BMC Biology

RESEARCH

Open Access

Leveraging explainable multi-scale features for fine-grained circRNA-miRNA interaction prediction

Li Peng[1*], Wang Wang[1], Zongyi Yang[1], Xiangzheng Fu[2*], Wei Liang[1] and Dongsheng Cao[3]

Abstract

Background Circular RNAs (circRNAs) and microRNAs (miRNAs) interactions have essential implications in various biological processes and diseases. Computational science approaches have emerged as powerful tools for studying and predicting these intricate molecular interactions, garnering considerable attention. Current methods face two significant limitations: the lack of precise interpretable models and insufficient representation of homogeneous and heterogeneous molecules.

Results We propose a novel method, MFERL, that addresses both limitations through multi-scale representation learning and an explainable fine-grained model for predicting circRNA-miRNA interactions (CMI). MFERL learns multi-scale representations by aggregating homogeneous node features and interacting with heterogeneous node features, as well as through novel dual-convolution attention mechanisms and contrastive learning to enhance features.

Conclusions We utilize a manifold-based method to examine model performance in detail, revealing that MFERL exhibits robust generalization, robustness, and interpretability. Extensive experiments show that MFERL outperforms state-of-the-art models and offers a promising direction for understanding CMI intrinsic mechanisms.

Keywords circRNA, miRNA, Explainability, Multi-scale feature, Representation learning

Background

An increasing number of studies have demonstrated that circRNAs and miRNAs interact through the competing endogenous RNA (ceRNA) network mechanism and exert their respective miRNA sponge functions through competition [1–6]. As research on circRNAs and miRNAs has advanced, a growing number of biological experiments have validated their interactions. For instance, Song et al. discovered that FOXO3a directly induced the expression of miR-29b-2 and miR-338 in breast cancer, suggesting their potential as therapeutic targets for this disease [7]. Zhou et al. found that the presence of miR-130a-5p in CircVAPA could enhance the migration and invasion capabilities of breast cancer cells [8].

The identification of circRNA-miRNA interaction (CMI) through traditional biological experiments is costly and time-consuming [9–11]. However, with the improvement of algorithms and the expansion of datasets, computational methods have emerged as a faster and more efficient approach for predicting CMI. Recently, many models for predicting cMI have been proposed [12–15].

*Correspondence:
Li Peng
plpeng@hnu.edu.cn
Xiangzheng Fu
fxz326@hnu.edu.cn
[1] School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411100, China
[2] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410012, Hunan, China
[3] Xiangya School of Pharmaceutical Sciences, Central South University, Changsha 410003, Hunan, China

Currently, computational approaches for CMI prediction can be categorized into three main groups.

The first category consists of traditional machine learning methods that utilize principles of matrix completion. Qian et al. proposed a computational framework, CMIVGSD [16], for predicting CMI by employing a singular value decomposition algorithm to extract linear features from matrices and using a Light-GBM classifier for prediction. Lan et al. introduced the NECMA model [17], which predicts CMI through inner product and neighborhood regularization logic matrix factorization. Yao et al. proposed the IIMCCMA model [18], which utilizes NetMF to extract latent feature vector representations based on similarity, followed by an inductive matrix completion algorithm to identify potential CMI. These methods primarily rely on known association data for feature extraction and prediction but overlook essential RNA sequence information, potentially compromising both prediction accuracy and comprehensiveness. The second category is based on the principle of ensemble learning. Qian et al. proposed a computational model called CMASG, which utilizes graph neural networks and singular value decomposition for CMI prediction [19]. In parallel, Ma et al. introduced a novel deep learning algorithm for CMI prediction that integrates Node2vec, graph attention networks, a conditional random field layer, and inductive matrix completion [20]. Guo et al. presented a new model called BGF-CMAP that combines gradient boosting decision trees with natural language processing and graph embedding techniques to infer potential CMI [21]. While some models employ natural language processing to extract RNA sequence features, others derive network features for prediction; however, they neglect the aggregation of neighbor information by only aggregating information within the network constructed from known interactions, ignoring the similarity between nodes of the same type. The third category is based on deep learning methods that utilize neural networks to learn node features. Guo et al. proposed a model called WSCD, which extracts attribute features from circRNAs (miRNAs) sequences and behavior features from CMA networks [22]. In the same year, Yu et al. introduced a computational model (SGCNCMI) that identifies circRNA-miRNA interactions by integrating multimodal information with graph convolutional networks [23]. The JSNDCMI model developed by Wang et al. integrates functional similarity and local topological features of nodes, strengthening feature representation with a denoising autoencoder [24]. The CA-CMA model introduced combines natural language features with interaction features, fine-tuning network parameters using labeled samples, and predicting CMI with a deep neural network classifier [25]. However, these methods have certain limitations: during feature fusion, they may encounter feature redundancy or loss while aggregating neighborhood information, they often only aggregate either heterogeneous or homogeneous neighbor information without considering the interactive learning of heterogeneous information.

In summary, the existing CMI prediction methods suffer from several limitations: (i) insufficient consideration of feature information and neglect of the importance of different features; (ii) sole reliance on association network features without aggregating homogeneous neighbor information; (iii) lack of addressing heterogeneous information interaction; (iv) potential redundancy or loss due to feature aggregation. To address these issues, we propose MFERL, a method that utilizes explainable multi-scale features for precise circRNA-miRNA interaction prediction. Specifically, MFERL offers the following advantages:

- To account for the diverse feature information of RNA sequences, we extracted multiple feature representations of RNA from different fine-grained sequence dimensions. Simultaneously, to adjust and balance the various features, we applied enhanced learning to the different features. During the process of feature aggregation learning, we separately considered homogeneous information aggregation as well as heterogeneous information interactive learning.
- To reduce the likelihood of feature redundancy or loss, we employed contrastive learning to optimize the feature vector representations, thereby obtaining high-quality feature embeddings. Ultimately, to enhance the information contained within node features, we concatenated the learned and aggregated feature representations from various perspectives as the ultimate node embeddings for prediction.
- To validate the interpretability of the model, we used t-SNE [26] and MDA [27] for visualization analysis. Simultaneously, to investigate the robustness of the model, we conducted experiments under different positive-to-negative sample ratios.

Results

Model design and training

The overall architecture of the proposed method is illustrated in Fig. 1. MFERL consists of three parts. Part I: multi-scale features extraction of miRNAs/circRNAs. Part II: feature learning of miRNAs/circRNAs. Feature aggregation and enhancement learning are performed on these five types of features from different perspectives: (1) homogeneous information aggregation: the similarity

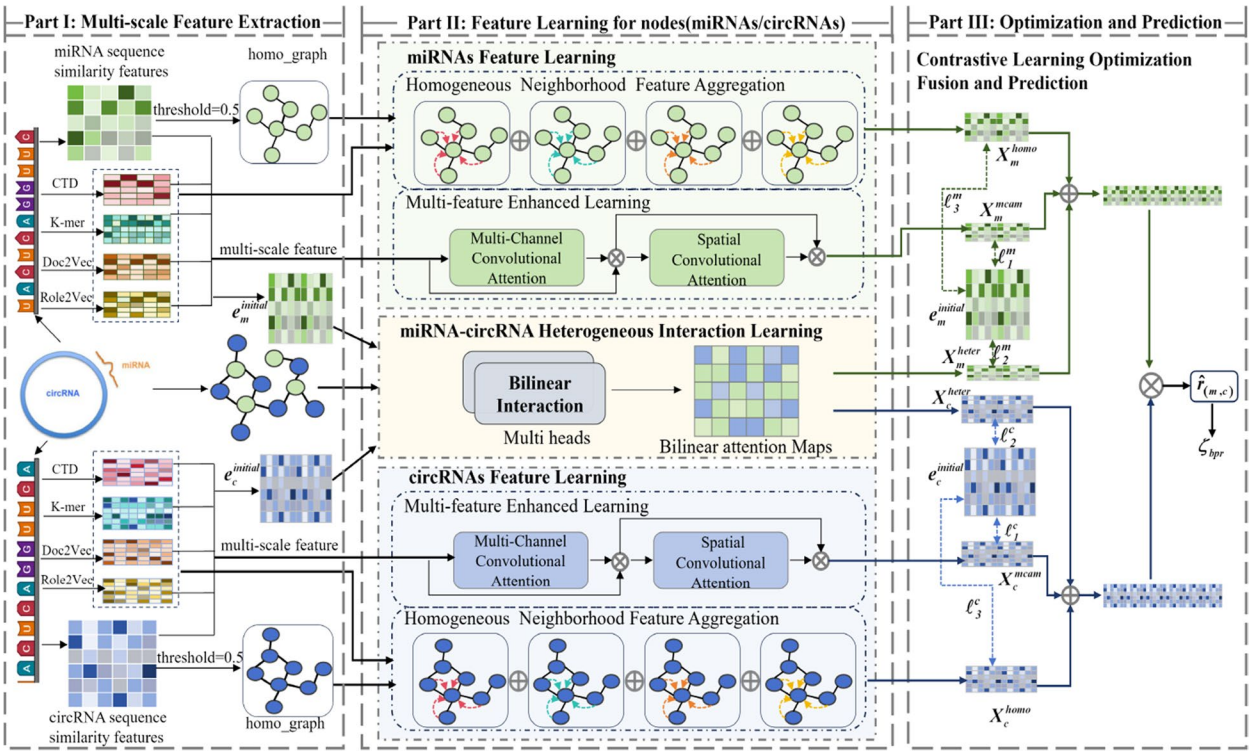

Fig. 1 The architecture of MFERL. Part I: multi-scale features extraction of miRNAs/circRNAs; part II: feature learning of miRNAs/circRNAs; Part III: model optimization and prediction

matrix is thresholded to construct a homogeneous graph, and GCN is applied to aggregate homogeneous neighbor information for the other four features on this graph. (2) Heterogeneous information interaction learning: the five features are concatenated and stretched as RNA feature representations, which are then input into a dual-attention module for heterogeneous information interaction learning. (3) Enhanced learning between features: the five features are fed into a dual-convolutional attention module to enhance learning across the different types of features. Part III: model optimization and prediction. To compare the high-dimensional embeddings of RNA after processing by the model, the embeddings obtained from the three different perspectives are combined with the original features, and contrastive learning is applied using a contrastive loss function to optimize the vector representations. Finally, the three sets of feature vectors are concatenated to form the final feature embedding used for prediction.

The main aim of our study was to develop a predictor for circRNA-miRNA interaction scores. Our model evaluation was conducted using three datasets, where known CMI were treated as positive samples, labeled as 1, and unverified CMI were considered negative samples, labeled as 0. Then, from the samples labeled as 0, we randomly selected an equal number of negative samples to match the number of positive samples, ensuring sample balance. In the 5-fold CV, all positive samples and the selected negative samples were randomly divided into five equal parts, with four parts used for training and the remaining part used for testing. To assess the model's performance, we utilized six common metrics, including AUC (area under the receiver operating characteristic curve), AUPR (area under the precision-recall curve), accuracy, recall, precision, and F1_score. The AUC and AUPR results for Dataset1 in the 5-fold CV are shown in Fig. 2A and B, respectively.

In addition, to verify the universality of the model, we also conducted independent tests on three datasets, as shown in Section 1 and Table S1 in Additional file 1. We also conducted five-fold cross-validation experiments on circRNAs and miRNAs on the three datasets, as shown in Section 2 and Figs. S1, S2, and Table S2 in Additional file 1 [28, 29].

In MFERL, there are three significant hyperparameters considered: di (embedding dimension), lr (learning rate), and τ (temperature hyperparameter in contrastive learning). A series of experiments were conducted on Dataset1 using different hyperparameters to evaluate the sensitivity of the model to these parameters.

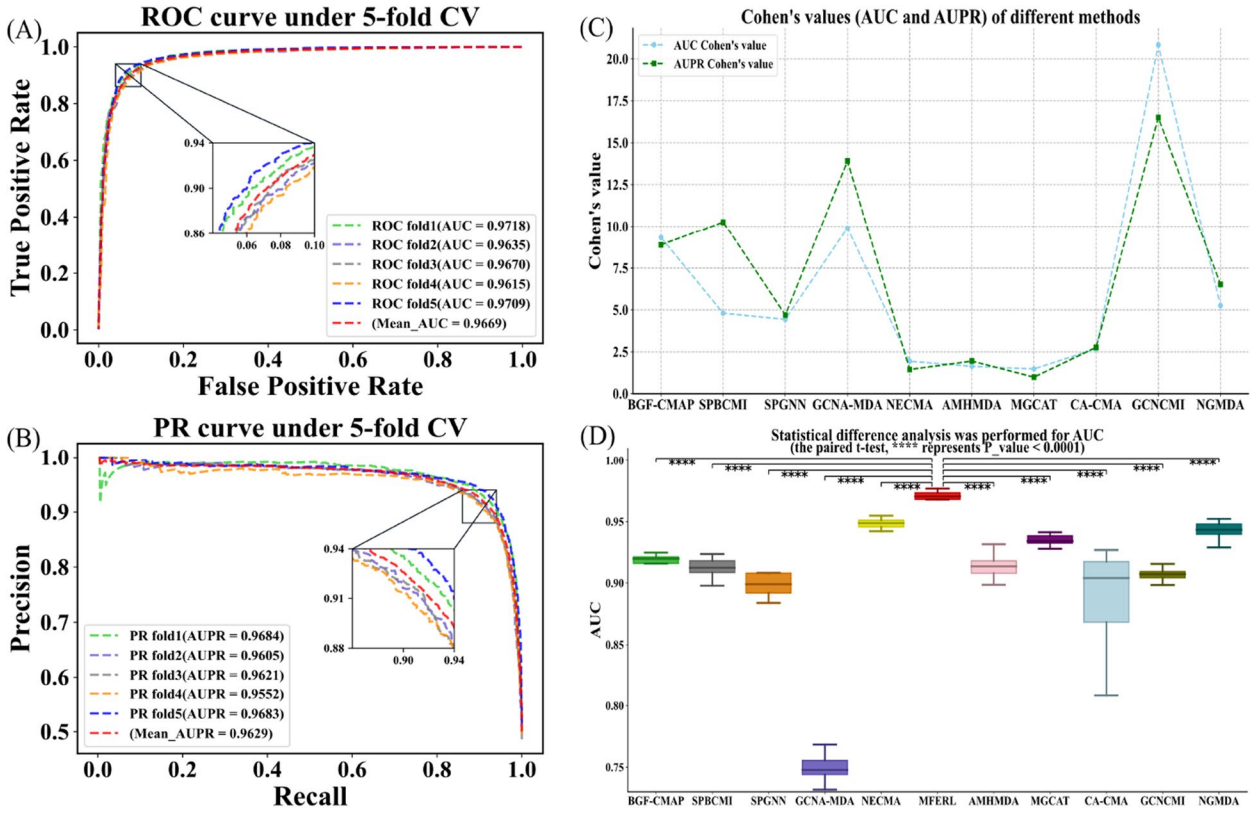

Fig. 2 Performance and statistical analysis of the MFERL model. **A** and **B** ROC and PR curves of MFERL on Dataset1 under 5-fold CV. **C** Statistical significance analysis of Cohen's value between MFERL and 10 methods across three datasets. **D** MFERL outperforms other methods (the paired *t*-test, $P < 0.0001$)

We held other parameters constant while varying the embedding dimension di(32, 64, 128, 256). Using 5-fold cross-validation, we measured and visualized the AUC, AUPR, F1_score, accuracy, recall, and precision, presenting the results in a heatmap as shown in Fig. 4A. Notably, we observed an improvement in MFERL's performance as the embedding dimension increased, with the best AUC and AUPR observed at a dimension of 128. However, the performance started to decline when the dimension reached 256, leading us to select 128 as the optimal value. Next, by adjusting the learning rate, we aimed to optimize the model's predictive capability. Therefore, we varied the learning rate lr(0.0001, 0.001, 0.01, 0.1) and conducted 5-fold cross-validation. As shown in Fig. 4B, the optimal performance was achieved when the learning rate was set to 0.01. To obtain high-quality feature embeddings, we introduced contrastive learning to optimize feature vector representations. Consequently, the temperature parameter τ in contrastive learning was treated as a hyperparameter, with values ranging from 0.01 to 0.15. As depicted in Fig. 4C, the best performance was achieved when τ was set to 0.1.

The proposed MFERL outperforms the state-of-the-art methods

To validate the performance of our model, we compared MFERL with the following methods across three datasets: BGF-CMAP [21], CA-CMA [25], GCNCMI [12], SPBCMI [30], NECMA [17], SPGNN [31], GCNA-MDA [32], NGMDA [33], MGCAT [34], and AMHMDA [35]. All comparisons were conducted under identical experimental settings, with the parameters for the comparative methods set to the optimal values recommended in their respective original studies.

Figure 3 illustrates the ROC and PR curves for the ten comparative methods and MFERL, evaluated through 5-fold CV on three datasets. Additionally, we employed Cohen's value to assess the statistical differences in AUC and AUPR between MFERL and the other methods across the three datasets. A Cohen's value greater than 0.8 indicates a substantial difference. The results in Fig. 2C demonstrate that the Cohen's value for MFERL compared to other methods exceeds 0.8, leading to the conclusion that there is a significant difference between MFERL and the other ten methods across the different datasets.

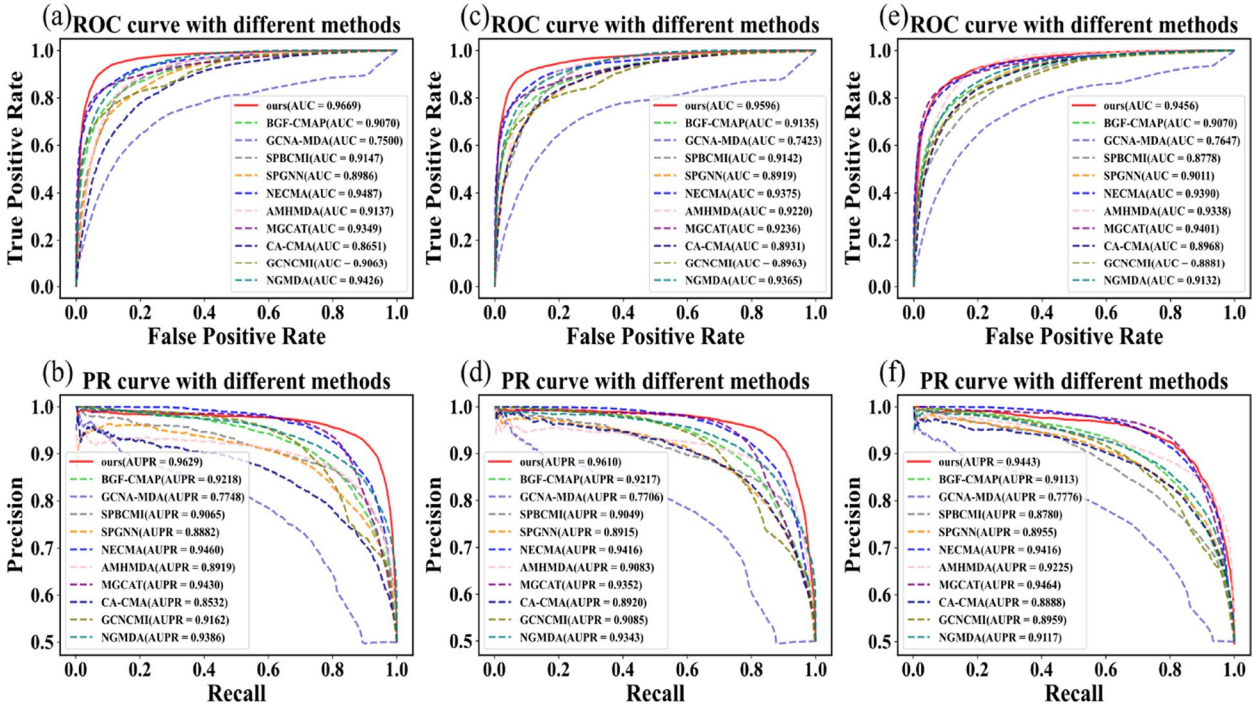

Fig. 3 ROC and PR curves for eleven methods under 5-fold CV. **a** and **b** are on Dataset1; **c** and **d** are on Dataset2; **e** and **f** are on Dataset3

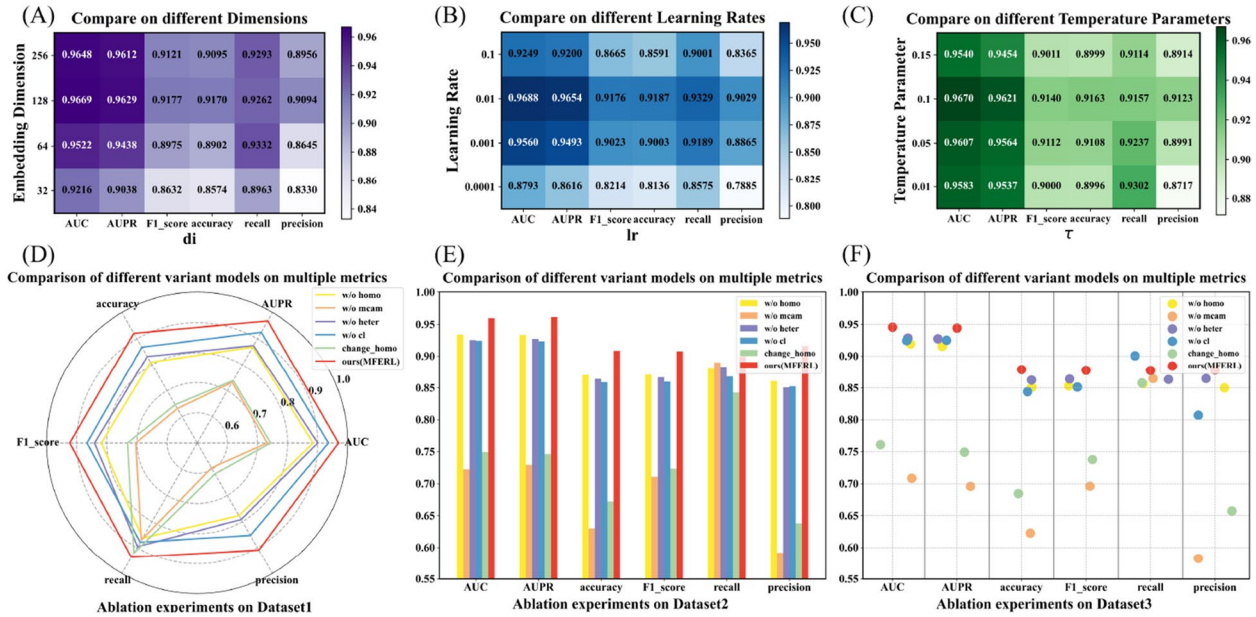

Fig. 4 Evaluation indicators for parameter sensitivity analysis and ablation experiments. **A**, **B**, and **C** The performance of 5-fold CV is compared on Dataset1 with different parameter values. **D**, **E**, and **F** Ablation experiments of MFERL on Dataset1, Dataset2, and Dataset3

Table 1 comprehensively summarizes our experimental analysis, highlighting the superior performance of our proposed model, MFERL, across six key evaluation metrics. Notably, on Dataset1, MFERL achieved impressive performance metrics, with an AUC of 0.9669, AUPR of 0.9629, F1_score of 0.9177, accuracy of 0.9170, recall of 0.9262, and precision of 0.9094. These values show a significant improvement over the corresponding metrics

Table 1 Performance of eleven methods in terms of AUC, AUPR, F1_score, accuracy, recall, and precision under 5-fold CV on Dataset1, Dataset2, and Dataset3

Datasets	Methods	AUC	AUPR	F1_score	Accuracy	Recall	Precision
Dataset1	BGF-CMAP	0.9196±0.0033	0.9218±0.0028	0.8432±0.0059	0.8418±0.0058	0.8507±0.0162	0.8363±0.0127
	GCNA-MDA	0.7508±0.0088	0.7748±0.0072	0.7258±0.0061	0.7156±0.0108	0.7525±0.0121	0.7014±0.0171
	SPBCMI	0.9147±0.0054	0.9065±0.0054	0.8520±0.0050	0.8432±0.0092	0.8929±0.0155	0.8153±0.0207
	SPGNN	0.8986±0.0093	0.8882±0.0134	0.8265±0.0113	0.8156±0.0136	0.8776±0.0142	0.7815±0.0221
	NECMA	0.9487±0.0034	0.9460±0.0042	0.8776±0.0065	0.8755±0.0074	0.8930±0.0105	0.8630±0.0134
	AMHMDA	0.9137±0.0116	0.8919±0.0146	0.8568±0.0081	0.8513±0.0090	0.8895±0.0226	0.8271±0.0143
	MGCAT	0.9349±0.0066	0.9430±0.0050	0.8731±0.0077	0.8773±0.0068	0.8447±0.0185	0.9038±0.0154
	CA-CMA	0.8651±0.0375	0.8532±0.0414	0.8014±0.0277	0.7845±0.0354	0.8659±0.0398	0.7487±0.0564
	GCNCMI	0.9063±0.0018	0.9162±0.0023	0.8214±0.0022	0.8273±0.0034	0.7941±0.0131	0.8510±0.0162
	NGMDA	0.9426±0.0044	0.9386±0.0050	0.8726±0.0055	0.8680±0.0077	0.9040±0.0163	0.8438±0.0176
	Ours(MFERL)	**0.9669** ±0.0040	**0.9629** ±0.0050	**0.9177** ±0.0063	**0.9170** ±0.0055	**0.9262** ±0.0055	**0.9094** ±0.0110
Dataset2	BGF-CMAP	0.9135± 0.0017	0.9217±0.0012	0.8378±0.0016	0.8365±0.0046	0.8440±0.0152	0.8323±0.0173
	GCNA-MDA	0.7423± 0.0051	0.7706±0.0099	0.7245±0.0018	0.7184±0.0068	0.7404±0.0210	0.7102±0.0178
	SPBCMI	0.9142± 0.0058	0.9049±0.0063	0.8490±0.0066	0.8380±0.0097	**0.9087** ±0.0200	0.7975±0.0206
	SPGNN	0.8919±0.0043	0.8915±0.0029	0.8200±0.0027	0.8122±0.0041	0.8555±0.0214	0.7879±0.0139
	NECMA	0.9375±0.0029	0.9416±0.0028	0.8741±0.0037	0.8720±0.0037	0.8886±0.0102	0.8601±0.0066
	AMHMDA	0.9220±0.0127	0.9083±0.0149	0.8672±0.0142	0.8618±0.0148	0.9024±0.0112	0.8349±0.0250
	MGCAT	0.9236±0.0040	0.9352±0.0031	0.8549±0.0040	0.8610±0.0034	0.8192±0.0102	0.8941±0.0044
	CA-CMA	0.8931±0.0168	0.8920±0.0136	0.8236±0.0238	0.8201±0.0234	0.8396±0.0267	0.8085±0.0271
	GCNCMI	0.8963±0.0056	0.9085±0.0053	0.8096±0.0071	0.8117±0.0067	0.8012±0.0224	0.8188±0.0155
	NGMDA	0.9365±0.0048	0.9343±0.0054	0.8631±0.0057	0.8610±0.0069	0.8765±0.0160	0.8508±0.0195
	Ours(MFERL)	**0.9596** ±0.0028	**0.9610** ±0.0041	**0.9074** ±0.0065	**0.9084** ±0.0054	0.8992±0.0053	**0.9159** ±0.0105
Dataset3	BGF-CMAP	0.9070± 0.0020	0.9113±0.0029	0.8308±0.0024	0.8269±0.0025	0.8499±0.0096	0.8127±0.0071
	GCNA-MDA	0.7647± 0.0102	0.7776±0.0118	0.7302±0.0078	0.7098±0.0116	0.7853±0.0162	0.6828±0.0156
	SPBCMI	0.8778±0.0038	0.8780±0.0038	0.8025±0.0052	0.7891±0.0106	0.8572±0.0182	0.7551±0.0222
	SPGNN	0.9011±0.0034	0.8955±0.0035	0.8292±0.0028	0.8173±0.0051	0.8869±0.0185	0.7791±0.0156
	NECMA	0.9390±0.0040	0.9416±0.0040	0.8694±0.0057	0.8674±0.0048	0.8835±0.0141	0.8560±0.0070
	AMHMDA	0.9338±0.0073	0.9225±0.0083	0.8701±0.0113	0.8630±0.0122	**0.9183** ±0.0100	0.8268±0.0178
	MGCAT	0.9401±0.0028	**0.9464** ±0.0020	0.8746±0.0027	0.8764±0.0032	0.8623±0.0137	**0.8876** ±0.0138
	CA-CMA	0.8968±0.0214	0.8888±0.0226	0.8243±0.0241	0.8183±0.0250	0.8527±0.0267	0.7983±0.0251
	GCNCMI	0.8881±0.0015	0.8959±0.0028	0.8098±0.0040	0.8088±0.0028	0.8146±0.0114	0.8053±0.0061
	NGMDA	0.9132±0.0039	0.9117±0.0017	0.8408±0.0037	0.8327±0.0063	0.8830±0.0182	0.8029±0.0144
	Ours(MFERL)	**0.9456** ±0.0034	0.9443±0.0030	**0.8778** ±0.0056	**0.8789** ±0.0059	0.8775±0.0083	0.8781±0.0102

Best and second-best results are bolded and underlined

achieved by the second-best method, exceeding them by 1.82%, 1.69%, 4.01%, 3.97%, 2.22%, and 0.56%, respectively. On Dataset2 and Dataset3, MFERL also consistently ranked either first or second in performance. Additionally, we evaluated MFERL's performance against the ten comparative methods on Dataset1 using ten rounds of 5-fold CV, collecting 50 AUC values for each method. The paired *t*-test was then conducted to statistically compare the AUC and AUPR values of MFERL against those of the ten comparative methods, highlighting the significant differences between MFERL and the other approaches. As shown in Fig. 2D, the results clearly indicate the superior effectiveness of our method compared to the others.

Ablation study of the feature learning

To verify the importance of feature learning from different perspectives in MFERL, we conducted comparative experiments on five variant models using Dataset1 in this section. The five variant models are introduced as follows:

- **"w/o cl"** removed the contrastive learning component from the original model.

- **"w/o homo"** indicated that the model did not consider the aggregation of homogeneous neighbor information.
- **"w/o mcam"** removed the multi-feature enhancement learning module from the original model.
- **"w/o heter"** indicated that the model lacked the heterogeneous information interaction learning module.
- **"change_homo"** represented the model where homogeneous information aggregation was performed by first fusing and then aggregating homogeneous neighbor information.

Figure 4D, E, and F present the comparison of evaluation metrics between the original model and the five variant models on three datasets. It is noteworthy that we found the multi-feature enhancement learning module had a significant impact on the model, highlighting the need to consider the interactions between different features when utilizing multi-scale features to enrich node information.

Explore the impact of the ratio of positive and negative samples in training data on model performance

In comparative experiments conducted on three datasets, it was found that MFERL performed well across all three, demonstrating the model's robustness and generalization capability. In real-world scenarios, the high cost of biological experiments often results in a limited number of verified CMI, leading to an imbalance between positive and negative samples. To further validate the model's generalization ability and robustness under different positive-to-negative sample ratios, we conducted 5-fold CV experiments on Dataset1 with varying positive-to-negative sample ratios (1:1, 1:2, 1:5, 1:10). As shown in Table 2, the changes in AUC for MFERL were minimal, and the accuracy remained above 0.91. Notably, the AUPR, which is most sensitive to the positive-to-negative sample ratio, remains above 0.87, demonstrating the model's strong generalization capability.

Visualization analysis and explainability

To visually demonstrate the model's capability in learning features, we employed t-SNE on Dataset1 to transform the embeddings of circRNA-miRNA pairs learned by our model into a two-dimensional space. As shown in Fig. 5A, it can be observed that as the number of epochs increased, the positive samples (in red) and negative samples (in blue) were gradually distinguished. When the epoch reached 800, the resulting embeddings exhibited good intra-class similarity and a clear boundary between the classes. This result indicates that the model's feature learning is both distinguishable and interpretable, thereby demonstrating the effectiveness of MFERL.

Through the ablation experiments conducted in the previous section, it was evident that the dual-convolution attention module used for multi-feature enhancement learning had a significant impact on model performance. To further explore the feature enhancement learning capability of this module, we conducted a visualization experiment using MDA. Specifically, MDA can display the arrangement of node features in low-dimensional space, where a more continuous color distribution indicates better preservation of the geometric relationships within the feature space. Additionally, MDA can analyze the influence of specific layers on specific behaviors. Therefore, we used MDA to analyze the impact of the dual-convolution attention module on feature enhancement learning behavior. As shown in Fig. 5B, it can be observed that as the training epochs increased, the distribution of the manifold structure in the visualization became more orderly, and the color patterns displayed a trend of clustering and gradual transition, while maintaining a continuous and uniform shape. This indicates the effectiveness of the dual-convolution attention module in feature enhancement learning, further demonstrating that the MFERL model possesses a certain degree of interpretability.

Case validation based on experimental results in the literature

To evaluate MFERL's capability in predicting circRNA-miRNA interaction pairs, we conducted a case study based on Dataset1. In this study, we trained the model using known interactions and an equal number of negative samples, then applied the trained model to predict unknown CMI. Among the top 20 predictions supported by experimental data, 14 CMI pairs were verified, as detailed in Table 3. We provide the top 50 case analysis

Table 2 Comparison of different positive and negative sample ratios during MFERL training

Ratios	AUC	AUPR	F1_score	Accuracy	Recall	Precision
1:1	0.9682	0.9667	0.9191	0.9182	0.9265	0.9119
1:2	0.9710	0.9445	0.8913	0.9265	0.9040	0.8790
1:5	0.9759	0.9117	0.8539	0.9512	0.8543	0.8536
1:10	0.9756	0.8771	0.8148	0.9666	0.8078	0.8219

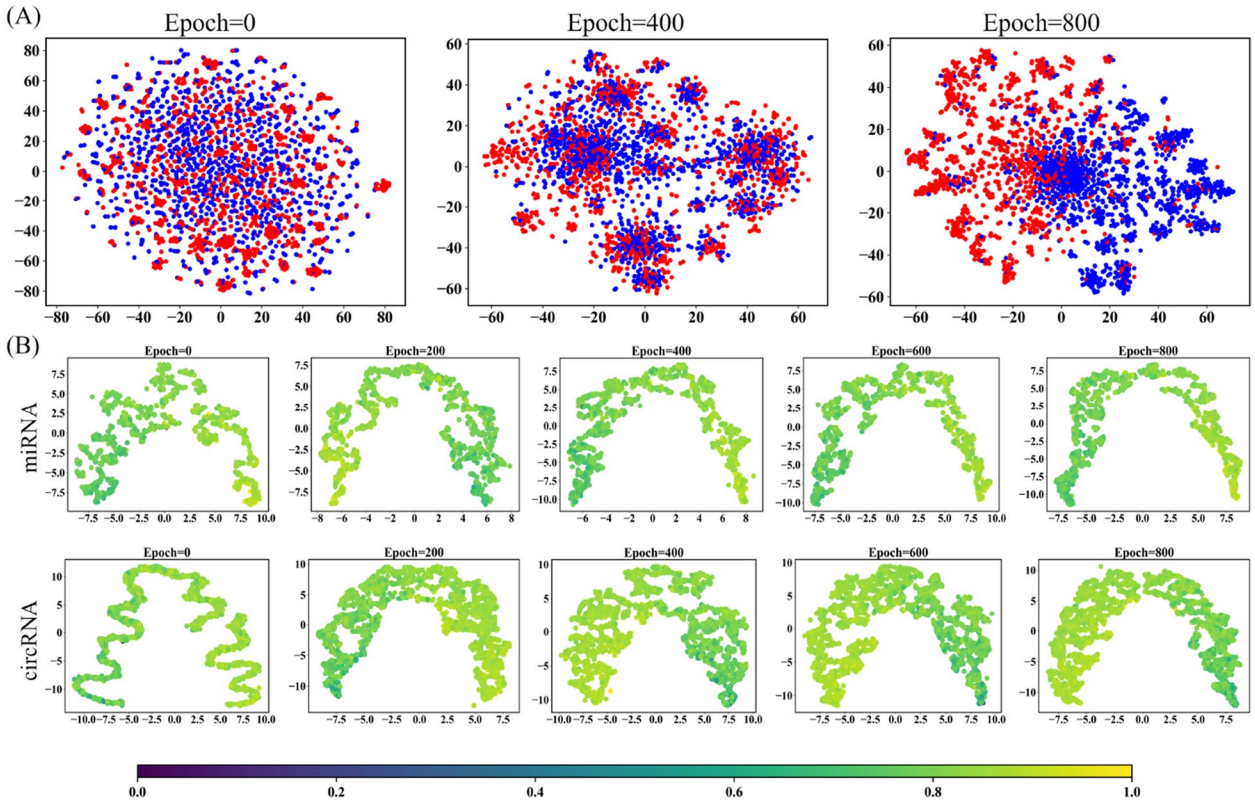

Fig. 5 Visualization analysis and explainability. **A** t-SNE visualization of circRNA-miRNA pairs embeddings learned at different epochs of MFERL training. **B** Visualization of feature-enhanced learning using MDA in dual-convolutional attention modules

Table 3 The top 20 prediction scores among unknown interactions

circRNAs	miRNAs	Evidence
hsa_circ_0001666	hsa-miR-1184	PMID: 35284630
hsa_circ_0041116	hsa-miR-103a-3p	PMID: 27484176
hsa_circ_0000527	hsa-miR-27a-3p	PMID: 34823425
hsa_circ_0013871	hsa-miR-3925-3p	Unconfirmed
hsa_circ_0021030	hsa-miR-1270	PMID:32107851
hsa_circ_0037997	hsa-miR-762	Unconfirmed
hsa_circ_0022342	hsa-miR-942-5p	PMID: 28682884
hsa_circ_0050101	hsa-miR-378a-3p	PMID: 37722013
hsa_circ_0041099	hsa-miR-3125	Unconfirmed
hsa_circ_0019687	hsa-miR-647	PMID: 33490086
hsa_circ_0041891	hsa-miR-378a-3p	PMID: 35093879
hsa_circ_0008234	hsa-miR-574-5p	PMID: 34050132
hsa_circ_0014209	hsa-miR-4685-3p	Unconfirmed
hsa_circ_0019689	hsa-miR-212-5p	PMID: 37870214
hsa_circ_0041089	hsa-miR-103a-3p	PMID: 27484176
hsa_circ_0039186	hsa-miR-1277-5p	Unconfirmed
hsa_circ_0032499	hsa-miR-210	PMID: 2607445
hsa_circ_0012069	hsa-miR-378i	PMID: 36174034
hsa_circ_0089776	hsa-miR-6769b-5p	Unconfirmed
hsa_circ_0068783	hsa-miR-93-5p	PMID: 33585208

and prediction scores in Additional file 1 and show the distribution of some prediction scores. For details, see Section 3, Table S2, and Fig. S3 in Additional file 1.

The results demonstrate that MFERL has strong identification performance. It is important to note that unverified CMI predictions do not necessarily indicate errors. Given MFERL's superior performance in comparative experiments and across three datasets, it can be inferred that the unverified CMI in this case study likely have a high probability of being accurate. Therefore, these predictions urgently require experimental validation.

Discussion

The growing importance of circRNA-miRNA interactions in disease mechanisms highlights the need for accurate and interpretable prediction models. In this study, we proposed MFERL, a novel multi-scale feature learning framework that integrates homogeneous and heterogeneous node learning to improve CMI prediction performance. While MFERL achieved competitive results, some challenges and open questions remain.

First, the inherent class imbalance in CMI datasets poses a significant challenge. Although MFERL maintains robust performance under different negative sampling

ratios, the selection of negative samples remains a critical factor influencing model accuracy. Further research is needed to develop more effective negative sampling strategies, potentially incorporating biological priors or knowledge-based constraints. Second, the model currently emphasizes node-level features without fully considering biological context, such as tissue specificity, disease state, or dynamic regulatory environments. These factors can significantly impact the behavior and interaction patterns of circRNAs and miRNAs. Future work should explore integrating this contextual information, possibly through the construction of knowledge graphs or hypergraphs that capture richer biological relationships. Moreover, although visualization techniques like t-SNE and MDA confirm the model's discriminative power, more advanced interpretability techniques could be applied to provide deeper insights into the learned features and the biological relevance of predictions. Specifically, while MDA helps visualize the organization of learned features and reveals their structural evolution during training, this interpretability remains at the computational level. It does not directly explain the biological mechanisms behind circRNA-miRNA interactions. This limitation stems from our model's reliance on raw sequence data, without integrating explicit biological annotations or motif-level supervision. Nonetheless, the clustering patterns and gradual transitions observed in MDA suggest that the model may implicitly capture biologically relevant signals, such as nucleotide preferences or interaction-related patterns. In future work, we aim to incorporate attention weight analysis, motif discovery techniques, and comparisons with experimentally validated binding sites to better align computational representations with biological meaning. This would further support the model's application in guiding experimental validation.

In conclusion, MFERL offers a solid foundation for CMI prediction, and with further enhancements—especially in data quality, context modeling, and sample selection—it holds strong potential for facilitating biological discovery in the era of RNA research.

Conclusions

In recent years, numerous clinical studies have demonstrated that circRNA-miRNA interaction plays a crucial role in disease development and treatment, drawing significant attention from researchers. In this paper, we propose a CMI prediction model, MFERL, based on multi-scale features learning. The model comprehensively considers features at different scales and employs feature learning from various perspectives. Specifically, we perform homogeneous node aggregation learning and heterogeneous node interaction learning, along with enhanced learning of multi-scale features. The results show that MFERL significantly outperforms other classic methods. Visualization analysis using t-SNE and MDA confirms that MFERL offers inter-class distinguishability and interpretability in feature learning. Moreover, by adjusting the ratio of positive and negative samples during training, we demonstrate the model's strong generalization capability. Case studies further indicate that MFERL is a reliable tool for predicting potential CMI, providing valuable insights for biological experiments.

Although MFERL achieved superior performance in comparative experiments, there are still some limitations. For example, the imbalance between positive and negative samples in the dataset affects the model's ability to accurately select negative samples during training. While this imbalance impacts performance to some extent, our model still demonstrates strong predictive capability, maintaining high precision (PPV) even when the negative sample ratio is increased. However, the varying characteristics of different entities in biological networks—such as the influence of cell types, disease states, or other contextual factors—can affect node feature information. Currently, our model primarily considers the features of target nodes, without incorporating these contextual influences. To address these limitations, we plan to integrate richer biological information (e.g., diseases [36], drugs [37, 38], cell types) in future work, and explore the construction of a comprehensive knowledge graph or a hypergraph to facilitate feature aggregation and improve prediction accuracy. Additionally, we will further investigate strategies for negative sample selection, aiming to develop more robust methods for model training in imbalanced datasets.

Methods

Datasets

MFERL will be tested on three datasets. Table 4 provides a summary of the detailed information for these datasets.

- Dataset1: Circbank [39] is a public database containing five features of circRNAs. Circbank includes approximately 140,000 human circRNAs and 1917 human miRNAs. After removing redundant data, we obtained 9589 circRNA-miRNA interaction pairs from the Circbank database, involving 2115 circRNAs and 821 miRNAs.

Table 4 Specific data of datasets

Datasets	circRNAs	miRNAs	Interactions
Dataset1	2115	821	9589
Dataset2	2346	926	9905
Dataset3	3569	1152	20208

- Dataset2: CMI-9905 was compiled by Wang et al. [24], consisting of 9905 interaction pairs between 2346 circRNAs and 926 miRNAs.
- Dataset3: It was obtained from BGF-CMA [21] and contains 20,208 experimentally validated CMI, involving 3569 circRNAs and 1152 miRNAs.

Overview of MFERL

The overall architecture of the proposed method is illustrated in Fig. 1. MFERL consists of three parts. Part I: multi-scale features extraction of miRNAs/circRNAs. We calculate sequence similarity, statistical features (CTD and K-mer), pre-trained distributed features (Doc2 Vec), and graph structural features (Role2 Vec) for circRNAs and miRNAs. Part II: feature learning of miRNAs/circRNAs. Feature aggregation and enhancement learning are performed on these five types of features from different perspectives. It includes three aspects: aggregation of homogeneous neighborhood features of multi-scale features, dual-convolutional attention module for enhanced learning between multi-scale features, and interactive learning of heterogeneous information of miRNAs and circRNAs. Part III: model optimization and prediction: contrastive learning optimizes vector representation, feature splicing enriches embedded information, and inner product obtains prediction results.

Multi-scale features extraction

To comprehensively characterize circRNAs and miRNAs, we extracted different fine-grained features from multiple perspectives. First, considering the previous work related to circRNAs [40], the Levenshtein distance method was applied to the RNA sequences to calculate the sequence similarity among the same type of RNA. Subsequently, drawing on existing research on interactions between ncRNAs [41], four types of features based on RNA sequences were extracted: statistical features (CTD and K-mer), pre-trained distributed features (Doc2 Vec), and graph structural features (Role2 Vec). The detailed process of feature extraction was then described.

Sequence similarity:

The evaluation of similarity between two circRNA sequences was based on the Levenshtein distance [42], which represents the minimum number of edit operations required to transform one circRNAs sequence into another. The edit operations included not only character substitutions but also the insertion and deletion of characters. Consequently, the sequence similarity for circRNAs was denoted as x_c^s, and similarly, for miRNA, it was denoted as x_m^s.

Composition/transition/distribution (CTD) features:

In this study, CTD features were employed to represent the sequence structural information of RNA. CTD features include nucleotide composition, nucleotide transition, and nucleotide distribution [43]. Currently, CTD features are rarely used for predicting interactions between circRNAs and miRNAs. Here, we utilized CTD features to supplement the structural information of RNA, denoted as x_c^c and x_m^c, respectively.

K-mer features of RNA sequences:

K-mer is a widely used RNA sequence descriptor, which has been successfully applied in enhancer recognition [44] and lncRNA prediction [45]. In this study, we employed four K-mer features, including 1-mer, 2-mer, 3-mer, and 4-mer. For circRNAs (miRNAs) sequence, the four K-mer features were concatenated into a feature vector, with the K-mer features represented as x_c^k and x_m^k.

Pre-trained distributed features by Doc2 Vec:

In this study, Doc2 Vec was utilized to obtain the distributed embeddings [46]. Each RNA sequence was treated as a sentence, and Doc2 Vec learned sentence representations by combining local context and global information. In this context, the distributed features for circRNAs (miRNAs) sequence were represented as x_c^d and x_m^d, respectively.

Graph structural features by Role2 Vec:

Role2 Vec was employed to learn graph structural information by utilizing attributed random walks to capture role-based embeddings. Following the approach in [34], the Role2 Vec embedding method was used to encode nodes within the interaction graph. Similarly, we obtained vector representations for circRNAs, where the Role2 Vec features for circRNAs and miRNAs were represented as x_c^r and x_m^r, respectively.

Feature learning of miRNAs/circRNAs

In MFERL, feature learning involved feature aggregation and enhancement learning of multi-scale features from different perspectives. This section encompassed three aspects: aggregation of homogeneous neighborhood features of multi-scale features, dual-convolutional attention module for enhanced learning between multi-scale features, and interactive learning of heterogeneous information of miRNAs and circRNAs.

Aggregation of homogeneous neighborhood features of multi-scale features

To aggregate different features of similar nodes, we opted to use a similarity matrix as a homogeneous graph and applied graph convolution to aggregate the features of homogeneous neighbors. Unlike traditional feature aggregation methods, we performed feature aggregation separately on the homogeneous graph for each feature type, and then fused the aggregated features by

concatenation and projection. Taking miRNAs as an example: first, we constructed a homogeneous graph G_m by applying a threshold on the similarity matrix x_m^s (for fairness, we set the threshold at 0.5); then, on graph G_m, we used graph convolution to aggregate the CTD features, K-mer features, Doc2 Vec features, and Role2 Vec features (as the similarity matrix has already been used to construct the homogeneous graph, sequence similarity features were not considered during the aggregation of homogeneous neighbor information). The feature aggregation process for CTD features, as an example, is detailed as follows:

$$X_{ctd}^{(l+1)} = f_{conv}\left(X_{ctd}^{l}, A\right) \tag{1}$$

$$f_{conv}\left(X_{ctd}^{l}, A\right) = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X_{ctd}^{l}W_{ctd}^{l}\right) \tag{2}$$

here, A is the adjacency matrix of graph G_m, $X_{ctd}^{0} = x_c^m$, $\tilde{A} = A + I$ represents the adjacency matrix A of the node plus self-loop, and $\tilde{D}$ is the degree matrix of $\tilde{A}$, which is a symmetric matrix. Here, we set the number of layers l of GCN to 1 by default. Similarly, we can obtain homogeneous information features X_m^{kmer}, X_m^{doc}, and X_m^{role} based on k-mer, doc2vec, and role2vec. Finally, the different homogeneous information features are spliced and projected as the homogeneous features of miRNAs and embedded into X_m^{homo}. Similarly, the homogeneous features of circRNAs can be embedded into X_c^{homo}.

Dual-convolutional attention module for enhanced learning between multi-scale features

To adjust and balance the weights of different features, we adopted a dual-convolution attention mechanism for enhanced learning of the various features. Taking miRNAs as an example, we first used the five-layer embedding $X_m = \left\{x_m^k, x_m^c, x_m^d, x_m^r, x_m^s\right\}$ as the input to the channel convolution attention block ($X_m \in \mathbb{R}^{C_m \times N_m \times D_m}$). This process sequentially produced the channel attention $\partial_{mc} \in \mathbb{R}^{5\times1\times1}$ and spatial attention $\partial_{mc} \in \mathbb{R}^{5\times N_m \times D_m}$, where C_m represents the number of channels or embedding layers (default is 5), N_m is the number of miRNAs nodes, and D_m is the embedding dimension. The overall attention process can be summarized as follows:

$$X_m' = \partial_{mc} \otimes X_m \tag{3}$$

$$X_m^{mcam} = \partial_{ms} \otimes X_m' \tag{4}$$

In this context, $\otimes$ denotes element-wise multiplication. During the multiplication process, the attention values are broadcasted accordingly, where X_m^{mcam} represents the embeddings for miRNAs. Similarly, the embeddings for circRNAs can be obtained as X_c^{mcam}. The detailed descriptions of the channel attention and spatial attention modules are provided below, as shown in Fig. 6.

Channel attention module:

We utilized five different features as five layers of feature embeddings input into the channel attention block.

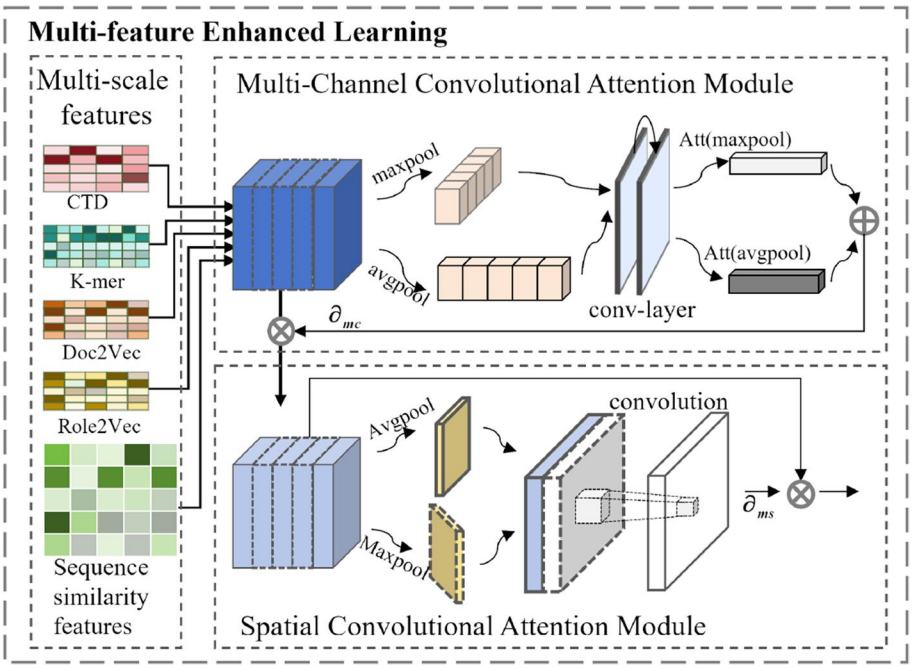

Fig. 6 Dual-convolutional attention module (miRNAs as an example)

Each feature embedding layer was considered as a distinct representation of miRNAs. To capture the influence of different layers on the overall embedding, we used convolutional layers to compress and restore the input embeddings. Additionally, to aggregate attention information, both average pooling and max pooling were employed to achieve more accurate channel attention. The process of the channel attention module is represented as follows:

$$\partial_{mc} = \sigma\left(f_{channel}\left(X_m^{max}\right) + f_{channel}\left(X_m^{avg}\right)\right) \tag{5}$$

here, σ represents the sigmoid function, $f_{channel}(\cdot) = Conv2d(ReLu(Conv2d(\cdot)))$.

Spatial attention module:

We generated a spatial attention map by leveraging the spatial relationships within the feature embeddings. The spatial attention block aimed to enhance channel attention by focusing more on specific parts of the information within the channels. To compute the spatial attention, average pooling and max pooling were applied along the channel axis to the input, and the results were concatenated. A convolutional layer was then applied to generate the spatial attention, which represents the regions where attention should be enhanced or diminished. The process of the spatial attention block is represented as follows:

$$\partial_{ms} = \sigma\left(f_{spatial}\left(\left[X'_{mAvg}; X'_{mMax}\right]\right)\right) \tag{6}$$

here, σ represents the sigmoid function and $f_{spatial}$ represents a convolution operation with a filter size of 3×3.

Interactive learning of heterogeneous information of miRNAs and circRNAs

To enhance the modeling of the complex relationships between the feature representations of miRNAs and circRNAs, we employed a bilinear interaction mechanism for heterogeneous information. This approach aids in extracting the joint representations of miRNAs and circRNAs. Specifically, we fused the five types of features as the miRNAs representation h_m and the circRNAs representation h_c to construct a bilinear interaction mapping, resulting in the interaction feature matrix $Z \in \mathbb{R}^{N_m \times N_c}$. The process is represented as follows:

$$\alpha = \mathrm{norm}\left(W_1\left(\sigma\left((h_m)^T W_m\right) \otimes \sigma\left((W_c)^T h_c\right)\right) + b_1\right) \tag{7}$$

here, $W_m \in \mathbb{R}^{N_m \times d}$ and $W_c \in \mathbb{R}^{N_c \times d}$ represent the learnable weight matrices for miRNAs and circRNAs, respectively. norm$(\cdot)$ represents the weight normalization operation. $\otimes$ represents the outer product, which is used to calculate the product of two vectors. W_1 is the weight matrix of the linear projection; b_1 is the bias term.

The bilinear interaction can be understood as mapping the miRNAs and circRNAs embedding vectors into a shared feature space using the weight matrix W_m and W_c. Subsequently, these mapped representations undergo vector multiplication, resulting in a high-dimensional interaction feature matrix. Additionally, the interaction feature matrix is subjected to a linear projection operation, which maps the high-dimensional features into a low-dimensional representation space, generating the linear projection feature vectors X_m^{heter} and X_c^{heter}. This process enables MFERL to effectively capture the nonlinear relationships between the input features, thereby enhancing its ability to model higher-order interactions between miRNAs and circRNAs features.

Model optimization and prediction

To capture higher-order relationships between nodes, we fully leveraged the advantages of contrastive learning and developed contrastive objectives to obtain high-quality feature embeddings. We treated the fused representation of the original features as the initial node embeddings $e_m^{initial}$ and $X_c^{initial}$. The embeddings X_m^{homo} and X_c^{homo}, obtained by aggregating different features on the homogeneous graph via graph convolution, represented the homogeneous embeddings that incorporate information from homogeneous neighbors. The embeddings X_m^{mcam} and X_c^{mcam}, obtained through dual-convolution enhanced learning of different features, represented the aggregated embeddings of the nodes. The embeddings X_m^{heter} and X_c^{heter}, derived from bilinear interactions of heterogeneous neighbors, represented the heterogeneous embeddings. We utilized these embeddings to model the high-dimensional embedding relationships between RNAs. Specifically, we considered the embeddings of the nodes and those obtained from the aforementioned three feature extraction methods as positive pairs and used the InfoNCE loss function to minimize the distance between positive samples. Taking miRNAs as an example:

$$\begin{aligned}\ell^m = &\ \mathrm{InfoNCE}\left(X_m^{homo}, e_m^{initial}\right) + \mathrm{InfoNCE}\left(X_m^{mcam}, e_m^{initial}\right) \\ &+ \mathrm{InfoNCE}\left(X_m^{heter}, e_m^{initial}\right)\end{aligned} \tag{8}$$

here, $\mathrm{InfoNCE}(x, y) = \sum_{n \in N_m} -\log \frac{\exp\left(x_{m_n} \cdot y_{m_n}/\tau\right)}{\sum_{i \in N_m} \exp\left(x_{m_n} \cdot y_{m_i}/\tau\right)}$, ℓ^m is the local contrast loss of miRNAs, τ is the temperature hyperparameter of softmax. Similarly, the local contrast loss ℓ^c of circRNAs can be obtained. The final local contrast target loss is the weighted sum of ℓ_{local}^m and ℓ_{local}^c as follows:

$$\ell_{cl} = \ell_{local}^m + \alpha \ell_{local}^c \tag{9}$$

here, α is the weight parameter used to balance ℓ^m_{local} and ℓ^c_{local}, default is 1.

After obtaining the node representations aggregated from different features and perspectives, we considered that the diverse node information could enrich the node feature information and contribute to the prediction results. Therefore, we concatenated the different feature representations to obtain the final node representation:

$$M_i = \mathrm{CNN}\left(\mathrm{concatenate}\left(X^{homo}_{m_i}, X^{mcam}_{m_i}, X^{heter}_{m_i}\right)\right) \tag{10}$$

$$C_j = \mathrm{CNN}\left(\mathrm{concatenate}\left(X^{homo}_{c_j}, X^{mcam}_{c_j}, X^{heter}_{c_j}\right)\right) \tag{11}$$

here, CNN$(\cdot)$ is a one-dimensional CNN. After that, we calculated the element-wise product of miRNAs node embedding and circRNAs node embedding. Then, we predicted the probability of interaction of cicRNA-miRNA pairs through FNN:

$$\hat{r}_{ij} = \mathrm{FNN}\left(M_i \odot C_j\right) \tag{12}$$

here, $\odot$ is the element-wise product of the miRNAs node vector and the circRNAs node vector. FNN$(\cdot)$ is a single-layer FNN whose output is activated by the Sigmoid activation function.

To optimize the model, we applied the cross-entropy loss function during model training to calculate the node classification loss:

$$\ell = -\sum_{i,j\in\mathcal{Y}\cup\mathcal{Y}^+}\left[r_{ij}\log\hat{r}_{ij} + (1-r_{ij})\log(1-\hat{r}_{ij})\right] \tag{13}$$

here, r_{ij} indicates node label, $\hat{r}_{ij}$ represents the prediction score, and introduced a contrastive learning objective as an auxiliary task. The final loss of the entire model is formulated as follows:

$$\mathrm{loss} = \ell + \lambda_1\ell_{cl} + \|\theta\|_2^2 \tag{14}$$

here, λ_1 is a hyperparameter that balances the weight of the loss function. θ is the $L2$ regularization parameter.

Abbreviations

circRNAs	Circular RNAs
miRNAs	microRNAs
CMI	circRNA-miRNA interaction
AUR	Area under the receiver operating characteristic curve
AUPR	Area under the precision-recall curve
di	Embedding dimension
lr	Learning rate
τ	Temperature hyperparameter in contrastive learning

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-025-02227-6.

Additional file 1: Table S1-S3 and Figure S1-S3. This document supplements the experiments: independent test experiments, 5-fold CV experiments of circRNAs, case study. Table S1: Performance of the MFERL model independently tested on three datasets. Table S2: Performance of the MFERL model under 5-fold CV for circRNAs and miRNAs on three datasets. Table S3: The top 50 prediction scores among unknown interactions. Figure S1: ROC and PR curves of the MFERL model under 5-fold CV for circRNAs.andare on Dataset1;andare on Dataset2;andare on Dataset3. Figure S2: ROC and PR curves of the MFERL model under 5-fold CV for miRNAs. andare on Dataset1;andare on Dataset2;andare on Dataset3. Figure S3: Sample score distribution.

Acknowledgements

We would like to thank the National Natural Science Foundation of China (Nos. 62372158, 62402533), Natural Science Foundation of Hunan Province (No. 2023JJ30264), and Scientific Research Project of Hunan Education Department (Nos. 22A0350, 23B0237, 22C0267). We also appreciate the contributions of all authors involved in this study, whose collaboration made this research possible.

Authors' contributions

L.P. and W.W. contributed to the initial draft and the design and implementation of the experiments; Z.Y. and X.F. were responsible for data collection and reference preparation; W.L. and D.C. provided experimental guidance and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 62372158, 62402533), Natural Science Foundation of Hunan Province (No. 2023JJ30264), and Scientific Research Project of Hunan Education Department (Nos. 22A0350, 23B0237, 22C0267).

Data availability

All data generated or analyzed during this study are included in this published article, its supplementary information files, and publicly available repositories. The code and datasets of MFERL are freely available at the repository Zenodo (https://doi.org/10.5281/zenodo.15265950) [47] and Github (https://github.com/biohnuster/MFERL) [48].

Declarations

Ethics approval and consent to participate

This study involves computational experiments that are non-invasive and do not directly intervene with any human or animal subjects. Therefore, ethical approval from an institutional review board is not required. The consent of participants and the protection of personal information are not applicable, as the experimental data are sourced from publicly available datasets.

Consent for publication

All authors have provided their consent for publication of this study. There are no identifable individuals or personal data included in this manuscript, ensuring compliance with publication ethics.

Competing interests

The authors declare no competing interests.

Received: 20 November 2024 Accepted: 29 April 2025
Published online: 09 May 2025

References

1. Chen W, Xu J, Wu Y, Liang B, Yan M, Sun C, et al. The potential role and mechanism of circRNA/miRNA axis in cholesterol synthesis. Int J Biol Sci. 2023;19(9):2879.
2. Hill M, Tran N. miRNA interplay: mechanisms and consequences in cancer. Dis Model Mech. 2021;14(4):dmm047662.
3. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011;146(3):353–8.
4. Dakal TC, Kumar A, Maurya PK. CircRNA-miRNA-mRNA interactome analysis in endometrial cancer. J Biomol Struct Dyn. 2025;43(3):1486-97.
5. Costa MC, Cortez-Dias N, Gabriel A, De Sousa J, Fiuza M, Gallego J, et al. circRNA-miRNA cross-talk in the transition from paroxysmal to permanent atrial fibrillation. Int J Cardiol. 2019;290:134–7.
6. Costa MC, Kurc S, Drożdż A, Cortez-Dias N, Enguita FJ, et al. The circulating non-coding RNA landscape for biomarker research: lessons and prospects from cardiovascular diseases. Acta Pharmacol Sin. 2018;39(7):1085–99.
7. Song Y, Zeng S, Zheng G, Chen D, Li P, Yang M, et al. FOXO3a-driven miRNA signatures suppresses VEGF-A/NRP1 signaling and breast cancer metastasis. Oncogene. 2021;40(4):777–90.
8. Zhou Sy, Chen W, Yang Sj, Li J, Zhang Jy, Zhang Hd, et al. Circular RNA circVAPA regulates breast cancer cell migration and invasion via sponging miR-130a-5p. Epigenomics. 2020;12(4):303–17.
9. Tong KL, Tan KE, Lim YY, Tien XY, Wong PF. CircRNA-miRNA interactions in atherogenesis. Mol Cell Biochem. 2022;477(12):2703–33.
10. Nishita-Hiresha V, Varsha R, Jayasuriya R, Ramkumar KM. The role of circRNA-miRNA-mRNA interaction network in endothelial dysfunction. Gene. 2023;851:146950.
11. Ma B, Wang S, Wu W, Shan P, Chen Y, Meng J, et al. Mechanisms of circRNA/lncRNA-miRNA interactions and applications in disease and drug research. BioMed Pharmacother. 2023;162:114672.
12. He J, Xiao P, Chen C, Zhu Z, Zhang J, Deng L. GCNCMI: a graph convolutional neural network approach for predicting circRNA-miRNA interactions. Front Genet. 2022;13:959701.
13. Yu CQ, Wang XF, Li LP, You ZH, Ren ZH, Chu P, et al. RBNE-CMI: an efficient method for predicting circRNA-miRNA interactions via multiattribute incomplete heterogeneous network embedding. J Chem Inf Model. 2024;64(18):7163–72.
14. Wang XF, Yu CQ, You ZH, Qiao Y, Li ZW, Huang WZ. An efficient circRNA-miRNA interaction prediction model by combining biological text mining and wavelet diffusion-based sparse network structure embedding. Comput Biol Med. 2023;165:107421.
15. Li YC, You ZH, Yu CQ, Wang L, Hu L, Hu PW, et al. DeepCMI: a graph-based model for accurate prediction of circRNA-miRNA interactions with multiple information. Brief Funct Genom. 2024;23(3):276–85.
16. Qian Y, Zheng J, Zhang Z, Jiang Y, Zhang J, Deng L. CMIVGSD: circRNA-miRNA interaction prediction based on variational graph auto-encoder and singular value decomposition. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2021. pp. 205–210.
17. Lan W, Zhu M, Chen Q, Chen J, Ye J, Liu J, et al. Prediction of circRNA-miRNA associations based on network embedding. Complexity. 2021;2021(1):6659695.
18. Yao D, Nong L, Qin M, Wu S, Yao S. Identifying circRNA-miRNA interaction based on multi-biological interaction fusion. Front Microbiol. 2022;13:987930.
19. Qian Y, Zheng J, Jiang Y, Li S, Deng L. Prediction of circRNA-miRNA association using singular value decomposition and graph neural networks. IEEE/ACM Trans Comput Biol Bioinforma. 2022;20(6):3461–8.
20. Ma Z, Kuang Z, Deng L. NGCICM: a novel deep learning-based method for predicting circRNA-miRNA interactions. IEEE/ACM Trans Comput Biol Bioinforma. 2023;20(5):3080–92.
21. Guo LX, Wang L, You ZH, Yu CQ, Hu ML, Zhao BW, et al. Biolinguistic graph fusion model for circRNA–miRNA association prediction. Brief Bioinform. 2024;25(2):bbae058.
22. Guo LX, You ZH, Wang L, Yu CQ, Zhao BW, Ren ZH, et al. A novel circRNA-miRNA association prediction model based on structural deep neural network embedding. Brief Bioinform. 2022;23(5):bbac391.
23. Yu CQ, Wang XF, Li LP, You ZH, Huang WZ, Li YC, et al. SGCNCMI: a new model combining multi-modal information to predict circRNA-related miRNAs, diseases and genes. Biology. 2022;11(9):1350.
24. Wang XF, Yu CQ, You ZH, Li LP, Huang WZ, Ren ZH, et al. A feature extraction method based on noise reduction for circRNA-miRNA interaction prediction combining multi-structure features in the association networks. Brief Bioinform. 2023;24(3):bbad111.
25. Guo LX, Wang L, You ZH, Yu CQ, Hu ML, Zhao BW, et al. Likelihood-based feature representation learning combined with neighborhood information for predicting circRNA–miRNA associations. Brief Bioinform. 2024;25(2):bbae020.
26. Maaten LV, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9(11):2579-605.
27. Islam MT, Zhou Z, Ren H, Khuzani MB, Kapp D, Zou J, et al. Revealing hidden patterns in deep neural network feature space continuum via manifold learning. Nat Commun. 2023;14(1):8506.
28. Bai P, Miljković F, John B, Lu H. Interpretable bilinear attention network with domain adaptation improves drug-target prediction. Nat Mach Intel. 2023;5(2):126–36.
29. Long S, Tang X, Si X, Kong T, Zhu Y, Wang C, et al. TriFusion enables accurate prediction of miRNA-disease association by a tri-channel fusion neural network. Commun Biol. 2024;7(1):1067.
30. Zhou J, Wang X, Niu R, Shang X, Wen J. Predicting circRNA-miRNA interactions utilizing transformer-based RNA sequential learning and high-order proximity preserved embedding. Iscience. 2024;27(1):108592.
31. Wang Z, Liang S, Liu S, Meng Z, Wang J, Liang S. Sequence pre-training-based graph neural network for predicting lncRNA-miRNA associations. Brief Bioinform. 2023;24(5):bbad317.
32. Liao Q, Ye Y, Li Z, Chen H, Zhuo L. Prediction of miRNA-disease associations in microbes based on graph convolutional networks and autoencoders. Front Microbiol. 2023;14:1170559.
33. Xuan P, Gu J, Cui H, Wang S, Toshiya N, Liu C, et al. Multi-scale topology and position feature learning and relationship-aware graph reasoning for prediction of drug-related microbes. Bioinformatics. 2024;40(2):btae025.
34. Li H, Wu B, Sun M, Ye Y, Zhu Z, Chen K. Multi-view graph neural network with cascaded attention for lncRNA-miRNA interaction prediction. Knowl-Based Syst. 2023;268:110492.
35. Ning Q, Zhao Y, Gao J, Chen C, Li X, Li T, et al. AMHMDA: attention aware multi-view similarity networks and hypergraph learning for miRNA–disease associations identification. Brief Bioinform. 2023;24(2):bbad094.
36. Peng L, Li H, Yuan S, Meng T, Chen Y, Fu X, Cao D. metaCDA: A Novel Framework for CircRNA-Driven Drug Discovery Utilizing Adaptive Aggregation and Meta-Knowledge Learning. J Chem Inf Model. 2025;65(4):2129-44.
37. Peng L, Yang C, Yang J, Tu Y, Yu Q, Li Z, et al. Drug repositioning via multi-view representation learning with heterogeneous graph neural network. IEEE J Biomed Health Inform. 2025;29(3):1668–79.
38. Peng L, Wang W, Yang C, Xiao W, Fu X, Chen Y. Dual-stream heterogeneous graph neural network based on zero-shot embeddings for predicting miRNA-drug sensitivity. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2024. pp. 1122–1128.
39. Liu M, Wang Q, Shen J, Yang BB, Ding X. Circbank: a comprehensive database for circRNA with standard nomenclature. RNA Biol. 2019;16(7):899–905.
40. Peng L, Yang C, Chen Y, Liu W. Predicting CircRNA-disease associations via feature convolution learning with heterogeneous graph attention network. IEEE J Biomed Health Inform. 2023;27(6):3072–82.
41. Yang S, Wang Y, Lin Y, Shao D, He K, Huang L. LncMirNet: predicting LncRNA-miRNA interaction based on deep learning of ribonucleic acid sequences. Molecules. 2020;25(19):4372.
42. Levenshtein VI, et al. Binary codes capable of correcting deletions, insertions, and reversals. In: Soviet physics doklady, vol. 10. Soviet Union; 1966. pp. 707–710.
43. Tong X, Liu S. CPPred: coding potential prediction based on the global description of RNA sequence. Nucleic Acids Res. 2019;47(8):e43–e43.
44. Lee D, Karchin R, Beer MA. Discriminative prediction of mammalian enhancers from DNA sequence. Genome Res. 2011;21(12):2167–80.

45. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res. 2013;41(17):e166–e166.
46. Lau JH, Baldwin T. An empirical evaluation of doc2vec with practical insights into document embedding generation. 2016. arXiv preprint arXiv:1607.05368.
47. Peng L, Wang W. MFERL. Zenodo. 2025. https://doi.org/10.5281/zenodo.15265950.
48. Peng L, Wang W. MFERL. 2025. GitHub https://github.com/biohnuster/MFERL.

Publisher's Note