: **Input:** N miRNA-mRNA pairs $(S_{mi}^1, S_m^1), ..., (S_{mi}^N, S_m^N)$ and labels $(l_1, ..., l_m)$, where in the ith pair comprises one miRNA sequence S_{mi}^i, one mRNA sequence S_m^i, and the corresponding labels l_i.

2: **Output:** y_pred

3: 1: For each miRNA and mRNA pair, this study uses the pre-trained DeepLocLnc model to extract miRNA features and the DNABERT model to extract mRNA features, resulting in the encoding matrices $\boldsymbol{X\_mi}$ and $\boldsymbol{X\_m}$ (where L_{mi} and L_m represent the lengths of the miRNA and mRNA sequences, and k represents the k-mer size, i.e., the number of bases in each segment). $\triangleright [x_{mi}^1; ...; x_{mi}^{L_{mi}-k+1}]$ and $[x_m^1; ...; x_m^{L_m-k+1}]$

4: 2: Input $\boldsymbol{X\_mi}$ and $\boldsymbol{X\_m}$ into TextCNN and BiLSTM to obtain the encoded feature matrix $\boldsymbol{H\_mi}$ and $\boldsymbol{H\_m}$, respectively. $\triangleright [h_{mi}^1; ...; h_{mi}^{L_{mi}-k+1}]$ and $[h_m^1; ...; h_m^{L_m-k+1}]$

5: 3: Input $\boldsymbol{H\_mi}$ and $\boldsymbol{H\_m}$ into the fusion convolutional multi-head attention and feedforward neural network to generate a comprehensive feature representation that captures the potential interactions between miRNAs and their mRNApairs. $\triangleright [O_{dec}^1; ...; O_{dec}^{L_{\max\_mi}-k+1}]$

6: 4: Obtain the final predicted probability y_{pred} through the softmax function. $\triangleright y_{pred}$

7: 5: During training, the cross-entropy loss function is used to minimize the difference between the model predictions and the true interaction labels. The loss function is defined as $Loss = -(y \times \log(y_{pred}) + (1-y) \times \log(1-y_{pred}))$, where y is the true label and y_{pred} is the model's predicted output.

8: 6: Optimize the model parameters through multiple iterations (i.e., training epochs) until the loss function converges or reaches the preset maximum number of iterations.

The model utilizes convolutional layers to extract key local features and integrates BiLSTM to capture the bidirectional dependencies within sequence data. Additionally, a self-attention mechanism is implemented to thoroughly analyze global interaction patterns across the entire sequence, incorporating spatial convolution techniques to account for the spatial properties and secondary structure. This multidimensional deep learning strategy significantly enhances the understanding of sequence structure and spatial characteristics, thereby improving the accuracy of miRNA-mRNA interaction predictions.

2.3.1. Encoder

The first sub-layer of the encoder adopts an improved TextCNN architecture (**Figure 5**). Traditionally, the classic TextCNN structure includes one-dimensional convolutional layers and max-pooling layers to effectively capture key sequence features [30]. However, to better retain information from miRNA and target sequences, the max-pooling layers are omitted from the TextCNN design. Various scales of convolutional kernels sample miRNA and target sequences through a sliding window,

obtaining local features of different ranges to capture semantic information at multiple levels within the sequences.
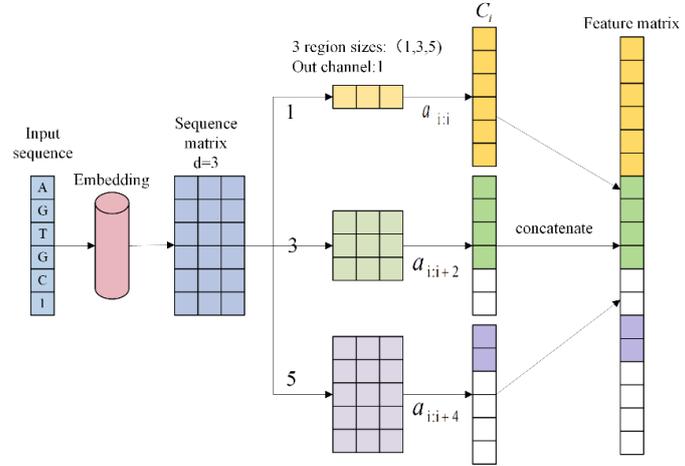
Figure 5. The structure diagram of TextCNN network.

The convolutional layer is fundamental to TextCNN. In experiments, three convolutional kernel sizes—1, 3, and 5—were chosen to extract diverse text features. When the kernel size is 1, it can capture fine-grained local features; with kernel sizes of 3 and 5, it can capture broader contextual information, thereby enhancing the diversity of feature representation. Padding is applied using the formula $padding = k\_size//2$ to maintain the same output length post-convolution for subsequent operations. The convolutional layer computation formula is as follows:

$$c_i = f\big(\omega_i \times a_{i:i+k\_size-1} + b_i\big) \tag{3}$$

Here, f denotes a nonlinear function, typically ReLU. ω_i represents the weights corresponding to the i-th node of the output matrix. $a_{i:i+k\_size-1}$ is the subsequence of the input miRNA and target sequences, with a length equal to k_size, corresponding to the receptive field of the convolutional kernel. b_i is the corresponding bias, and c_i is the result after convolution. The feature vectors of the miRNA sequence S_{mi} and the target sequence S_m processed by TextCNN can be formally described as follows:

$$C_{mi} = T(S_{mi}) = TextCNN(x_{mi}) \tag{4}$$

$$C_m = T(S_m) = TextCNN(x_m) \tag{5}$$

Here, $C_{mi}, C_m \in \mathbb{R}^{d_t}$ represent the output feature vectors of S_{mi} and S_m respectively, where d_t is the dimension of the feature vector. T denotes the improved TextCNN model. x_{mi} is the output of the miRNA sequence S_{mi} after being processed by the DeepLncLoc model, and x_m is the output of the target sequence S_m after being processed by the DNABERT model.

The second sub-layer of the encoder in the DeepmiRNATar model employs a BiLSTM. BiLSTM, which enhances the original LSTM by considering both forward and backward inputs simultaneously, is crucial for capturing contextual relationships in sequence data. Thus, DeepmiRNATar utilizes BiLSTM to capture deep sequential features, thereby improving expressiveness. The feature vectors of the miRNA

sequence *S_mi* and the target sequence *S_m* processed by BiLSTM can be formally described as:

$$h_{mi} = L(S_{mi}) = \overrightarrow{LSTM}(c_{mi}) \oplus \overleftarrow{LSTM}(c_{mi}) \tag{6}$$

$$h_m = L(S_m) = \overrightarrow{LSTM}(c_m) \oplus \overleftarrow{LSTM}(c_m) \tag{7}$$

Here, $h_{mi}, h_m \in \mathbb{R}^{d_l}$ represent the output feature vectors of S_{mi} and S_m, respectively, where d_l is the dimension of the feature vector. L denotes the BiLSTM model. C_{mi} and C_m are the feature vectors of the miRNA sequence S_{mi} and the target sequence S_m after being processed by TextCNN. $\overrightarrow{LSTM}(\cdot)$ and $\overleftarrow{LSTM}(\cdot)$ capture the latent interactions in the context in the forward and backward directions, respectively. The symbol $\oplus$ denotes the concatenation of the outputs from the forward and backward LSTM units.

2.3.2. Decoder

When miRNA binds to its target, a double-stranded structure indicating the spatial properties of miRNA-mRNA interactions forms (**Figure 6**). CNNs are designed to capture spatial features. To effectively discern complex miRNA and target sequence relationships, the decoder was enhanced by integrating convolution operations into the standard multi-head attention layer, naming this innovative module SpatialConv Attention (**Figure 7**).
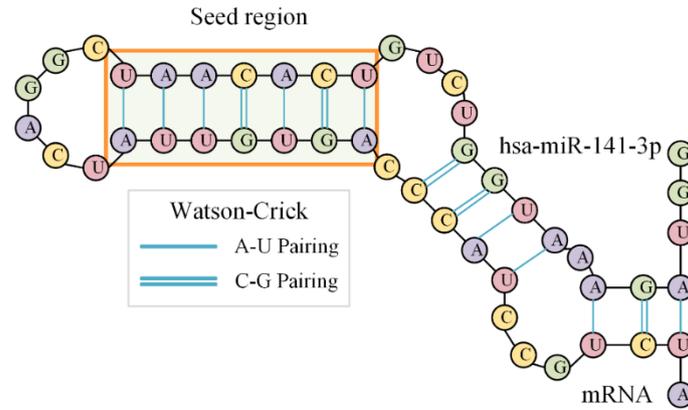
Figure 6. Secondary structure diagram of miRNA and target binding.

In the DeepmiRNATar model, SpatialConv Attention is a core sub-layer of the decoder. This design merges the multi-head attention mechanism with convolution operations to enhance feature capture capabilities, allowing the model to focus on multiple key areas of the encoder output [31]. Through an additive fusion operation, multi-head attention results are combined with convolution output. This fusion retains fine-grained information from multi-head attention while integrating spatial feature recognition from convolution.
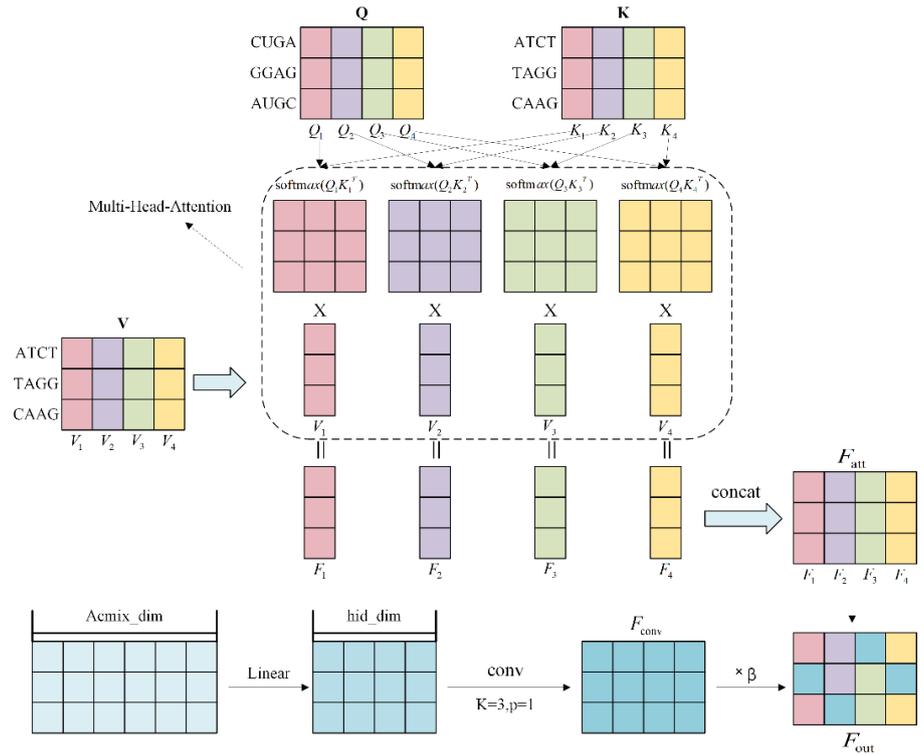
Figure 7. The structural diagram of the SpatialConv attention network.

This design ensures the model balances multi-head attention and convolution contributions, enhancing representation of complex miRNA-target interactions. A learnable scaling parameter β adjustably weights the contributions of multi-head attention and convolution, enabling the model to optimally extract key features from miRNA and target sequences.

$$F_{out} = F_{att} + \beta F_{conv} \tag{8}$$

The decoder's second sub-layer is a feedforward neural network layer. This layer, consisting of two fully connected layers and utilizing the ReLU function as the nonlinear activation function, is depicted in Equation (9). The output from the SpatialConv Attention operation is further refined by the feedforward neural network layer. This process amplifies the model's capacity to spatially represent miRNA sequences and their target sequences through increased nonlinear processing.

$$FFN = \text{Max}(0, XW_1 + b_1)W_2 + b_2 \tag{9}$$

Here, X represents the output matrix after SpatialConv Attention, W_1 and W_2 represent the weight matrices, and b_1 and b_2 represent the biases of the network.

As the network depth increases, significant changes in the internal data distribution may occur, sometimes resulting in gradient vanishing or explosion, which can compromise training stability. To address these challenges, the model incorporates two residual connections and two layer normalization (LN) steps to enhance training stability. Residual connections combine the input and output of a sub-layer via addition, helping to prevent gradient issues in deep networks. Layer normalization standardizes the sub-layer output distribution, hastening the model's convergence.

These operations for residual connections and layer normalization are detailed in Equation (10):

$$Output = LN\big(X + SubLayer(X)\big) \tag{10}$$

Here, X represents the input to each sub-layer, and $SubLayer(X)$ represents the output of the sub-layer itself. The calculation method for LN is described by Equation (11):

$$f(x) = \alpha \frac{x - \mu}{\sigma + \varepsilon} + \beta \tag{11}$$

where μ and σ are the mean and standard deviation of X, respectively, ε is a small constant added for numerical stability, and α and β are learnable parameters that scale and shift the normalized output. This normalization helps stabilize the training process by maintaining consistent distributions of the sub-layer outputs.

2.4. Performance evaluation

To comprehensively assess the classification model's performance, multiple metrics were employed, covering various aspects of model efficacy. This study evaluates DeepmiRNATar's performance on the test set from four perspectives: Area Under the Curve (*AUC*), Precision-Recall Curve (*PRC*), sensitivity, and *F*-measure.

2.4.1. AUC

As the area under the ROC curve, *AUC* gauges the model's capability to distinguish between positive and negative samples across various classification thresholds. It is a threshold-independent metric. The horizontal axis denotes the false positive rate (*FPR*), while the vertical axis denotes the true positive rate (*TPR*). The *FPR* and *TPR* are calculated using Equations (12) and (13):

$$TPR = \frac{TP}{TP + FN} \tag{12}$$

$$FPR = \frac{FP}{FP + TN} \tag{13}$$

In this study, true positive (*TP*) denotes instances where the model correctly predicts an interaction between miRNA and the target (predicted as 1, actually 1). True negative (*TN*) denotes instances where the model accurately predicts no interaction (predicted as 0, actually 0). False positive (*FP*) represents instances where the model erroneously predicts an interaction (predicted as 1, actually 0). False negative (*FN*) represents instances where the model falsely predicts no interaction (predicted as 0, actually 1).

2.4.2. PRC

This reflects the relationship between Precision and Recall, particularly suitable for datasets with highly imbalanced positive and negative samples. Recall is calculated as shown in Equation (12), while Precision is calculated as shown in Equation (14):

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

2.4.3. Sensitivity

In predicting miRNA-mRNA interactions, high sensitivity indicates the model's efficacy in correctly identifying true interactions, crucial for avoiding the omission of potentially significant biological signals. The sensitivity calculation method is shown in Equation (15):

$$sensitivity = \frac{TP}{TP + FN} \tag{15}$$

2.4.4. F-measure

F-measure, the harmonic mean of Precision and Recall, offers a balanced metric between precision and recall. The *F*-measure calculation method is shown in Equation (16):

$$F - measure = \frac{2TP}{2TP + FN + FP} \tag{16}$$

Through these metrics, the model's comprehensive performance can be thoroughly evaluated, encompassing its ability to identify true targets and manage sample imbalance, thereby ensuring its reliability and accuracy in practical applications.

3. Result

### 3.1. Performance evaluation and comparison of DeepmiRNATar

The DeepmiRNATar model leverages an encoder-decoder architecture with enhancements over the conventional Transformer framework. It integrates TextCNN and BiLSTM techniques within the encoder, while the decoder employs a combination of multi-head self-attention layers and convolution. A series of ablation experiments were conducted to assess the significance of each module, using a dataset split into 60% for training, 20% for testing, and 20% for validation.

The DeepmiRNATar-sca variant replaces the standard multi-head attention mechanism with SpatialConv Attention. This variant is designed to evaluate the impact of SpatialConv Attention on model performance and compare it with the traditional multi-head attention mechanism to clarify its role within the model. The DeepmiRNATar-textcnn variant excludes the TextCNN module while retaining other components. This variant helps in exploring the specific contribution of the TextCNN module to model performance and assessing its importance within the overall architecture. The DeepmiRNATar-bilstm variant removes the BiLSTM module, keeping other components intact. This variant aids in understanding the role of BiLSTM in the model and its impact on performance. Additionally, the basic Transformer model employs standard Transformer methods. The DeepmiRNATar model, which integrates TextCNN, BiLSTM, and SpatialConv Attention, is compared with the above-mentioned individual variants to evaluate the effect of combining these modules.

Figure 8A Plot showing the trend of decreasing loss on the validation set as epoch increases for different models. Changes in loss across different models (**Figure**

8B). The results of *AUC*, *PRC*, sensitivity, and *F*-measure for different models demonstrate the effectiveness of each component in the DeepmiRNATar model.

Figure 8A depicts the loss reduction trend during training for various models, showing that DeepmiRNATar converges more rapidly than the original Transformer, thereby enhancing learning efficiency. Post 100 epochs, the loss variance stabilizes, indicating model performance stabilization with no overfitting. **Figure 8B** illustrates the comprehensive performance of DeepmiRNATar across several metrics, surpassing the baseline Transformer in all areas, most notably in *AUC* and *F*-measure scores, underscoring its superiority in predicting miRNA-target interactions.
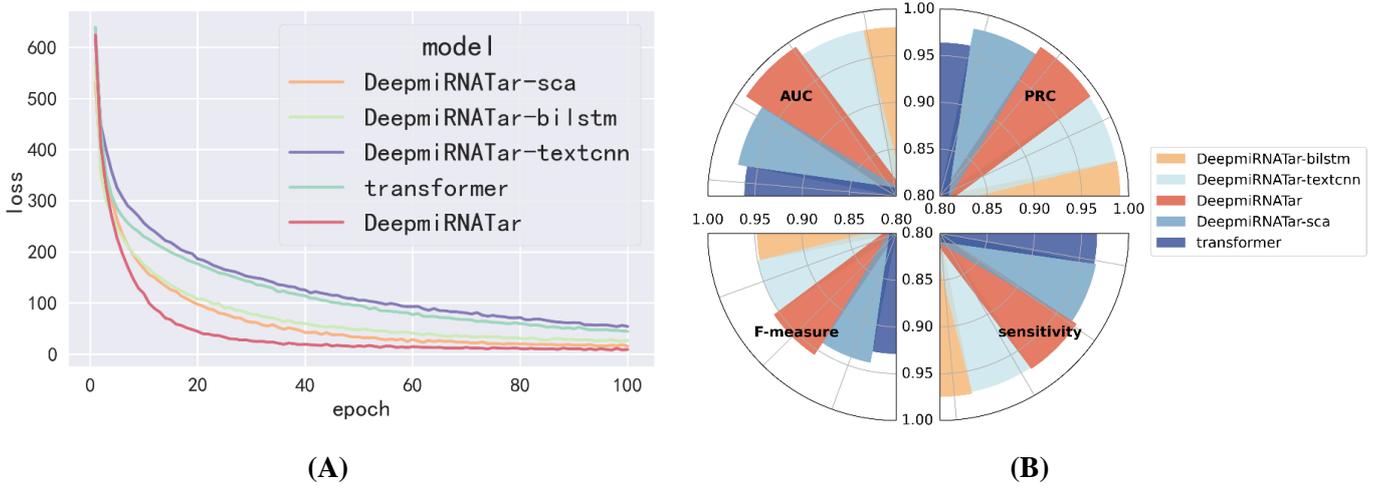
(A)   **(B)**

Figure 8. (A) Plot showing the trend of decreasing loss on the validation set as epoch increases for different models.Changes in loss across different models; (**B**) the results of AUC, PRC, sensitivity, and F-measure for different models demonstrate the effectiveness of each component in the DeepmiRNATar model.

3.2. Hyperparameter tunning

To obtain the optimal model parameters, we conducted a series of controlled variable experiments. The analysis primarily focused on batch sizes and dropout. The basic parameter settings of the model are shown in **Table 2**.

Table 2. DeepmiRNATar model parameter settings.

parameters	value
TextCNN convolutional kernel size	[1,3,5]
Number of decoder layers	4
Number of heads in multi-head attention	8
Dimension of the hidden layer	64
Batch size	32
Dropout	0.1
Learning rate	$1e^{-4}$
epoch	100

3.2.1. The impact of batch size on the model

Batch size refers to the count of miRNA-mRNA pairs engaged simultaneously in each training iteration. Excessively large batch sizes require substantial GPU memory and can diminish the model's generalization capabilities. Conversely, too small a batch size prolongs training duration and can cause gradient oscillations, impeding model convergence. Therefore, selecting an appropriate batch size is crucial for model training [32]. In our experiments, we tested batch sizes of 16, 32, and 64.

As illustrated in **Figure 9**, the model's performance was evaluated against key metrics for batch sizes of 16, 32, and 64. The scores and their error bars clearly demonstrate model efficiency. While all batch sizes presented high and similar scores in AUC and PRC, a batch size of 32 provided an optimal balance in *F*-measure scores. The *F*-measure, the harmonic mean of precision and recall, which is crucial for assessing classification performance, is particularly vital for our model's performance as it accounts for both false positives and false negatives. Moreover, although the sensitivity score for a batch size of 32 was marginally lower than that for 64, considering resource expenditure and computational efficiency, a smaller batch size proves more practical. Therefore, considering both performance evaluation and resource efficiency, a batch size of 32 is a judicious choice.
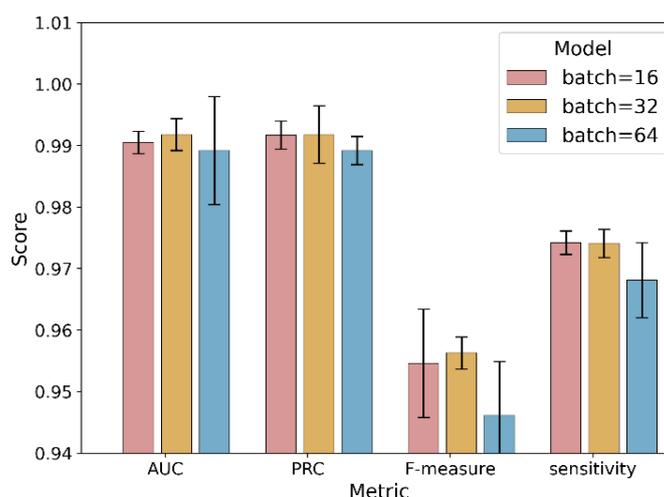

Figure 9. The impact of different batch sizes on model performance.

3.2.2. Choice of dropout

Overfitting is a significant challenge in deep learning methods. Dropout is a technique to avert overfitting by randomly omitting some neurons from the network with certain probabilities. In this study, we experimented with various dropout rates (0.1, 0.2, and 0.3). The experimental results are presented in **Figure 10**.

Regarding AUC and PRC metrics, the model performance was consistent across different dropout rates. However, in terms of *F*-measure and sensitivity metrics, the model with a dropout rate of 0.1 exhibited superior scores. Although a higher dropout rate can help mitigate overfitting, within the context of the current dataset and model architecture, a dropout rate of 0.1 offers adequate regularization while preserving model performance.
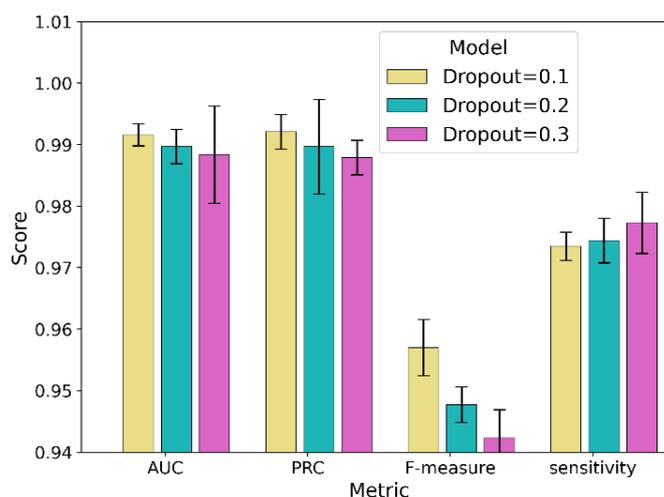
Figure 10. The impact of different dropouts on model performance.

Utilizing the optimal parameters identified, experiments with 100 iterations of training were conducted.

Figure 11A presents the ROC curve of the DeepmiRNATar model, revealing that DeepmiRNATar demonstrates excellent classification ability on the test set with an AUC value of 0.9915.

Figure 11B shows the Precision-Recall (PR) curve of the DeepmiRNATar model, depicting a significant balance between precision and recall. The PRC value of 0.9918 confirms the model's high reliability in positive class predictions, indicating that the model accurately identifies true miRNA targets with minimal false negative rates.
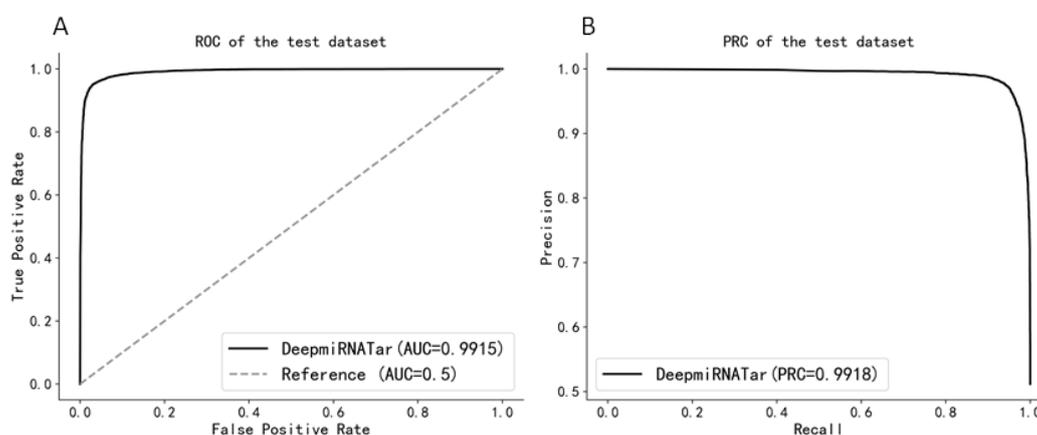

Figure 11. (A) ROC curve for the test dataset; **(B)** PRC curve for the test dataset.

3.3. The impact of RNN variants on the performance of DeepmiRNATar

To assess the effectiveness of BiLSTM in the feature extraction process, this research compared two variants of recurrent neural networks (RNNs): Bidirectional Long Short-Term Memory (BiLSTM) and Bidirectional Gated Recurrent Unit (BiGRU). The performance of these two methods was evaluated while keeping other modules of DeepmiRNATar constant. The study conducted a quantitative analysis of the performance of these two methods, focusing on key metrics such as AUC, PRC,

sensitivity, and *F*-measure to gain a comprehensive understanding of their performance in feature extraction and pattern recognition tasks.

As illustrated in **Figure 12**, the DeepmiRNATar model employing BiLSTM delivers more consistent and efficient results across the testing metrics. This finding underscores the potency of BiLSTM in capturing long-term dependencies in miRNA and target sequences. The superior performance of BiLSTM further validates the potential of deep learning methodologies in accurately predicting miRNA-mRNA interactions.
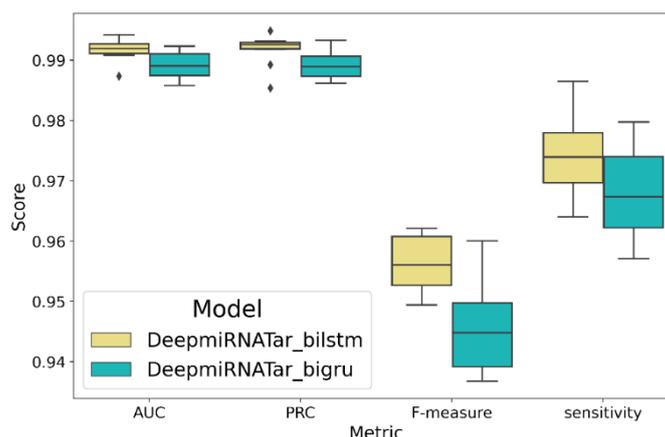
Figure 12. Performance comparison of RNN variants in the miRNA target prediction.

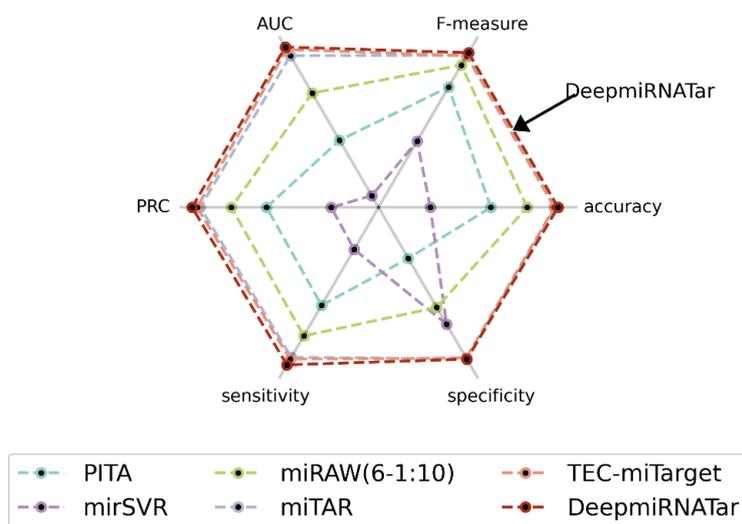### 3.4. Comparison with existing methods

In the experiments, 20% of the dataset was allocated as the test set. The performance of the DeepmiRNATar model was compared with the rule-based software simulation approach PITA [8], the machine learning-based method mirSVR [10], and deep learning-based models miRAW [15], miTAR [16] and TEC-mitar [17]. As shown in **Table 3**, DeepmiRNATar outperformed the other models in nearly all evaluation metrics, further validating its capability in predicting miRNA targets.

Table 3. Comparison results of DeepmiRNATar with existing methods in the field.

DataSet	Test Dataset						Independent set
Methods	AUC	PRC	sensitivity	*F*-measure	ACC	specificity	TPR
PITA	0.4145	0.5957	0.606	0.744	0.597	0.316	23/48
miSVR	0.0723	0.2517	0.26	0.41	0.275	0.723	17/48
miRAW(6-1:10)	0.7064	0.7825	0.795	0.88	0.79	0.619	24/48
miTAR	0.9481	0.9479	0.9274	0.9398	0.9394	0.9276	36/48
TEC-miTarget	0.977	0.965	0.9388	0.9369	0.9353	0.9317	39/48
DeepmiRNATar	0.9915	0.9918	0.9743	0.9547	0.9527	0.9358	41/48

Figure 13 illustrates the performance of DeepmiRNATar and other methods in a radar chart, which provides an intuitive reflection of the models' performance across six key metrics: AUC, PRC, sensitivity, specificity, accuracy (ACC), and *F*-measure. The comprehensive performance of DeepmiRNATar on these metrics highlights its

practicality and efficiency in predicting miRNA-mRNA interactions, especially regarding AUC and *F*-measure. This comparison exhibits the strengths and limitations of DeepmiRNATar across multiple dimensions, laying the groundwork for future model enhancements and selection.


Figure 13. Performance comparison of DeepmiRNATar with existing methods in the field.

For the PAR-CLIP independent dataset, the model in this study was able to correctly predict 85% of the data, with an accuracy higher than other models included in the comparative experiments. Since the positive sample data in the PAR-CLIP independent dataset has already been verified through biological experiments, this also demonstrates the model's ability to identify and recognize microRNA target genes.

3.5. Case study

To further evaluate the generalization capability of the DeepmiRNATar model, a case study was conducted on eight disease-related genes. Among them, GNG7 is a tumor suppressor gene linked to esophageal cancer, gastric cancer, pancreatic cancer, and other diseases [33]. The DNMT1 gene is associated with various diseases, including hereditary sensory neuropathy type Ie and autosomal dominant cerebellar ataxia, deafness, and narcolepsy [34]. The KLF12 gene encodes a transcription repressor and is tied to tumor metastasis and resistance to cancer treatment [35].

First, the model was trained using all known association data from the experimental dataset, and then predictions were made on unknown associations. The prediction results are presented in **Figure 14**, where blue squares indicate that the miRNA regulates the gene, and yellow squares indicate no relationship. All ten predicted interaction pairs were correct, demonstrating the effectiveness of the DeepmiRNATar model in predicting miRNA-mRNA relationships.
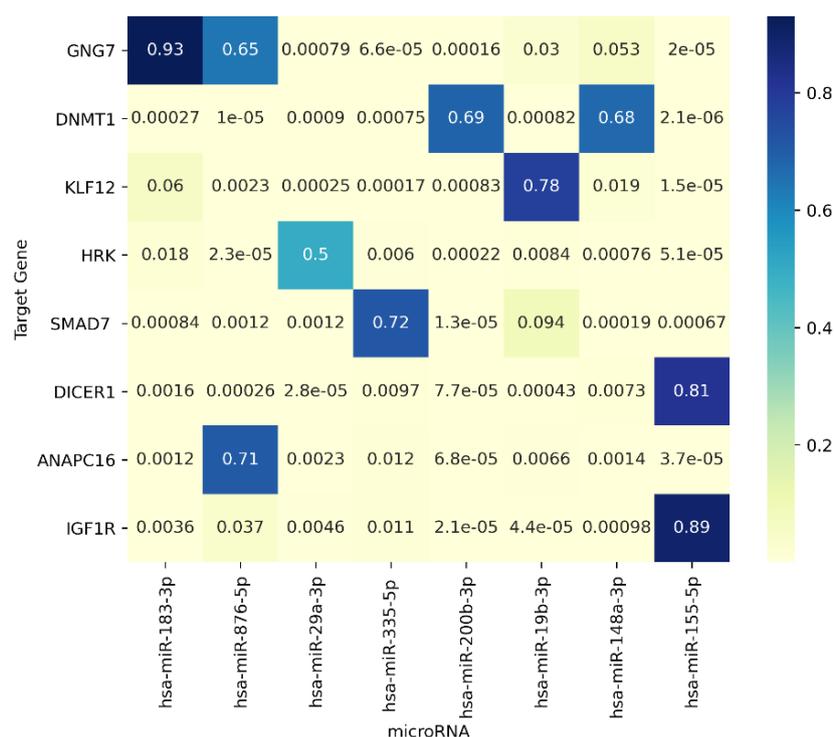
Figure 14. Probability heatmap of 8 miRNAs and their disease-related targets.

4. Dicussion

By integrating various deep learning techniques, this paper presents an innovative miRNA-mRNA interactions prediction model named DeepmiRNATar. Compared to existing methods, DeepmiRNATar shows superior performance on multiple metrics, such as AUC and AUPRC. DeepmiRNATar uniquely combines TextCNN, BiLSTM, and SpatialConv Attention technologies, enabling the model to more accurately capture the complex interactions between miRNAs and targets.

By integrating TextCNN for local feature extraction and BiLSTM for global sequence dependency analysis, the DeepmiRNATar model optimizes sequence data processing. This combined strategy allows the model to detect subtle interaction differences between miRNAs and their targets, which traditional miRNA-mRNA prediction models often overlook. Furthermore, the SpatialConv Attention design in DeepmiRNATar enhances the understanding of spatial features within sequences. This design preserves the fine-grained advantages of the multi-head attention mechanism, thereby providing a more comprehensive understanding of the interaction mechanisms between miRNA and their targets.

The development of the DeepmiRNATar model offers a robust method for predicting miRNA target interactions, showcasing the tremendous potential of deep learning technologies in the field of bioinformatics. With ongoing technological improvements and algorithmic innovations, it is expected to play an increasingly significant role in disease mechanism research and clinical applications in the future.

The success of this research not only deepens our understanding of the complex regulatory mechanisms between miRNAs and their target genes but also has significant practical applications. For example, in precision medicine, the DeepmiRNATar model can help identify new biomarkers, thereby enhancing the

accuracy of disease diagnosis and treatment. In drug development, the DeepmiRNATar model can accelerate the screening and validation of new drug targets, reducing the research and development cycle and costs. Additionally, this model is also crucial in formulating personalized treatment plans, allowing more effective therapeutic strategies based on specific miRNA expression profiles of patients. In summary, the success of the DeepmiRNATar model is a strong validation of the application of deep learning technologies and provides new tools and perspectives for future research in the biomedical field.

Despite the excellent performance of DeepmiRNATar in predicting miRNA-mRNA interactions, the model does have some limitations. The datasets we used are derived from the studies by miRAW and DeepmiRTar. miRAW primarily utilizes human miRNA-mRNA interaction data, while DeepmiRTar encompasses interaction data from multiple species[14,6]. Although the dataset from DeepmiRTar has cross-species characteristics[36], our model training and validation have mainly focused on the species represented in the dataset, and were not specifically designed or tested for cross-species prediction. Therefore, the model's cross-species prediction capability has not been specifically validated. Differences in miRNA and mRNA sequences between species and the coverage of the dataset make cross-species prediction complex. We plan to consider and validate the model's predictive performance on unseen species in future work.

Additionally, the current model primarily targets interactions between miRNAs and mRNAs. Although the model theoretically could be extended to predict interactions between miRNAs and other types of non-coding RNAs (such as lncRNAs and circRNAs), this would require further expansion of the dataset and optimization of the model to capture these complex interactions. Thus, we plan to consider these extensions in future research to enhance the model's applicability in predicting non-coding RNAs.

Author contributions: Conceptualization, DL and HP; methodology, DL; software, YL; validation, YL; formal analysis, CL; investigation, YL; resources, YL; data curation, CL; writing—original draft preparation, HP; writing—review and editing, DL; visualization, HP; supervision, DL; project administration, DL. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Alshalalfa M, Alhajj R. Using context-specific effect of miRNAs to identify functional associations between miRNAs and gene signatures. BMC Bioinformatics. 2013;14 Suppl 12(Suppl 12): S1.
2. Taguchi YH. Inference of Target Gene Regulation via miRNAs during Cell Senescence by Using the MiRaGE Server. Aging Dis. 2012;3(4):301-6.
3. Plaisance-Bonstaff K, Renne R. Viral miRNAs. Antiviral RNAi: Concepts, Methods, and Applications. 2011:43-66.
4. Shaikh MAJ, Altamimi ASA, Afzal M, Gupta G, Singla N, Gilhotra R, et al. Unraveling the impact of miR-21 on apoptosis regulation in glioblastoma. Pathol Res Pract. 2024; 254:155121.
5. Ng MY, Yu C, Chen S-H, Liao YW, Lin T. Er:YAG Laser Alleviates Inflammaging in Diabetes-Associated Periodontitis via Activation CTBP1-AS2/miR-155/SIRT1 Axis. International Journal of Molecular Sciences. 2024;25.
6. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in Drosophila. Genome Biol. 2003;5(1):R1.
7. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120(1):15-20.
8. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nat Genet. 2007;39(10):1278-84.
9. Sturm M, Hackenberg M, Langenberger D, Frishman D. TargetSpy: a supervised machine learning approach for microRNA target prediction. BMC Bioinformatics. 2010; 11: 292.
10. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biology. 2010;11(8): R90.
11. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol. 2015;33(8):831-8.
12. Shuang C, Maozu G, Chunyu W, Xiaoyan L, Yang L, Xuejian W. MiRTDL: A Deep Learning Approach for miRNA Target Prediction. IEEE/ACM Trans Comput Biol Bioinform. 2016;13(6):1161-9.
13. Lee B, Baek J, Park S, Yoon S, editors. deepTarget: end-to-end learning framework for microRNA target prediction using deep recurrent neural networks. Proceedings of the 7th ACM international conference on bioinformatics, computational biology, and health informatics; 2016.
14. Wen M, Cong P, Zhang Z, Lu H, Li T. DeepMirTar: a deep-learning approach for predicting human miRNA targets. Bioinformatics. 2018;34(22):3781-7.
15. Pla A, Zhong X, Rayner S. miRAW: A deep learning-based approach to predict microRNA targets by analyzing whole microRNA transcripts. PLoS computational biology. 2018;14(7): e1006185.
16. Gu T, Zhao X, Barbazuk WB, Lee JH. miTAR: a hybrid deep learning-based approach for predicting miRNA targets. BMC Bioinformatics. 2021;22(1):96.
17. Yang T, Wang Y, He Y. TEC-miTarget: enhancing microRNA target prediction based on deep learning of ribonucleic acid sequences. BMC bioinformatics. 2024;25(1):159.
18. Gallego A, Melé M, Balcells I, García-Ramallo E, Torruella-Loran I, Fernández-Bellon H, et al. Functional Implications of Human-Specific Changes in Great Ape microRNAs. PLoS One. 2016;11(4): e0154194.
19. Moore MJ, Scheel TK, Luna JM, Park CY, Fak JJ, Nishiuchi E, et al. miRNA–target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. Nature communications. 2015;6(1):8864.
20. Kim D, Sung YM, Park J, Kim S, Kim J, Park J, et al. General rules for functional microRNA targeting. Nat Genet. 2016;48(12):1517-26.
21. Vlachos IS, Paraskevopoulou MD, Karagkouni D, Georgakilas G, Vergoulis T, Kanellos I, et al. DIANA-TarBase v7. 0: indexing more than half a million experimentally supported miRNA: mRNA interactions. Nucleic acids research. 2015;43(D1): D153-D9.
22. Chou C-H, Chang N-W, Shrestha S, Hsu S-D, Lin Y-L, Lee W-H, et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic acids research. 2016;44(D1): D239-D47.
23. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. 2010;141(1):129-41.

24. Ding J, Li X, Hu H. TarPmiR: a new approach for microRNA target site prediction. Bioinformatics. 2016;32(18):2768-75.

25. Wang X. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies. Bioinformatics. 2016;32(9):1316-22.

26. Zeng M, Wu Y, Lu C, Zhang F, Wu F-X, Li M. DeepLncLoc: a deep learning framework for long non-coding RNA subcellular localization prediction based on subsequence embedding. Briefings in Bioinformatics. 2022;23(1): bbab360.

27. Edgar R. Syncmers are more sensitive than minimizers for selecting conserved k-mers in biological sequences. PeerJ. 2021;9: e10805.

28. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764-70.

29. Ji Y, Zhou Z, Liu H, Davuluri RV. DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. Bioinformatics. 2021;37(15):2112-20.

30. Chen Y. Convolutional neural network for sentence classification: University of Waterloo; 2015.

31. Subakan C, Ravanelli M, Cornell S, Bronzi M, Zhong J, editors. Attention is all you need in speech separation. ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2021: IEEE.

32. Shen S, Yao Z, Gholami A, Mahoney M, Keutzer K, editors. Powernorm: Rethinking batch normalization in transformers. International conference on machine learning; 2020: PMLR.

33. Duan H, Chen B, Wang W, Luo H. Identification of GNG7 as a novel biomarker and potential therapeutic target for gastric cancer via bioinformatic analysis and in vitro experiments. Aging (Albany NY). 2023;15(5):1445.

34. Singh D, Khan MA, Siddique HR. Role of epigenetic drugs in sensitizing cancers to anticancer therapies: emerging trends and clinical advancements. Epigenomics. 2023;15(8):517-37.

35. Pan X, Zhang W, Wang L, Guo H, Zheng M, Wu H, et al. KLF12 transcriptionally regulates PD-L1 expression in non-small cell lung cancer. Molecular Oncology. 2023;17(12):2659-74.

36. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic acids research. 2014;42(D1): D68-D73.