



# KGDCMI: A New Approach for Predicting circRNA–miRNA Interactions From Multi-Source Information Extraction and Deep Learning

Xin-Fei Wang<sup>1</sup>, Chang-Qing Yu<sup>1\*</sup>, Li-Ping Li<sup>1,2\*</sup>, Zhu-Hong You<sup>3</sup>, Wen-Zhun Huang<sup>1</sup>, Yue-Chao Li<sup>1</sup>, Zhong-Hao Ren<sup>1</sup> and Yong-Jian Guan<sup>1</sup>

<sup>1</sup>School of Information Engineering, Xijing University, Xi'an, China, <sup>2</sup>College of Grassland and Environment Sciences, Xinjiang Agricultural University, Urumqi, China, <sup>3</sup>School of Computer Science, Northwestern Polytechnical University, Xi'an, China

## OPEN ACCESS

### Edited by:

Chunhou Zheng,  
Anhui University, China

### Reviewed by:

Jin-Xing Liu,  
Qufu Normal University, China  
Yin Sun,  
University of Rochester, United States

### \*Correspondence:

Chang-Qing Yu  
xaycq@163.com  
Li-Ping Li  
cs2bioinformatics@gmail.com

### Specialty section:

This article was submitted to  
Computational Genomics,  
a section of the journal Frontiers in  
Genetics.

**Received:** 31 May 2022

**Accepted:** 23 June 2022

**Published:** 16 August 2022

### Citation:

Wang X-F, Yu C-Q, Li L-P, You Z-H,  
Huang W-Z, Li Y-C, Ren Z-H and  
Guan Y-J (2022) KGDCMI: A New  
Approach for Predicting  
circRNA–miRNA Interactions From  
Multi-Source Information Extraction  
and Deep Learning.  
*Front. Genet.* 13:958096.  
doi: 10.3389/fgene.2022.958096

Emerging evidence has revealed that circular RNA (circRNA) is widely distributed in mammalian cells and functions as microRNA (miRNA) sponges involved in transcriptional and posttranscriptional regulation of gene expression. Recognizing the circRNA–miRNA interaction provides a new perspective for the detection and treatment of human complex diseases. Compared with the traditional biological experimental methods used to predict the association of molecules, which are limited to the small-scale and are time-consuming and laborious, computing models can provide a basis for biological experiments at low cost. Considering that the proposed calculation model is limited, it is necessary to develop an effective computational method to predict the circRNA–miRNA interaction. This study thus proposed a novel computing method, named KGDCMI, to predict the interactions between circRNA and miRNA based on multi-source information extraction and fusion. The KGDCMI obtains RNA attribute information from sequence and similarity, capturing the behavior information in RNA association through a graph-embedding algorithm. Then, the obtained feature vector is extracted further by principal component analysis and sent to the deep neural network for information fusion and prediction. At last, KGDCMI obtains the prediction accuracy (area under the curve [AUC] = 89.30% and area under the precision–recall curve [AUPR] = 87.67%). Meanwhile, with the same dataset, KGDCMI is 2.37% and 3.08%, respectively, higher than the only existing model, and we conducted three groups of comparative experiments, obtaining the best classification strategy, feature extraction parameters, and dimensions. In addition, in the performed case study, 7 of the top 10 interaction pairs were confirmed in PubMed. These results suggest that KGDCMI is a feasible and useful method to predict the circRNA–miRNA interaction and can act as a reliable candidate for related RNA biological experiments.

**Keywords:** circRNA–miRNA interaction, circRNA, deep neural network, graph embedding, K-mer

## INTRODUCTION

Circular RNA (circRNA) is a kind of special single-stranded circular endogenous non-coding RNA molecule. Numerous studies have shown that circRNAs are highly conserved and biostable throughout the organism. CircRNAs were first found in RNA viruses in 1976. Sanger *et al.* (1976) thus described viroids as highly base-paired rod-like and single-stranded closed circular structures. In 1979, the study published by Hsu *et al.* provided electron microscopic evidence for the circular form of RNA (Hsu and Coca-Prados 1979). During the following 30 years, research on circRNAs has achieved some results, and a small number of circRNAs were discovered by accident (Nigro *et al.*, 1991; Cocquerelle *et al.*, 1992). However, during the early stages of research, circRNAs were considered to be abnormal splicing of genes. Although these early studies found and documented the existence of circRNAs and put forward guesses about their functions, the potential impact of circRNAs may have been largely ignored.

A surge in circRNAs research began around 2010 with the progress of RNA-seq technology and specialized computational pipelines, thus bringing circRNAs' testing and sequence analysis back to biological research. CircRNAs' importance gradually emerged, and many studies have indicated that circRNA has a higher stability structure than linear RNAs, which makes circRNA present the tissue-specific expression pattern and play a crucial role in some cell activities (Armakola *et al.*, 2012; Li *et al.*, 2015; Xu *et al.*, 2015). In the meantime, the important biological functions of circRNA also connect it with the diagnosis and treatment of human diseases (Qu *et al.*, 2015; Kulcheski *et al.*, 2016). The circRNA-miRNA interaction is a classic and important aspect of circRNA-mediated gene regulation (Memczak *et al.*, 2013). The circRNAs contain many miRNAs response elements making circRNA work as miRNA sponges (Hansen *et al.*, 2013; Memczak, Jens, Elefsinioti, Torti, Krueger, Rybak, Maier, Mackowiak, Gregersen and Munschauer 2013). This leads to the circRNA binding to miRNA and repressing their function. MiRNA is one of the most important kinds of ncRNAs involved in several aspects of gene regulation in eukaryotes (Grishok *et al.*, 2001; Siomi and Siomi 2010; Hayes *et al.*, 2014). At present, there is numerous evidence showing that circRNAs are miRNA sponges resulting in the up-regulation of downstream proteins, which is closely associated with a variety of human diseases, such as type-2 diabetes, cardiovascular diseases, atherosclerotic vascular disease risk, and cancers (Zhang *et al.*, 2018; Chen and Shan 2021). For example, circRNA can act as an oncogene in some tumors, causing the proliferation and metastasis of cancer cells. Yang *et al.* applied the RT-qPCR assay, CCK-8, wound-healing, and cell colony formation assay to detect the expression level and the effect of RNA molecular on cancer cell proliferation and metastasis, finally finding that circSPECC1 could promote the proliferation and migration of Bca and may be used as a new diagnostic biomarker and effective therapeutic target for some cancers (Yang *et al.*, 2022). Chen *et al.* used quantitative real-time PCR and established the stable knockdown of circXRCC5 in U87 and U251 cells to assess the functions of RNA and detect the

expression of circXRCC5 in glioma tissues, which proved that circXRCC5, as the sponge of miR-490-3p, regulates the expression of the downstream gene and promotes the progression of glioma (Chen *et al.*, 2022). Tao *et al.* (2022) treated human lens epithelial cells with high glucose and detected gene expression by quantitative real-time polymerase chain reaction (PCR). The Cell Counting Kit-8 test, EdU test, and Western blot assay were used to detect cell proliferation and viability. The dual-luciferase reporter method and RNA immunoprecipitation assay were used to validate the target interactions. They finally clarified that circPAG1 can mediate the miR-211-5p/E2F3 axis to protect human lens epithelial cells from damage induced by high glucose (Tao *et al.*, 2022).

The above studies indicate that the research of the circRNA-miRNA interaction can be a new biomarker for the treatment and diagnosis of diseases. However, the traditional biological experimental method is often expensive, time-consuming, and requires a lot of labor. Thus, predicting the interaction between miRNA and circRNA by computational methods is crucial for relevant research. At present, many prediction models have been applied in related fields. For example, Li *et al.* combined a variety of similarities and improved the traditional nonnegative matrix algorithm to predict disease-related miRNAs (Li *et al.*, 2021). Ren *et al.* (2022) proposed a model named BioDKG-DDI, which combines multi-characteristic biochemical information and uses an attention machine to predict potential drug-drug interactions. Yin *et al.* (2020) proposed a computational method, NCPLP, which is based on the network consistency projection and label propagation to predict disease-associated microbes. Zhou *et al.* (2021) proposed a novel method that learns features through multiple kernel learning and deep autoencoder, finally predicting new microRNA-disease associations by reconstruction error. Pan *et al.* (2022) combined protein attribute and behavior vectors and used a deep neural network (DNN) to fuse the protein feature vector to predict the protein-protein interactions. Such computational models have achieved very successful results and provided an experimental basis for further studies.

Because the naming norms and studies on circRNA are not mature, the computational models regarding the circRNA-miRNA interaction are scarce compared with related fields. In recent years, as the amount of circRNAs data increased, a large number of databases have been developed to store information about circRNAs, like circR2Disease (Fan *et al.*, 2018), circBase (Glažar *et al.*, 2014), circRNA disease (Zhao *et al.*, 2018), and Circbank (Liu *et al.*, 2019). CircR2Disease is a manually curated database containing 661 circRNAs, 100 diseases, and 725 experimentally verified associations of circRNA-diseases. Each pair of associations in the circR2Disease database contains a brief description of the circRNA-disease relationship, gene symbol, expression patterns of circRNA, circRNA and disease name, experimental techniques, year of publication, and the PubMed ID. Circbase is an online database that includes thousands of circular RNAs in animals, which users can search the circRNA sequence, gene description, and circRNA ID. CircRNA disease is a database that

documents a total of 354 circRNA-disease interaction pairs between 330 circRNAs and 48 diseases. Each pair includes the circRNA ID and circRNA name, the expression pattern and biological function of circRNA, disease name, and experimental detection techniques. Circbank is a comprehensive database with standard nomenclature, which includes approximately 140,000 human circRNAs, and the other five features of circRNAs. These databases are public and allow users to query the circRNA information based on different search criteria, which allows users to conveniently obtain data for research on circRNAs. This makes predicting the interaction of circRNA-miRNA by computational method possible.

For the purposes listed above, this study proposed a creative efficient computational method named KGDCMI to predict the circRNA-miRNA interactions based on multidimensional feature extraction and fusion. First, we use the *K*-mer algorithm to obtain the abundance of different fragment sequences in the whole RNA sequence, which represents the inherent attribute information of RNA molecules. To take full advantage of the biological information of circRNA and miRNA, we also constructed a Gaussian interaction profile kernel similarity matrix of each RNA. Second, the sparse autoencoder (SAE) was used to further extract features from redundant or sparse information to obtain the final attribute features. Next, we constructed a circRNA-miRNA bipartite graph to describe the associations between these molecules, in which each node represents an RNA molecule, and each link represents their interaction. Then we employed high-order proximity reserved embedding (HOPE), a graph-embedding algorithm to capture the description of behavior information between nodes from the interactions. At last, the DNN was used to objectively and automatically fuse multiple features and predict the potential circRNA-miRNA interactions effectively. To evaluate the performance of the proposed method comprehensively and fairly, the five-fold cross-validation was used in the experiment, and a variety of evaluation indicators were employed to evaluate the performance and practicability of the proposed method. As a result, an 89.30% area under the curve (AUC) and 87.67% area under the precision-recall curve (AUPR) were obtained; meanwhile, in comparison with the known predictive model with the same dataset, the KGDCMI achieved the best prediction accuracy. In addition, in three groups of comparative experiments, we verified an optimal classifier, the most suitable *K* value of *K*-mer for RNA fragment extraction, and feature extraction dimensions respectively. In the prediction results of the proposed method, seven of the 10 pairs with the highest predicted scores were confirmed in published literature, meaning that the method is powerful and feasible.

## MATERIALS AND METHODS

### Dataset

Circbank (Liu, Wang, Shen, Yang, and Ding 2019) is a publicly available database including the five features of circRNAs, such as

a novel naming system of circRNAs based on the circRNAs host genes and the conservation of circRNAs. The Circbank contains approximately 140,000 human circRNAs and 1917 human miRNAs. After removing redundant data, we obtained 9589 circRNA-miRNA interaction pairs, including 2115 circRNAs and 821 miRNAs from the Circbank database.

CircR2Cancer database (Lan et al., 2020) is a manually managed database containing 1439 circRNA-cancer associations between 1135 circRNAs and 82 cancers, and all these associations have been verified by experiments and existing literature. We obtained a total of 318 pairs of circRNA-miRNA between 238 circRNAs and 230 miRNAs from the circR2Cancer database.

After integrating these data from two databases, we finally obtained a total of high-quality 9905 circRNA-miRNA interaction pairs between 2346 circRNAs and 962 miRNAs.

### Extracting Sequence Features of RNAs by *K*-mer Algorithm

Counting all *K*-mers (substrings of length *K*) in RNA sequences is usually used as an important and common step in bioinformatics analysis, like variants detection, transcriptome assembly, and read error correction. Related studies have confirmed that RNA sequences contain abundant information. To obtain more comprehensive attribute information on RNAs, we used the *K*-mer algorithm to obtain a sparse matrix from the RNA sequences, which represent the RNA's attribute features.

For a given circRNA or miRNA sequence, *K*-mer was used to split them into subsequences. It scanned each RNA sequence from beginning to the end with a *k* nucleotides window, one nucleic acid once a time. For a sequence of length *K*, we could obtain  $4^K$  different possible *K*-mers. Different *K* values determine different lengths of vectors. For example, 3-mers of circRNA can be represented as AAA, AAC, . . . , TTG, and TTT, and the number of possible 3-mers is  $l-3 + 1$ , whereas 5-mers can be represented as AAAAA, AAAAC, . . . , TTTTG, TTTTT, and the possible 5-mers is  $l-5 + 1$ . After processing circRNA and miRNA sequences, we obtained the sequence characteristic representation matrices of each RNA, which can be represented as

$$K_{circRNA} = 2346 \times 4^5 \quad (1)$$

$$K_{miRNA} = 962 \times 4^2 \quad (2)$$

The detail of the *K*-mer algorithm is shown in **Figure 1**.

### Gaussian Interaction Profile Kernel Similarity for CircRNA and miRNA

Homologous RNAs with similar associations may have receptors with similar phenotypes; here, we increased the RNA Gaussian interaction profile kernel similarity to represent circRNA and miRNA similarity.

First, with the adjacent matrix,  $P_c \times m$  represents the associations of circRNAs and miRNAs, and the number of circRNAs and miRNAs is defined as *c* and *m*. When circRNA *i* is related to miRNA *j*, the value of the matrix  $P_{(ij)}$  equals 1, and is otherwise 0.

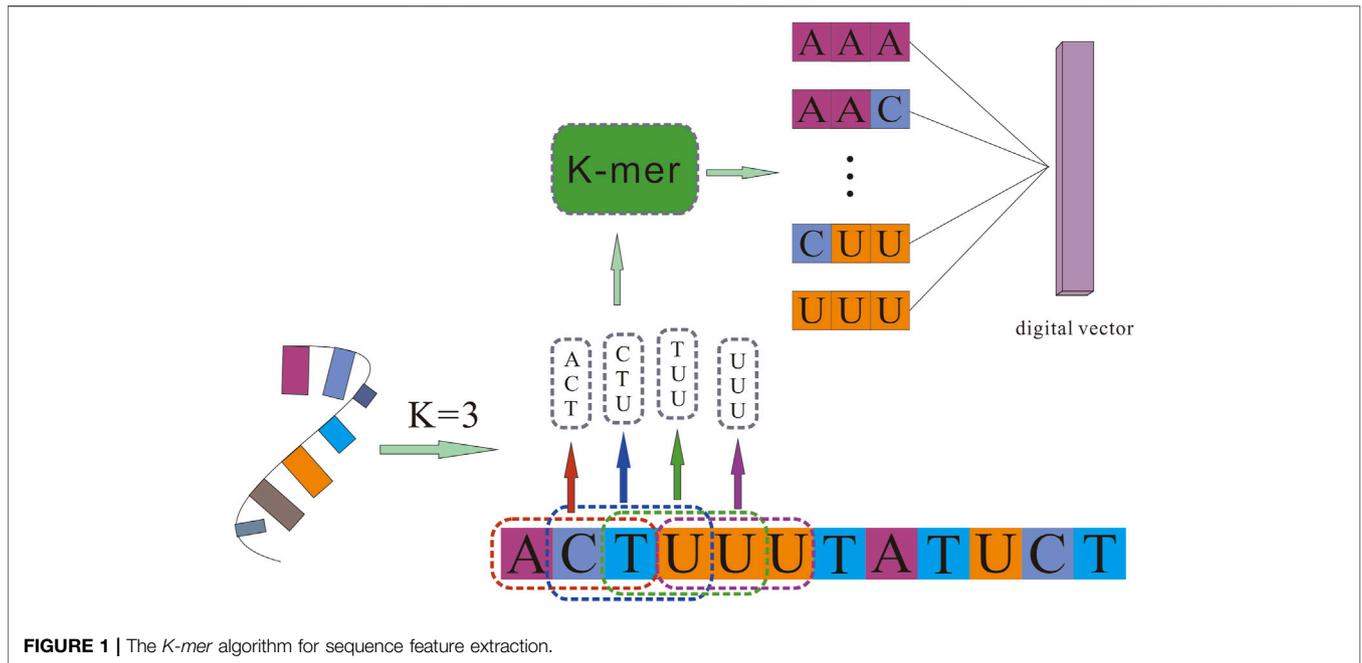


FIGURE 1 | The *K-mer* algorithm for sequence feature extraction.

Each row of matrix,  $P_c \times m$  represents a circRNA interaction profile, and the column represents a miRNA interaction profile. Furthermore, the gastric inhibitory polypeptide (GIP) kernel of circRNA can be calculated by the following formula:

$$G_C(c_i, c_j) = \exp(-\lambda_c \|LP(c_i) - LP(c_j)\|^2) \quad (3)$$

where  $G_C(c_i, c_j)$  denotes the GIP kernel similarity between circRNA  $c_i$  and circRNA  $c_j$ ,  $\lambda_c$  is an adjustable parameter, which controls the kernel bandwidth:

$$\lambda_c = \lambda_c / \left( \frac{1}{nc} \sum_{i=1}^{nc} \|LP(c_i)\|^2 \right) \quad (4)$$

$\lambda_c$  is defined as one according to the previous study (Van Laarhoven et al., 2011).

Likewise, the GIP kernel similarity between miRNA  $m_i$  and miRNA  $m_j$  is calculated as

$$G_M(m_i, m_j) = \exp(-\lambda_m \|LP(m_i) - LP(m_j)\|^2) \quad (5)$$

$$\lambda_m = \lambda_m' / \left( \frac{1}{mm} \sum_{i=1}^{mm} \|LP(m_i)\|^2 \right) \quad (6)$$

### Sparse Autoencoder to Extract Features

RNA sequences contain a great deal of valuable information, and we believe that using the *K-mer* algorithm can effectively transform sequence information into a digital vector containing rich attribute information. When using the *K-mer* algorithm, the value of *K* ranges from 2 to 5, which is the most effective empirical parameter verified. However, the length of

RNA sequences is not uniform. When the value of *K* is too small, useful information will be lost. On the contrary, it will cause “noise” disturbance and increase the computational overhead.

Principal component analysis is a data dimensionality reduction and feature extraction method widely used before data are preprocessed by other algorithms. Furthermore, it removes noise and useless information about features to maximize the information value of features and improve the algorithm accuracy. In this work, we used the SAE (Ng 2011) with sparse penalty terms to obtain a more effective and comprehensive eigenvector.

SAE is a three-layer neural network including an input layer, an output layer, and a hidden layer that include two functions: encoding and decoding.

The function of encoding the layer is mapping the input feature  $x$  to the hidden layer  $H_2$

$$H_2 = \sigma(W_{H_1}x(i) + b_{H_1}) \quad (7)$$

where  $x$  is the origin-dimensional input data vector,  $W$  is the connection parameter between the input and the hidden layer, and  $b$  represents a function offset.

Select  $\sigma(\cdot)$  as the network’s activation function:

$$\sigma(x) = \frac{1}{(1 + e^{-x})} \quad (8)$$

$H_2 = (h_1, h_2, \dots, h_l)$  is the vector output from the hidden layer. The average activation amount of the activated hidden unit  $t$  is

$$\hat{\rho}_j = \frac{1}{n} \sum_{i=1}^n [a_j(x(i))] \quad (9)$$

$a_j(x)$  denotes the activation of hidden unit  $t$ .

SAE adds the sparsity penalty term to the target function, which constraints on the hidden layer to maintain low-average-activation values:

$$P = \sum_{j=1}^{h_2} KL(\rho || \hat{\rho}_j) \tag{10}$$

$h_2$  is the number of neurons in the hidden layer, and  $P$  reflects the degree of penalization  $\hat{\rho}_j$  deviating from  $\rho$ .

The sparsity penalty term is expressed as the *Kullback-Leibler* (*KL*) divergence. In general,  $\rho$  is a sparsity parameter, which is a small value close to 0:

$$KL(\rho || \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \tag{11}$$

If  $\hat{\rho} = \rho$ ,  $KL(\rho || \hat{\rho}_j) = 0$ ; Otherwise,  $KL$  increases monotonically as the gap grows.

The cost function of sparse penalty term is defined as

$$F_C(W, b) = F(W, b) + \kappa \sum KL(\rho || \hat{\rho}), \tag{12}$$

where  $F(W, b)$  is the cost function of the neural networks, and  $\kappa$  is the weight of the sparse penalty.

### High-Order Proximity Reserved Embedding

Graph-embedding algorithm is an effective method to mine for hidden information that can map graphs into the vector space and retain the structure and inherent attributes of graphs. In this part, we constructed a bipartite graph of the circRNA-miRNA interactions and used a graph-embedding algorithm, HOPE (Ou et al., 2016) to capture hidden information between nodes in a large-scale network.

For a given graph  $Gn = \langle V, E \rangle$ ,  $V$  is a collection of vertices, and  $E$  represents the directed edge sets. HOPE embeds  $Gn$  into a vector space, where the asymmetric transitivity and the structure of the graph are preserved. HOPE preserves the asymmetric transitivity by approximating high-order proximity, and it adopts the *L2-norm* below as the minimized loss function:

$$\min ||M - N^s \cdot N^t ||_F^2 \tag{13}$$

$M$  is a high-order proximity matrix, where  $M_{ij}$  is the proximity between  $v_i$  and  $v_j$ , and  $N = [N^s \cdot N^t]$  represents the embedding matrix,  $N^s, N^t \in \gamma^{\alpha \times \beta}$  are the source embedding vectors and target embedding vectors respectively, where  $\beta$  is the dimension of the embeddings. Many high-order approximations reflect asymmetric transfer properties, HOPE adopts a general formulation to facilitate the approximation of these proximities:

$$M = M_g^{-1} \cdot M_t \tag{14}$$

where  $M_g$  and  $M_t$  are polynomial matrices. Furthermore, HOPE uses singular value decomposition (SVD) to obtain an optimal rank- $K$  approximation of the proximity matrix  $M$  and corresponding singular vectors to construct embedding vectors:

$$M = \sum_{i=1}^N \sigma_i v_i^s v_i^{tT} \tag{15}$$

where  $\sigma_i$  is a singular value sorted in descending order, and  $v_i^s, v_i^t$  are corresponding singular vectors of  $\sigma_i$ . The obtained optimal embedding vectors are

$$N^s = [\sqrt{\sigma_1} \cdot v_1^s, \dots, \sqrt{\sigma_k} \cdot v_k^s] \tag{16}$$

$$N^t = [\sqrt{\sigma_1} \cdot v_1^t, \dots, \sqrt{\sigma_k} \cdot v_k^t] \tag{17}$$

### Deep Neural Network

In recent years, as a new machine learning technology, the neural network has become a widely used method in many fields and achieved satisfactory results. We used the DNN to effectively learn and fuse multidimensional features and predict potential circRNA-miRNA interactions.

After SAE extracted the attribute feature and HOPE processed the embedded feature, the extracted attribute feature and behavior feature were concatenated and formed a complete sample feature vector. Then, the complete feature vector was fed into the DNN composed of an input layer, multiple hidden layers, and an output layer to obtain the final prediction of the interaction between each circRNA and miRNA.

DNN was fed features from the input layer, and then the hidden layers transformed data in a non-linear way. At last, the learned features were calculated to the output layer. In the whole DNN model, the neuron units in layer  $n$  were connected to the layer  $(n - 1)$ , and the output data were generated by non-linear transformation function  $f(\cdot)$ :

$$D_{i+1} = f \left( \sum_{i=1}^H w_i d_i + b_i \right) \tag{18}$$

where  $H$  represents the number of hidden neurons, and  $w_i$  and  $b_i$  are the weight and bias of neurons respectively. In our method, the rectified linear unit was used to capture hidden patterns within the feature and reduce gradient vanishing:

$$F_{RELU}(x) = \max(0, x) \tag{19}$$

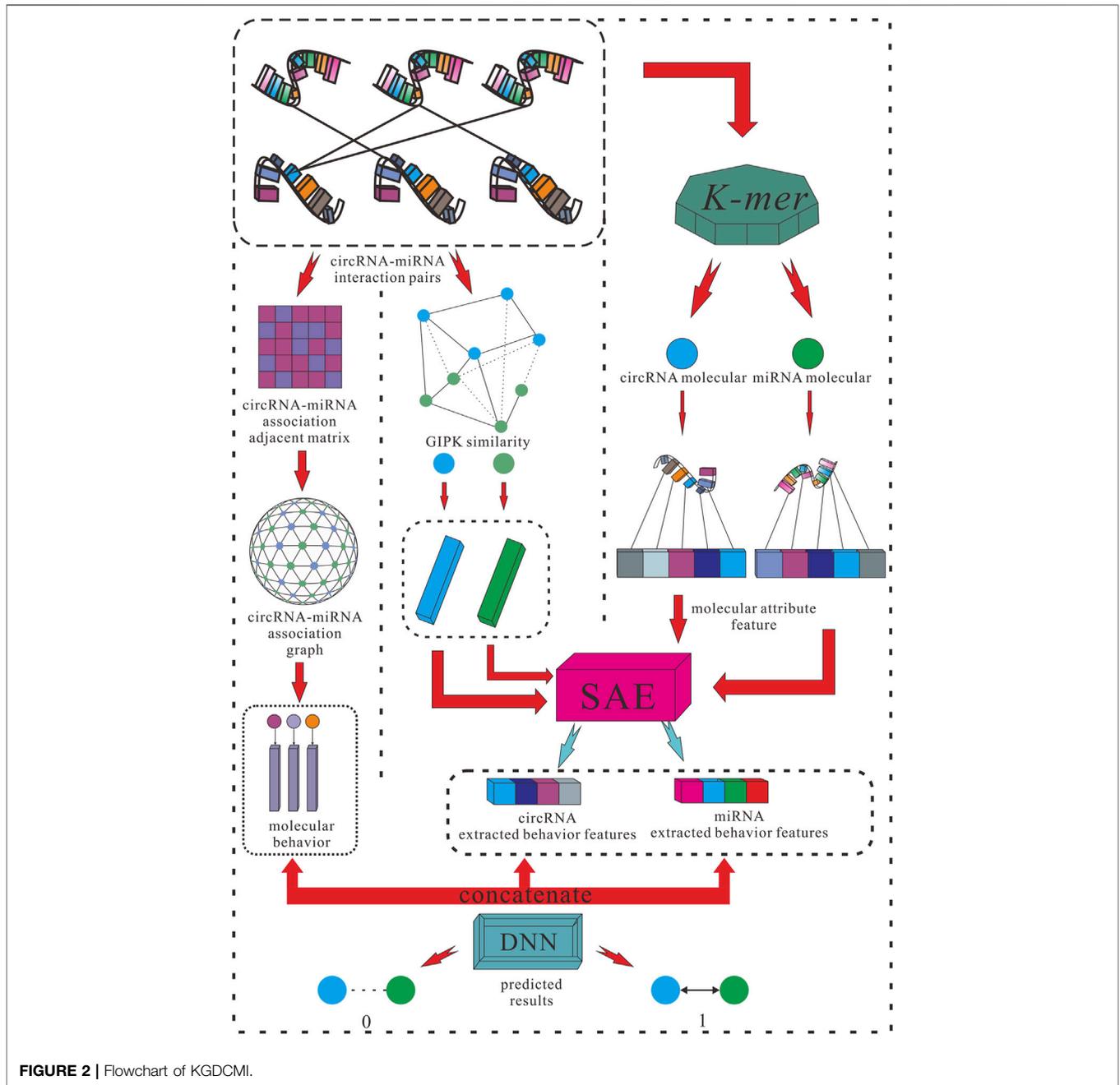
and *sigmoid* was utilized to project the predicted value within a reasonable range:

$$F_{sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{20}$$

We selected *binary\_crossentropy* as the cost function to minimize the objective function to minimize the loss, which is shown as follows:

$$F_{cost} = -\frac{1}{L} \sum_j y_j \cdot \log \hat{y}_j + (1 - y_j) \cdot \log(1 - \hat{y}_j) \tag{21}$$

where  $L$  is the length of the output size,  $y_j$  represents the true label for samples 0 or 1, respectively, and  $\hat{y}_j$  is the predicted probability of the point being positive. For a given positive sample  $y = 1$ , the closer the predicted value is to 1, the smaller the loss value is, and vice versa. In an ideal sense, we hope the loss value approaches 0 infinitely. The specific step is shown in **Figure 2**.



## RESULTS

### Evaluation Criteria

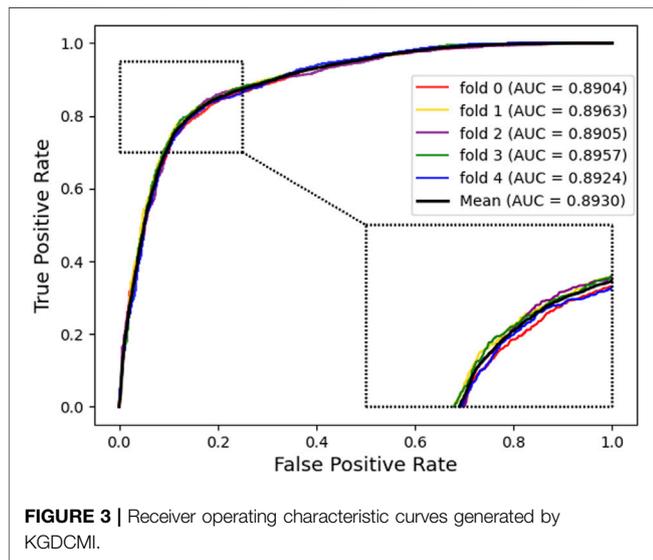
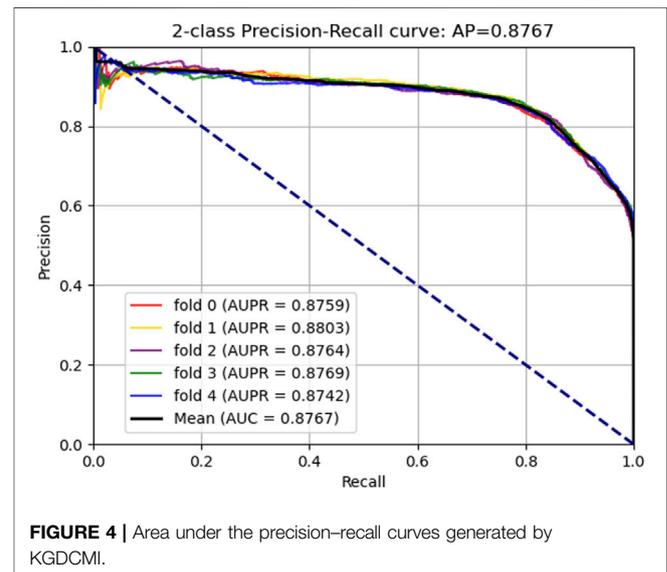
In this study, the samples were first divided into a training, validation, and test set. Taking 70% of samples as the training set to construct the representation vector and train the DNN classifier, 20% of samples were regarded as a validation set, and the remaining 10% of them were treated as a test set to evaluate the performance of the model. At last, the mean and standard deviation of the results of five experiments were calculated. Meanwhile, to fully and fairly demonstrate the

performance of our method, we introduced five evaluation criteria to comprehensively evaluate the results of the model, which assesses the robustness and effectiveness of the model. The evaluation criteria used include the *ACC.* (accuracy), *Sen.* (sensitivity), *Spec.* (specificity), *Prec.* (precision), and *MCC.* (Matthews correlation coefficient) to ensure comprehensiveness and fairness of the results. The calculation formula is as follows:

$$Acc. = \frac{(TN + TP)}{(TN + TP + FN + FP)} \tag{22}$$

**TABLE 1** | Five-fold cross-validation results performed by KGDCMI.

Test Set	ACC. (%)	Sen. (%)	Spec. (%)	Prec. (%)	MCC (%)
1	82.00	80.62	83.39	82.92	64.03
2	82.94	78.55	87.33	86.11	66.13
3	82.74	80.11	85.36	84.55	65.56
4	83.04	81.32	84.76	84.21	66.12
5	82.51	80.36	84.65	83.97	65.08
Average	82.65 ± 0.41	80.19 ± 1.02	85.10 ± 1.44	84.35 ± 1.16	65.38 ± 0.87

**FIGURE 3** | Receiver operating characteristic curves generated by KGDCMI.**FIGURE 4** | Area under the precision-recall curves generated by KGDCMI.

$$Sen. = \frac{TP}{(TP + FN)} \quad (23)$$

$$Spec. = \frac{TN}{(TN + FP)} \quad (24)$$

$$Prec. = \frac{TP}{(TP + FP)} \quad (25)$$

$$Mcc. = \frac{(TP \times TN - FP \times FN)}{\sqrt{((TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN))}} \quad (26)$$

where *TP* (true positive) is the count of true interactions predicted to have interacting circRNA-miRNA samples; *TN* (true negative) is the number of true interactions predicted to have non-interacting circRNA-miRNA samples; *FN* (false negative) is the count of interacting circRNA-miRNA samples that are predicted to have no interaction; and *FP* (false positives) refers to the number of non-interacting circRNA-miRNA samples predicted to interact.

## Assessment of Prediction Ability

In this part, to verify the practicability and performance of the proposed method, the dataset, including data based on the Circbank database and circR2Cancer database, has been used to evaluate the method. **Tables 1** list the experimental results of the proposed method, the results on the dataset were as follows: the average *ACC.* of experiments is 82.65%, *Sen.* is 80.19%, *Spec.* is

85.10%, *Prec.* is 84.35%, *MCC.* is 65.38%, *AUC* is 89.30%, and *AUPR* is 87.67%. The standard deviations of these evaluation criteria are 0.41, 1.02, 1.44, 1.16, 0.87, and 0.28 respectively. Among the five-experiment validation, the highest *AUC* was 89.63%, and the lowest was 89.04%.

In addition, we plotted the receiver operating characteristic curve (ROC) and precision-recall (PR) for visualization, while we calculated the area under ROC (*AUC*) and area under PR (*AUPR*) separately to facilitate comparison with other methods. The *AUC* and PR curves are shown in **Figures 3, 4**, respectively.

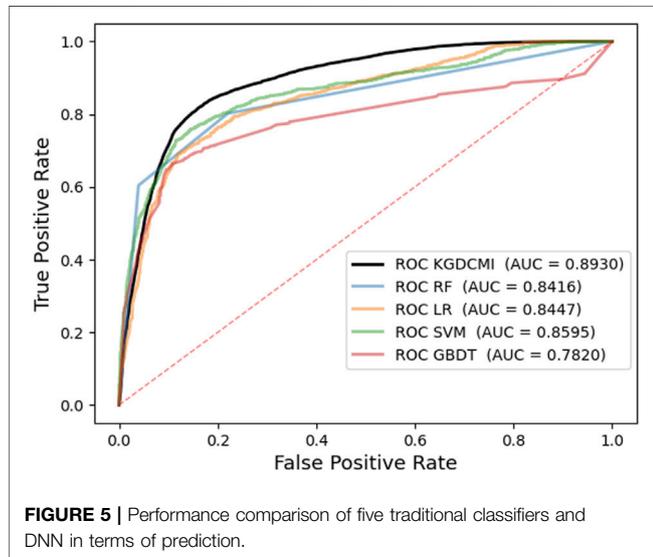
## Comparison With Related Models

To demonstrate the superiority of KGDCMI in the prediction of the circRNA-miRNA interaction, we compared the performance of the proposed method with the only existing model in this field. Since few models use computational methods to predict the circRNA-miRNA interactions, we also compared our method with several models in other highly related fields. In our comparative experiment, all models used 9589 pairs of data based on the Circbank and adopted five-fold cross-validation. At last, it was found that our method achieved the best results.

To the best of our knowledge, the CMIVGSD (Qian et al., 2021) is the only known computational framework for predicting the circRNA-miRNA interactions, which obtains the linear

**TABLE 2** | Performance comparison of five methods based on five-fold cross-validation.

Methods	DMFCDA	AE-RF	DMFMDA	NTSHMDA	CMIVGSD	KGDCMI
AUC	0.7321	0.7662	0.7922	0.8526	0.8804	0.9041
AUPR	0.7115	0.8239	0.8230	0.8772	0.8629	0.8937



features by SVD and uses the graph variational autoencoder to generate the non-linear features. At last, the LightGBM classifier is used to predict the circRNA-miRNA interactions. The deep matrix factorization prediction model (Liu et al., 2020) uses a neural network with an embedding layer to obtain a low-dimensional dense vector and realize matrix factorization, which is applied to obtain final scores of microbe-disease pairs. The NTSHMDA (Luo et al., 2018) constructs a heterogeneous network by similarity network and uses a method based on the random walk to predict the microbe-disease association. Furthermore, the AE-RF (Deepthi et al., 2021) integrates a similar network to construct features, uses a deep autoencoder to extract hidden features, and trains the random forest classifier to predict the association between circRNA and disease. The Deep Matrix Factorization CircRNA-Disease Association (Lu et al., 2020) uses a projection layer to learn latent features and applies a multi-layer neural network to predict the association between circRNA and disease.

**Table 2** shows the AUC and AUPR of our model and the other five models under the five-fold cross-validation. Our model obtains the best AUC and AUPR among the five models. Moreover, our model achieves the highest accuracy in the field of circRNA-miRNA interaction prediction, which is 2.37% higher than the second-best model.

## Comparison With Traditional Classifier

In the prediction method of the KGDCMI, we applied the DNN to fuse multidimensional features and classify the interaction between circRNA and miRNA. To reflect the superiority of our

classification strategy, we used some traditional classifiers to replace our classification method and compare the results. In a more concrete sense, we kept the multiple features of the KGDCMI unchanged and conducted independent experiments with four different classifiers, including the random forest classifier (Breiman 2001), logistic regression classifier (LaValley 2008), support vector machine classifier (Cortes and Vapnik 1995), and gradient boosting decision tree (Friedman 2001) to replace DNN for prediction, the prediction results are shown in **Figure 5**.

It can be seen in **Figure 5** that the results of the DNN classifier are significantly better than the five traditional classifiers. The results of these experiments indicated that DNN is applicable to the proposed method. The main reason for this is that the DNN has a better fusion for multidimensional features, which can give full play to the advantages of behavior characteristics and attribute characteristics in our method. Therefore, the DNN shows a better performance than the traditional classifiers when using the same feature description.

## Parameter Settings for Attribute Feature Dimension

This study used the  $K$ -mer method to extract attribute features from RNA sequences. For a given RNA, we can obtain  $4^k$ -dimensional vectors, which depend on the size of the value  $K$ . The value sizes {2,3,4,5} of  $K$  are frequently used values.

However, miRNA is a non-coding RNA transcript with an average length of 21 nucleotides (Buermans et al., 2010), and circRNA is a long-stranded RNA molecule, where over 14% of circRNAs in humans and over 10% of circRNA in mouse are more than 800 nucleotides in length (Wu et al., 2020). Due to the great difference in the sequence length between miRNA and circRNA, we need to adopt different  $K$  values for different RNAs to ensure that the most valuable features can be extracted.

In this part, we treated miRNA and circRNA with  $K \in \{2, 3, 4, 5\}$  respectively and compared the experimental results produced by different  $K$  values to obtain the most appropriate  $K$  value for each RNA.

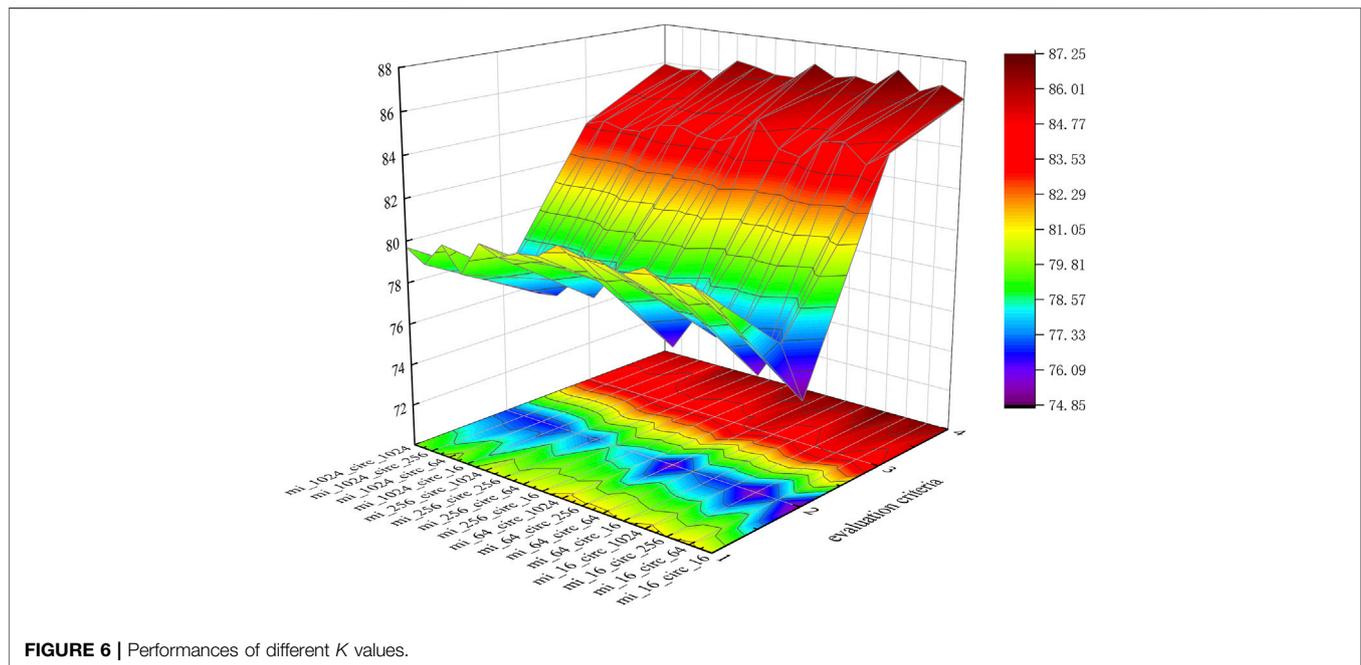
**Table 3** shows the performance of our model under different  $K$  values. To independently reflect the impact of different dimensions on the performance of the model, in the comparative experiment of  $K$  value, we do not add similarity descriptors and do not use SAE to extract features.

To intuitively represent each group of data, we visualize the data in three-dimensional space, as shown in **Figure 6**.

In **Figure 6**, evaluation criteria 1 through 4 represent Sen., Spec., Prec., and AUC., respectively. Meanwhile, mi\_16 indicates

**TABLE 3** | Performances of different  $K$  values.

$K$ -value	Sen. (%)	Spec. (%)	Prec. (%)	Acc. (%)	AUC. (%)
$K_{miRNA} = 2, K_{circRNA} = 2$	74.89	86.46	84.70	80.67	86.25
$K_{miRNA} = 2, K_{circRNA} = 3$	77.20	85.29	84.00	81.25	86.77
$K_{miRNA} = 2, K_{circRNA} = 4$	75.49	86.55	84.88	81.02	86.33
$K_{miRNA} = 2, K_{circRNA} = 5$	76.77	86.15	84.72	81.46	87.21
$K_{miRNA} = 3, K_{circRNA} = 2$	76.75	85.17	83.83	80.96	86.39
$K_{miRNA} = 3, K_{circRNA} = 3$	77.27	85.29	84.04	81.27	86.63
$K_{miRNA} = 3, K_{circRNA} = 4$	75.72	86.90	85.26	81.31	86.43
$K_{miRNA} = 3, K_{circRNA} = 5$	77.52	85.56	84.31	81.54	86.97
$K_{miRNA} = 4, K_{circRNA} = 2$	77.79	84.86	83.91	80.83	86.16
$K_{miRNA} = 4, K_{circRNA} = 3$	78.44	85.04	84.00	80.73	86.17
$K_{miRNA} = 4, K_{circRNA} = 4$	77.08	85.57	84.33	80.33	86.43
$K_{miRNA} = 4, K_{circRNA} = 5$	77.67	84.42	84.21	80.74	86.61
$K_{miRNA} = 5, K_{circRNA} = 2$	76.70	84.50	83.73	79.10	85.26
$K_{miRNA} = 5, K_{circRNA} = 3$	76.58	84.93	83.68	80.25	85.89
$K_{miRNA} = 5, K_{circRNA} = 4$	76.74	85.00	83.93	79.11	85.75
$K_{miRNA} = 5, K_{circRNA} = 5$	76.87	85.52	83.66	79.70	85.92

**FIGURE 6** | Performances of different  $K$  values.

that the dimensions of miRNA are 16, which means  $K_{miRNA} = 2$ , which is the same as for circRNA.

According to **Table 3** and **Figure 6**, when mi\_16\_circ\_1024 ( $K_{miRNA} = 2, K_{circRNA} = 5$ ), KGDCMI obtained the highest accurate value. When mi\_64\_circ\_1024 ( $K_{miRNA} = 3, K_{circRNA} = 5$ ) and mi\_16\_circ\_64 ( $K_{miRNA} = 2, K_{circRNA} = 3$ ), the method achieved the second and third highest results, respectively.

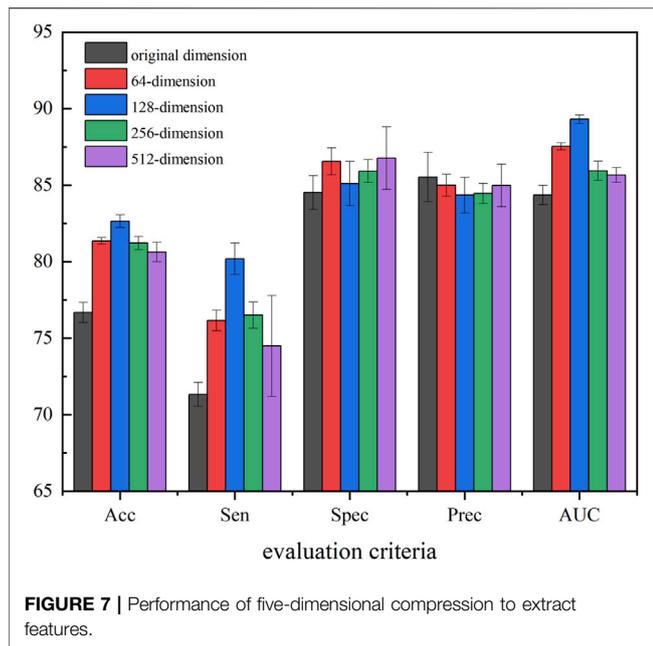
The  $K$  value of the  $K$ -mer algorithm represents the abundance of divided sequence fragments, larger  $K$  values tend to produce more representative features. However, for miRNA with a short sequence, when  $K > 3$ , the characteristic matrix generated by  $K$ -mer becomes sparser, and the performance of KGDCMI begins to decline. For circRNA with a long sequence, when  $3 < K \leq 5$ , the method achieves satisfactory results, which shows that the

algorithm can effectively extract features, but not sufficiently. When  $K = 5$ , we obtain the best performance.

According to our experiment, when  $K_{miRNA} = 2$  and  $K_{circRNA} = 5$ , attribute feature extraction based on  $K$ -mer can obtain the best effect. This is the first analysis of the different dimensions of feature extraction from RNA sequence length in the field of circRNA-miRNA interaction prediction, and we believe this research will provide a reference for future experiments.

## Different Dimensions Based on Sparse Autoencoder

In our method, to obtain the most representative feature descriptor, SAE was employed to process the resulting



eigenvectors to learn the hidden and relatively sparse features. To further evaluate the efficiency of the proposed feature extraction strategy and determine the best extraction dimension, we applied SAE to extract features with different dimensions as the final behavior feature. In detail, we selected 5-mers circRNA and 2-mers miRNA for which  $K$  showed the best performance in the previous section, and used 64-, 128-, 256-, and 512-dimension final behavior features extracted by SAE to perform experiments. Figure 7 provides experiment results of the different feature dimensions of the proposed method.

Figure 7 shows that compression of all four dimensions improves the accuracy of the model, which suggests that the use of SAE can effectively improve robustness, generalization ability, and accuracy. In all five evaluation criteria, 128-dimension compression achieves the best results in three evaluation criteria. Therefore, in this experiment, we set the SAE compression dimensions at 128.

## CASE STUDIES

To verify the performance of our model in the real environment, we implemented the case study based on our dataset. First, we used known circRNA-miRNA interactions to construct the features vectors and train the model. Second, the trained model was used to predict the unknown association interactions, and finally, we obtained the final score of prediction of each pair after sorting the results from large to small, the top 10 results are shown in Table 4. We can see from table 4 that 7 of the top 10 circRNA-miRNA interactions were confirmed in PubMed. The remaining three unconfirmed pairs of interaction were not verified by biological experiments, but there is no doubt that interaction is possible between them.

**TABLE 4 |** The top 10 results predicted in our model based on the dataset.

Num	CircRNA	miRNA	Evidence
1	hsa_circ_0006916	hsa-miR-522-3p	PMID:29726904
2	hsa_circ_0002142	hsa-miR-625-5p	PMID:30988674
3	hsa_circ_0000977	hsa-miR-874-3p	PMID:29454093
4	hsa_circ_0041089	hsa-miR-3192-5p	unconfirmed
5	hsa_circ_0041103	hsa-miR-103a-3p	PMID:27484176
6	hsa_circ_0007915	hsa-miR-106a-3p	PMID:28727484
7	hsa_circ_0000673	hsa-miR-767-3p	unconfirmed
8	hsa_circ_100242	hsa-miR-145-5p	PMID:32218853
9	hsa_circ_0092306	hsa-miR-197-3p	PMID:31689616
10	hsa_circ_0089776	hsa-miR-6752-5p	unconfirmed

## CONCLUSION

Due to the high labor cost and the production time of biological methods, computational methods have increasingly received attention and have been used to predict the association of different molecules. Predicting the interaction between circRNA and miRNA can not only save resources and time but also help to find the potential relationship between molecules and facilitate the understanding of complex disease mechanisms (Lei et al., 2021). At present, there is only one computational prediction model for circRNA and miRNA interaction, to the best of our knowledge. This study proposed a method called KGDCMI to predict the interaction between the circRNA and miRNA.

First, we used the  $K$ -mer algorithm and Gaussian kernel similarity to obtain the digital descriptor representing the attribute characteristics of RNA. Second, SAE was used to remove the redundancy and noise of attribute features to obtain the final attribute vectors. Next, we used HOPE to capture behavior information in molecular association networks. At last, we used a DNN for feature fusion and get the final predict score.

With the same dataset, our model obtained the highest AUC (90.41%) and AUPR (89.37%), which is 2.37% and 3.08% higher than the second-best model, respectively. Meanwhile, we compared and discussed the extraction of fragments with different lengths of molecular sequences for the first time, and we obtained the most suitable  $K$  value for miRNA and circRNA, which we believe will facilitate future research. The results predicted by the model were verified in case studies. There is no doubt that our model is an effective computational tool to predict the circRNA-miRNA interaction.

## DATA AVAILABILITY STATEMENT

The datasets for this paper can be found in the Circbank <http://www.circbank.cn/>, CircR2Cancer <http://www.biobdlab.cn:8000/>. The data and source code can be found at <https://github.com/1axin/KGDCMI>. KGDCMI is also publicly available as an online predictor at <http://120.77.11.78/KGDCMI/>.

## AUTHOR CONTRIBUTIONS

X-FW, C-QY L-PL, Z-HY, and W-ZH: conceptualization, methodology, software, validation, resources, and data curation. Z-HR Y-JG, and Y-CL: writing original draft preparation. All authors contributed to manuscript revision, read, and approved the submitted version.

## REFERENCES

- Armakola, M., Higgins, M. J., Figley, M. D., Barmada, S. J., Scarborough, E. A., Diaz, Z., et al. (2012). Inhibition of RNA Lariat Debranching Enzyme Suppresses TDP-43 Toxicity in ALS Disease Models. *Nat. Genet.* 44, 1302–1309. doi:10.1038/ng.2434
- Breiman, L. (2001). Random Forests. 45:5–32. doi:10.1023/a:1010933404324
- Buermans, H. P., Ariyurek, Y., van Ommen, G., den Dunnen, J. T., and 't Hoen, P. A. (2010). New Methods for Next Generation Sequencing Based microRNA Expression Profiling. *BMC Genomics* 11, 716–16. doi:10.1186/1471-2164-11-716
- Chen, L., and Shan, G. (2021). CircRNA in Cancer: Fundamental Mechanism and Clinical Potential. *Cancer Lett.* 505, 49–57. doi:10.1016/j.canlet.2021.02.004
- Chen, P., Nie, Z.-Y., Liu, X.-F., Zhou, M., Liu, X.-X., and Wang, B. J. N. (2022). *CircXRCC5, as a Potential Novel Biomarker, Promotes Glioma Progression via the miR-490-3p/XRCC5/CLC3 ceRNA Network.*
- Cocquerelle, C., Daubersies, P., Majerus, M. A., Kerckaert, J. P., and Bailleul, B. (1992). Splicing with Inverted Order of Exons Occurs Proximal to Large Introns. *EMBO J.* 11, 1095–1098. doi:10.1002/j.1460-2075.1992.tb05148.x
- Cortes, C., and Vapnik, V. (1995). Support-vector Networks. *Mach. Learn* 20, 273–297. doi:10.1007/bf00994018
- Deepthi, K., Jereesh, A. S., and Therapy (2021). Inferring Potential CircRNA-Disease Associations via Deep Autoencoder-Based Classification. *Mol. Diagn Ther.* 25, 87–97. doi:10.1007/s40291-020-00499-y
- Fan, C., Lei, X., Fang, Z., Jiang, Q., and Wu, F.-X. J. D. (20182018). *CircR2Disease: A Manually Curated Database for Experimentally Supported Circular RNAs Associated with Various Diseases.*
- Friedman, J. (2001). *Greedy Function Approximation: A Gradient Boosting Machine*, 1189–1232.
- Glažar, P., Papavasileiou, P., and Rajewsky, N. J. R. (2014). *circBase a database circular RNAs* 20, 1666–1670.
- Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., et al. (2001). Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs that Control *C. elegans* Developmental Timing. *Cell.* 106, 23–34. doi:10.1016/s0092-8674(01)00431-7
- Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA Circles Function as Efficient microRNA Sponges. *Nature* 495, 384–388. doi:10.1038/nature11993
- Hayes, J., Peruzzi, P. P., and Lawler, S. (2014). MicroRNAs in Cancer: Biomarkers, Functions and Therapy. *Trends Mol. Med.* 20, 460–469. doi:10.1016/j.molmed.2014.06.005
- Hsu, M.-T., and Coca-Prados, M. (1979). Electron Microscopic Evidence for the Circular Form of RNA in the Cytoplasm of Eukaryotic Cells. *Nature* 280, 339–340. doi:10.1038/280339a0
- Kulcheski, F. R., Christoff, A. P., and Margis, R. (2016). Circular RNAs Are miRNA Sponges and Can Be Used as a New Class of Biomarker. *J. Biotechnol.* 238, 42–51. doi:10.1016/j.jbiotec.2016.09.011
- Lan, W., Zhu, M., Chen, Q., Chen, B., Liu, J., Li, M., et al. (2020). *CircR2Cancer: A Manually Curated Database of Associations between circRNAs and cancers.2020.*
- LaValley, M. P. (2008). Logistic Regression. *Circulation* 117, 2395–2399. doi:10.1161/circulationaha.106.682658
- Lei, X., Mudiyansele, T. B., Zhang, Y., Bian, C., Lan, W., Yu, N., et al. (2021). *A Comprehensive Survey on Computational Methods of Non-coding RNA and Disease Association Prediction*, 22, bbaa350. doi:10.1093/bib/bbaa350
- Li, L., Gao, Z., Wang, Y.-T., Zhang, M.-W., Ni, J.-C., Zheng, C.-H., et al. (2021). SCMFMDA: Predicting microRNA-Disease Associations Based on Similarity Constrained Matrix Factorization. *PLoS Comput. Biol.* 17, e1009165. doi:10.1371/journal.pcbi.1009165

## FUNDING

This work was supported by the Science and Technology Innovation 2030-New Generation Artificial Intelligence Major Project (No.2018AAA0100103), and in part by the NSFC Program, under Grant 61873212, 62072378, and 62002297, and the National Natural Science Foundation of China (No. 62172338).

- Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., et al. (2015). Exon-intron Circular RNAs Regulate Transcription in the Nucleus. *Nat. Struct. Mol. Biol.* 22, 256–264. doi:10.1038/nsmb.2959
- Liu, M., Wang, Q., Shen, J., Yang, B. B., and Ding, X. (2019). Circbank: a Comprehensive Database for circRNA with Standard Nomenclature. *RNA Biol.* 16, 899–905. doi:10.1080/15476286.2019.1600395
- Liu, Y., Wang, S., Zhang, J., Zhang, W., Zhou, S., Li, W., et al. (2020). Dmfmda: Prediction of Microbe-Disease Associations Based on Deep Matrix Factorization Using Bayesian Personalized Ranking. *IEEE/ACM Trans. Comput. Biol. Bioinform* PP, 1763–1772. doi:10.1109/TCBB.2020.3018138
- Lu, C., Zeng, M., Zhang, F., Wu, F., Li, M., Wang, J., et al. (2020). Deep Matrix Factorization Improves Prediction of Human circRNA-Disease Associations. *IEEE J. Biomed. Health Inf.* PP, 891–899. doi:10.1109/JBHI.2020.2999638
- Luo, J., Long, Y., and bioinformatics (2020). NTSMDA: Prediction of Human Microbe-Disease Association Based on Random Walk by Integrating Network Topological Similarity. *IEEE/ACM Trans. Comput. Biol. Bioinform* 17, 1341–1351. doi:10.1109/TCBB.2018.2883041
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs Are a Large Class of Animal RNAs with Regulatory Potency. *Nature* 495, 333–338. doi:10.1038/nature11928
- Ng, W., and Ng, W. (2011). Wedding Dress. *Sparse autoencoder* 72, 1–19. doi:10.4016/31240.01
- Nigro, J. M., Cho, K. R., Fearon, E. R., Kern, S. E., Ruppert, J. M., Oliner, J. D., et al. (1991). Scrambled Exons. *Cell.* 64, 607–613. doi:10.1016/0092-8674(91)90244-s
- Ou, M., Cui, P., Pei, J., Zhang, Z., and Zhu, W. (2016). *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* doi:10.1145/2939672.2939751 Asymmetric Transitivity Preserving Graph Embedding
- Pan, J., You, Z.-H., Li, L.-P., Huang, W.-Z., Guo, J.-X., Yu, C.-Q., et al. (2022). *DWPPPI: A Deep Learning Approach for Predicting Protein-Protein Interactions in Plants Based on Multi-Source Information with a Large-Scale Biological Network*, 10. doi:10.3389/fbioe.2022.807522
- Qian, Y., Zheng, J., Zhang, Z., Jiang, Y., Zhang, J., and Deng, L. (2021). “CMIVGSD: circRNA-miRNA Interaction Prediction Based on Variational Graph Auto-Encoder and Singular Value Decomposition,” in *Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (IEEE)*. doi:10.1109/bibm52615.2021.9669875
- Qu, S., Yang, X., Li, X., Wang, J., Gao, Y., Shang, R., et al. (2015). Circular RNA: a New Star of Noncoding RNAs. *Cancer Lett.* 365, 141–148. doi:10.1016/j.canlet.2015.06.003
- Ren, Z.-H., Yu, C.-Q., Li, L.-P., You, Z.-H., Guan, Y.-J., Wang, X.-F., et al. (2022). *BioDKG-DDI: Predicting Drug-Drug Interactions Based on Drug Knowledge Graph Fusing Biochemical Information.*
- Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J., and Kleinschmidt, A. K. (1976). Viroids Are Single-Stranded Covalently Closed Circular RNA Molecules Existing as Highly Base-Paired Rod-like Structures. *Proc. Natl. Acad. Sci. U.S.A.* 73, 3852–3856. doi:10.1073/pnas.73.11.3852
- Siomi, H., and Siomi, M. C. (2010). Posttranscriptional Regulation of microRNA Biogenesis in Animals. *Mol. Cell.* 38, 323–332. doi:10.1016/j.molcel.2010.03.013
- Tao, D., Liu, Z., Wang, L., Li, C., Zhang, R., and Ni, N. (2022). CircPAG1 Interacts with miR-211-5p to Promote the E2F3 Expression and Inhibit the High Glucose-Induced Cell Apoptosis and Oxidative Stress in Diabetic Cataract. *Cell. Cycle* 21, 708–719. doi:10.1080/15384101.2021.2018213
- Van Laarhoven, T., Nabuurs, S. B., and Marchiori, E. (2011). Gaussian Interaction Profile Kernels for Predicting Drug-Target Interaction. 27:3036–3043. doi:10.1093/bioinformatics/btr500
- Wu, W., Ji, P., and Zhao, F. (2020). CircAtlas: an Integrated Resource of One Million Highly Accurate Circular RNAs from 1070 Vertebrate Transcriptomes. *Genome Biol.* 21, 101–114. doi:10.1186/s13059-020-02018-y

- Xu, H., Guo, S., Li, W., and Yu, P. (2015). The Circular RNA Cdr1as, via miR-7 and its Targets, Regulates Insulin Transcription and Secretion in Islet Cells. *Sci. Rep.* 5, 12453–12512. doi:10.1038/srep12453
- Yang, J., Qi, F., Tan, B., Dai, G., Chen, R., Wan, W., et al. (202215391). *circSPECC1 Promotes Bladder Cancer Progression via Regulating miR-136-5p/GNAS axis*.
- Yin, M-M., Liu, J-X., Gao, Y-L., Kong, X-Z., and Zheng, C-H. (2020). *NCPLP: A Novel Approach for Predicting Microbe-Associated Diseases with Network Consistency Projection and Label Propagation*.
- Zhang, Z., Yang, T., and Xiao, J. (2018). Circular RNAs: Promising Biomarkers for Human Diseases. *EBioMedicine* 34, 267–274. doi:10.1016/j.ebiom.2018.07.036
- Zhao, Z., Wang, K., Wu, F., Wang, W., Zhang, K., Hu, H., et al. (2018). circRNA Disease: a Manually Curated Database of Experimentally Supported circRNA-Disease Associations. *Cell. Death Dis.* 9, 475–482. doi:10.1038/s41419-018-0503-3
- Zhou, F., Yin, M-M., Jiao, C-N., Zhao, J-X., Zheng, C-H., Liu, J-X. J. I. ToN. N., et al. (2021). *Predicting miRNA-Disease Associations through Deep Autoencoder with Multiple Kernel Learning*.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Yu, Li, You, Huang, Li, Ren and Guan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.