A multi-scale neighbor topology guided transformer and Kolmogorov-Arnold network enhanced feature learning model for disease-related circRNA prediction

Ping Xuan, Haoyuan Li, Hui Cui, Zelong Xu, Toshiya Nakaguchi and Tiangang Zhang *

***Abstract*— As circular non-coding RNA (circRNA) is closely associated with various human diseases, identifying disease-related circRNAs can provide a deeper understanding of the mechanisms underlying disease pathogenesis. Advanced circRNA-disease association prediction methods mainly focus on graph learning techniques such as graph convolutional networks and graph attention networks. However, these methods do not fully encode the multi-scale neighbor topologies of each node, and the dependencies among the pairwise attributes. We propose a multi-scale neighbor topology-guided transformer with Kolmogorov-Arnold network (KAN) enhanced feature learning for circRNA and disease association prediction, termed MKCD. The model integrates multi-scale neighbor topology, complex relationships among multiple nodes, and the global and local dependencies of pairwise attributes. First, MKCD incorporates an adaptive multi-scale neighbor topology embedding construction strategy (AMNE), which generates neighbor topologies covering varying scopes of neighbors by performing random walks on a circRNA-disease-miRNA heterogeneous graph. Second, we design a dynamic multi-scale neighbor topology-guided transformer (DMTT) that leverages the multi-scale neighbor topologies to guide the learning of relationships among circRNA, miRNA, and disease nodes. The multi-scale neighbor topology is dynamically evolved, providing adaptive guidance to the transformer's learning process. Third, we establish a feature-gated network (FGN) to evaluate the importance of topological features derived from DMTT and the original node attributes. Finally, we propose an adaptive joint convolutional neural networks and KAN learning strategy (ACK) to learn the global and local dependencies of circRNA and disease node pair features. Comprehensive comparison experiments demonstrate the superior performance of our method over six state-of-the-art methods, and ablation experiments further validate the effectiveness of AMNE, DMTT, FGN and ACK innovations. Case studies on three diseases further validate the application value of our method in discovering reliable circRNA candidates for diseases of focus. The source code and datasets are freely available at https://github.com/pingxuan-hlju/MKCD.**

***Index Terms*— Multi-scale neighbor topology, neighbor topology-guided transformer, KAN enhanced feature learning, local and global dependencies of pairwise attributes, disease-related circRNA prediction.**

I. Introduction

CIRCRNA is a class of single-stranded circular non-coding RNA that lacks 5' and 3' polyadenylated tails [1]. Increasing evidence suggests that the abnormal expression of circRNAs is associated with the occurrence of various diseases [2], including cancers [3]–[5], immune system disorders [6], and cardiovascular diseases [7]–[9]. These associations provide important evidence for the diagnosis and treatment of diseases. CircRNAs are considered ideal biomarkers due to their high stability [10], [11]. For example, studies have found that the plasma expression level of hsa_circ_0001785 in breast cancer patients differs from that of healthy individuals before and after surgery, indicating that it can serve as an emerging biomarker for breast cancer diagnosis [12]. Li *et al.* [13] reported significantly elevated expression of hsa_circ_102584 in the peripheral blood of patients with systemic lupus erythematosus, indicating its potential as a biomarker for the disease. Therefore, identifying the associations between specific diseases and circRNAs can aid in the diagnosis and treatment of these diseases. Additionally, an increasing number of studies have shown that miRNAs interact with circRNAs and jointly participate in the pathogenesis of various diseases [14], [15]. One study points out that circGFRAI and miR-188–3P synergistically regulate the proliferation of non-small cell lung carcinoma (NSCLC) cells [16]. Therefore, the interaction between circRNAs and miRNAs, as well as the association between miRNAs and diseases, provide important auxiliary information for predicting circRNA-disease associations. As a result, our model integrates miRNA-related data. Recently, computational prediction methods have been

This work was supported by Natural Science Foundation of China (62172143, 62372282); Natural Science Foundation of Heilongjiang Province (LH2023F044); Guangdong Basic and Applied Basic Research Foundation (2024A1515010176); STU Scientific Research Initiation Grant (NTF22032).

Ping Xuan is with the Department of Computer Science and Technology, Shantou University, Shantou, 515063, China.

Haoyuan Li is with the Department of Computer Science and Technology, Shantou University, Shantou, 515063, China.

Hui Cui is with the Department of Computer Science and Information Technology, La Trobe University, Melbourne, 3083, Australia.

Zelong Xu is with the School of Information Science and Engineering, Yanshan University, Qinhuangdao, 066004, China.

Toshiya Nakaguchi is with the Center for Frontier Medical Engineering, Chiba University, Chiba, 2638522, Japan.

Tiangang Zhang is the corresponding author and he is with the School of Cyberspace Security, Hainan University, Haikou, 570228, China (e-mail: zhang@hlju.edu.cn).

proposed to predict the associations between miRNA and diseases [17], [18], the interactions between lncRNA and miRNA [19]–[21], and the interactions between circRNA and miRNA [22], [23]. Computational prediction methods can also discover associations between circRNAs and diseases, which can then be used to provide reliable disease-related circRNA candidates for subsequent biological experiments [24], [25].

Existing computational prediction methods can be categorized into three main types. The first type involves establishing network-based models to predict associations between circRNAs and diseases. KATZHCDA and PWCDA models calculate the association scores for each circRNA-disease pair based on the paths connecting them in heterogeneous networks [26], [27]. iCircDA-MF model proposed by Wei *et al.* incorporates gene information to construct a circRNA-gene-disease relationship network and predicts associations using matrix factorization [28]. However, these models suffer from limited association information.

The second category of prediction models is based on machine learning techniques to predict the associations between circRNAs and diseases. A couple of methods combine k-nearest neighbors to predict disease-associated circRNA candidates [29]–[31]. MLCDA model proposed by Wang *et al.* performs predictions using inductive matrix completion [32]. GBDTCDA and AE-RF are both decision tree-based prediction methods [33], [34], whereas CD-LNLP and RN-MFLP calculate the association scores between circRNAs and diseases through label propagation [35], [36]. However, these methods establish superficial prediction models, making it difficult to capture the deep relationships between circRNAs and diseases.

The third category focuses on developing models based on deep learning strategies, improving prediction performance by extracting complex and representative features. Several methods establish convolutional neural network-based models [37]–[40] or attention mechanism-based models [41] to predict disease-related circRNAs, yet they overlook the neighbor topology structure between multiple circRNA and disease nodes. Bi-SGTAR proposed by Li *et al.* employs an encoder with sparse gating to predict the propensity of all circRNA-disease associations [42]. This method also neglects the topology formed by circRNA and disease nodes. Other methods are based on graph convolutional networks [43]–[46] and graph attention networks [47], [48], and combinations of both [49], to learn deep relationships among nodes. However, these methods learn the features of each node from the entire graph, ignoring the global dependency learning of features between individual circRNA and disease node pairs.

We propose a novel association prediction model, **m**ulti-scale neighbor topology-guided transformer with **K**olmogorov-Arnold network enhanced feature learning for **c**ircRNA and **d**isease association prediction (MKCD), to learn multi-scale neighbor topologies, the relationships among circRNA, miRNA, and disease nodes, as well as the global and local dependencies between node pair features. The main contributions of this work are summarized as follows.

- We construct a circRNA-disease-miRNA heterogeneous graph that incorporates circRNA, disease, and miRNA nodes, along with their association, interaction, and similarity relationships. Each circRNA (disease, miRNA) node in this graph has multi-scale neighbors, with varying degrees of closeness to the target node. To effectively differentiate the contributions of neighbor topologies at different scales to node feature learning, we propose an adaptive multi-scale neighbor topology embedding construction strategy (AMNE). This strategy can adaptively determine the importance of each scale-specific neighbor topology and generate multi-scale neighbor topology embeddings.
- Most existing transformer-based models primarily focus on feature similarities between nodes while neglecting the topological structure among them. To address this limitation, we introduce dynamic multi-scale neighbor topology-guided transformer (DMTT), which encodes relationships among multiple circRNA, miRNA, and disease nodes. DMTT dynamically constructs an evolving neighbor topology and utilizes it to assess the importance of all circRNA, miRNA, and disease nodes in the heterogeneous graph.
- Multi-scale neighbor topology-introduced node features capture structural information, whereas the original features of circRNA, miRNA, and disease nodes provide rich biological details. To effectively integrate these complementary feature representations, we design a feature-gated network (FGN), which selectively determines the importance of topological and original features for downstream predictions.
- The features of circRNA and disease node pairs exhibit both local and global dependencies. To capture these dependencies, we propose a feature learning strategy for node pairs, ACK, which leverages Kolmogorov-Arnold network (KAN) [50] to learn global feature relationships and a multi-layer convolutional neural networks (CNNs) to extract local feature dependencies. Comparative experiments demonstrate that MKCD outperforms state-of-the-art methods in predicting circRNA-disease associations. Case studies validate that our method can effectively screen potential disease-related circRNA candidates.

II. Materials and methods

We propose a novel prediction model, MKCD (Figure 1), for predicting disease-related circRNA candidates. A circRNA-disease-miRNA heterogeneous graph is constructed to integrate the similarities, interactions, and associations among circRNA, disease, and miRNA (Figure 1(a)). MKCD consists of four components, each learning distinct information from the heterogeneous graph. The proposed DMTT encodes the complex relationships among multiple circRNA, disease, and miRNA nodes (Figure 1(c)). It further guides the transformer model's learning process through dynamically constructed multi-scale neighbor topology embeddings via AMNE (Figure 1(b)). To preserve more detailed information from node features, we introduce FGN (Figure 1(d)). The designed ACK is used to learn and fuse both global and local dependencies between the features of circRNA and disease node pairs

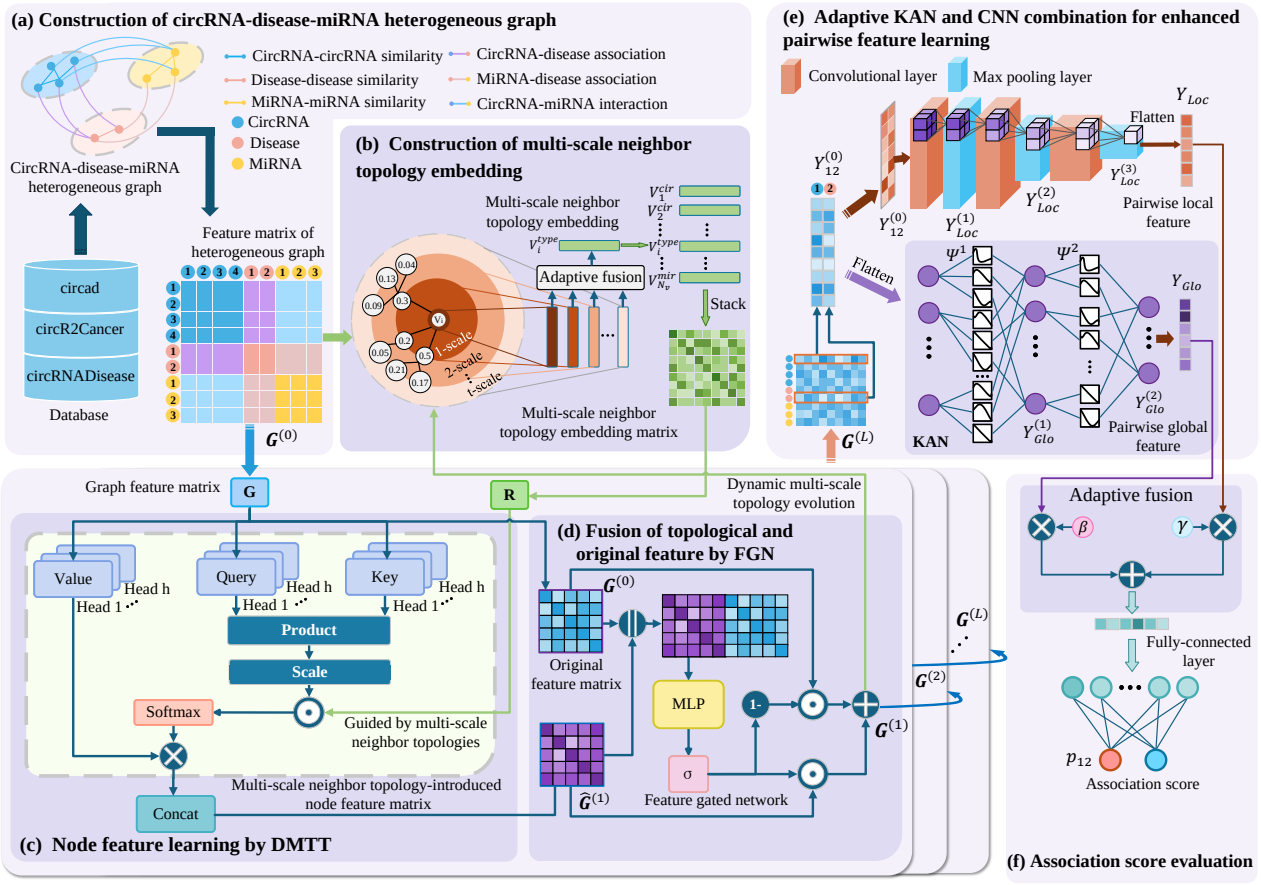

Fig. 1. The overall framework of MKCD. **(a)** Construction of the circRNA-disease-miRNA heterogeneous graph. **(b)** Adaptive multi-scale neighbor topology embedding construction strategy. **(c)** Learn the complex relationships among circRNA, miRNA, and disease nodes based on DMTT. **(d)** Fuse topological features and original node features by FGN. **(e)** Learn local and global dependencies of features of circRNA and disease node pairs based on ACK. **(f)** Adaptively fuse two representations and estimate circRNA-disease association scores.

(Figure 1(e)). These components work together to improve the capability of MKCD in predicting circRNA-disease associations.

A. Dataset

The dataset utilized in this study is derived from previous work [51], which comprises two datasets containing circRNA, disease and miRNA information. The first dataset covers 514 circRNAs, 62 diseases, and 564 miRNAs. The second dataset contains associations and interactions among 330 circRNAs, 79 diseases, and 245 miRNAs. We integrated these two datasets to form a new, larger dataset. This merged dataset consists of 989 circRNA-disease associations, 837 miRNA-disease associations, and 902 circRNA-miRNA interactions, covering a total of 834 circRNAs, 138 diseases, and 555 miRNAs. The original associations between circRNAs and diseases are obtained from the circR2Cancer database [52], the circad database [53], and the circRNADisease database [54].

B. CircRNA-disease-miRNA heterogeneous graph

We constructed a three-layer heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ using associations, interactions, and similarities among circRNAs, miRNAs, and diseases (Figure 2). The node set $\mathcal{V} = \{V^{cir} \cup V^{dis} \cup V^{mir}\}$ comprises the set of circRNA nodes V^{cir}, disease nodes set V^{dis}, and miRNA nodes set V^{mir}. An edge $e_{ij} \in \mathcal{E}$ connects a pair of nodes $v_i, v_j \in \mathcal{V}$, represented by the association and interaction matrix H and similarity matrix S.

The association and interaction matrix H among circRNAs, miRNAs, and diseases is defined as follows,

$$H = \begin{cases} H^{cir-dis} \in \mathbb{R}^{N_{cir} \times N_{dis}}, & \text{if } v_i \in V^{cir}, v_j \in V^{dis}; \\ H^{mir-dis} \in \mathbb{R}^{N_{mir} \times N_{dis}}, & \text{if } v_i \in V^{mir}, v_j \in V^{dis}; \\ H^{cir-mir} \in \mathbb{R}^{N_{cir} \times N_{mir}}, & \text{if } v_i \in V^{cir}, v_j \in V^{mir}; \end{cases} \tag{1}$$

where $H^{cir-dis}$, $H^{mir-dis}$, and $H^{cir-mir}$ denote the circRNA-disease association matrix, miRNA-disease association matrix, and circRNA-miRNA interaction matrix, respectively. N_{cir}, N_{dis}, and N_{mir} represent the number of circRNAs, diseases, and miRNAs in the dataset. $H_{ij} \in H^{cir-dis}(H^{mir-dis})$ represents the association between the circRNA (miRNA) node $v_i^{cir}(v_i^{mir})$ and the disease node v_j^{dis}. For a circRNA (miRNA) node v_i^{cir} (v_i^{mir}) and a disease node d_j^{dis}, if $H_{ij}^{cir-dis} = 1$ $(H_{ij}^{mir-dis} = 1)$, it indicates the existence of an association between them; conversely, $H_{ij}^{cir-dis} = 0$ $(H_{ij}^{mir-dis} = 0)$, no association has been

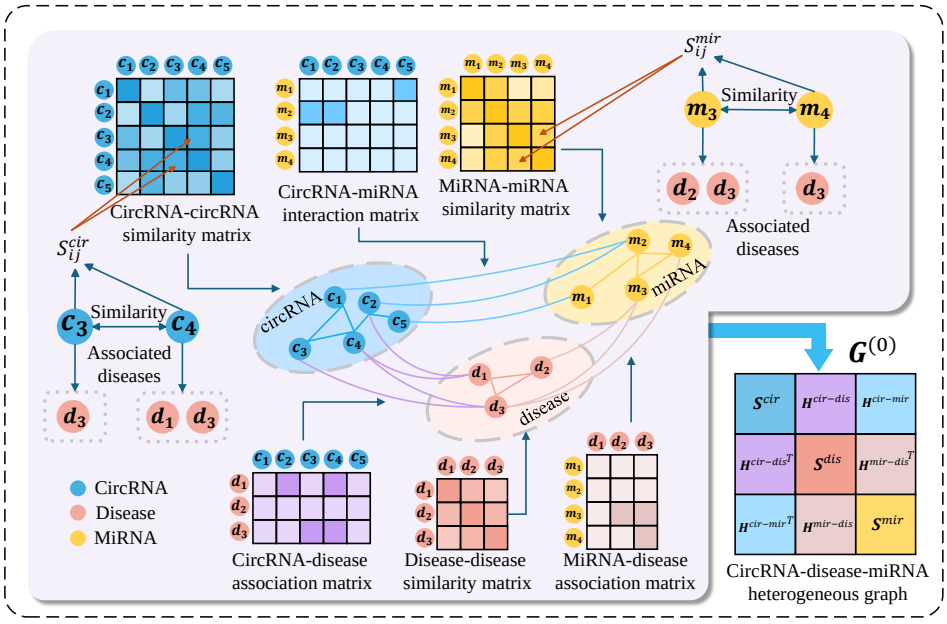

Fig. 2. Construction of the circRNA-disease-miRNA heterogeneous graph based on multi-source data.

observed. If $H_{ij}^{cir-mir} = 1$, an interaction is present between the circRNA node v_i^{cir} and the miRNA node v_j^{mir}; otherwise, $H_{ij}^{cir-mir} = 0$.

S represents the similarity matrix related to circRNAs, miRNAs, and diseases,

$$S = \begin{cases} S^{cir} \in \mathbb{R}^{N_{cir} \times N_{cir}}, & \text{if } v_i, v_j \in V^{cir}; \\ S^{dis} \in \mathbb{R}^{N_{dis} \times N_{dis}}, & \text{if } v_i, v_j \in V^{dis}; \\ S^{mir} \in \mathbb{R}^{N_{mir} \times N_{mir}}, & \text{if } v_i, v_j \in V^{mir}; \end{cases} \tag{2}$$

where S^{cir}, S^{dis}, and S^{mir} are the similarity matrix for circRNAs, diseases, and miRNAs, respectively. The values in S^{cir}, S^{dis}, and S^{mir} range from 0 to 1, reflecting similarity between two nodes of the same type, with higher values indicating greater similarity.

According to the method proposed by Wang *et al.* [55], the similarity between two disease nodes $[v_i^{dis}, v_j^{dis}]$ is calculated based on their directed acyclic graph (DAG). The similarities of circRNAs and miRNAs are calculated using the methods proposed by Wang *et al.* [55] and Chen *et al.* [56], respectively, where the similarity of a pair of circRNAs (miRNAs) is derived from the similarity between the two sets of diseases associated with them. For instance, suppose the i-th circRNA node v_i^{cir} is associated with N_i^{cir} diseases, which forms the set $\Omega_i^{cir} = \{d_{ik} | k = 1, \ldots, N_i^{cir}\}$, and the j-th circRNA node v_j^{cir} is associated with the disease set $\Omega_j^{cir} = \{d_{jl} | l = 1, \ldots, N_j^{cir}\}$. The similarity S_{ij}^{cir} between $[v_i^{cir}, v_j^{cir}]$ is determined by evaluating the similarity between Ω_i^{cir} and Ω_j^{cir},

$$S_{ij}^{cir} = \frac{\sum_{m=1}^{N_i^{cir}} S_m + \sum_{n=1}^{N_j^{cir}} S_n}{N_j^{cir} + N_j^{cir}}, \tag{3}$$

$$S_m = \max_{1 \leqslant n \leqslant N_j^{cir}} (DS(d_{im}, d_{jn})), \tag{4}$$

$$S_n = \max_{1 \leqslant m \leqslant N_i^{cir}} (DS(d_{jn}, d_{im})), \tag{5}$$

where $DS(d_{im}, d_{jn})$ is the semantic similarity between diseases d_{im} and d_{jn} which belong to Ω_i^{cir} and Ω_j^{cir} respectively. Similarly, we can calculate the similarity S_{ij}^{mir} for $[v_i^{mir}, v_j^{mir}]$.

Based on the constructed association (interaction) matrix H and the similarity matrix S, the original feature matrix of the heterogeneous graph $G^{(0)} \in \mathbb{R}^{N_v \times N_v}$ is defined as follows,

$$G^{(0)} = \begin{bmatrix} S^{cir} & H^{cir-dis} & H^{cir-mir} \\ H^{cir-dis^T} & S^{dis} & H^{mir-dis^T} \\ H^{cir-mir^T} & H^{mir-dis} & S^{mir} \end{bmatrix}, \tag{6}$$

where N_v denotes the total number of circRNAs, miRNAs, and diseases, and $H^{cir-dis^T}$ represents the transpose of the matrix $H^{cir-dis}$. The i-th row g_i of the matrix $G^{(0)}$ represents the node embedding of node $v_i \in \mathcal{V}$, which contains the associations and similarities involving v_i and all circRNAs, diseases, and miRNAs. The set $\{g_i | 0 \leqslant i < N_{cir}\}$ denotes the collection of node embeddings for all circRNAs. Furthermore, the sets $\{g_i | N_{cir} \leqslant i < N_{cir} + N_{dis}\}$ and $\{g_i | N_{cir} + N_{dis} \leqslant i < N_{cir} + N_{dis} + N_{mir}\}$ represent the collections of node embeddings for all diseases and miRNAs, respectively.

C. Adaptive multi-scale neighbor topology embedding construction strategy (AMNE)

In the circRNA-disease-miRNA heterogeneous graph, the node v_i has one-scale neighbors that can be reached in one step, or d-scale neighbors that can be reached in d $(d > 1)$ steps. The multi-scale neighbor topological structure formed by these neighboring nodes can provide important auxiliary information for predicting the associations between circRNAs and diseases. The contributions of low-scale (one-scale) neighbors and high-scale (d-scale) neighbors to the features learned for each node are different, thus, we proposed AMNE, which utilizes random walk with restart (RWR) to establish multi-scale neighbor topology embeddings (Figure 3). Taking v_i as

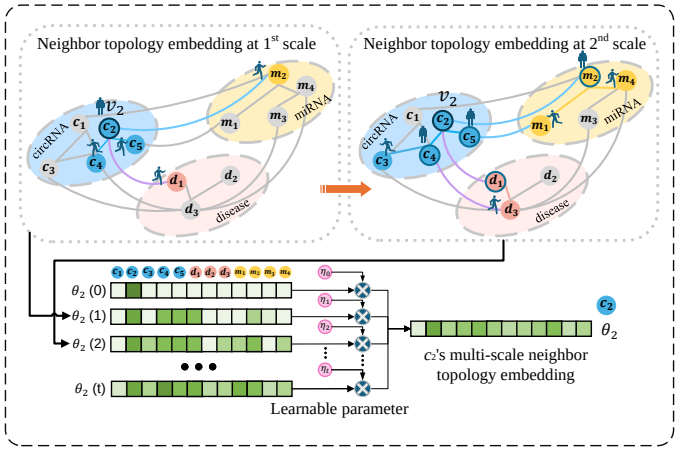

Fig. 3. Process of adaptively constructing multi-scale neighbor topology embeddings with RWR, illustrated by node c_2.

an example, the walker starts from v_i and performs random walks to travel to other nodes in the circRNA-disease-miRNA heterogeneous graph. The probability distribution of reaching all circRNA, disease, and miRNA nodes at time t is given by $\theta_i(t) \in \mathbb{R}^{1\times N_v}$,

$$\theta_i(t) = (1-\lambda)O^T\theta_i(t-1) + \lambda\theta_i(0), \tag{7}$$

where the j-th value of $\theta_i(t)$ denotes the probability that the walker starts from v_i reaches node v_j $(0 \leqslant j < N_v)$ after t steps. $\theta_i(0)$ is the initial one-hot vector, where the i-th position is 1 and all other positions are 0. λ is the probability that the walker restarts from the starting point; a larger value of λ results in a smaller movement range of the walker within the network. $O \in \mathbb{R}^{N_v\times N_v}$ is obtained from the row-normalized matrix of $G^{(0)}$, where $o_{ij} \in O$ represents the transition probability from v_i to v_j. $\theta_i(t)$ can be viewed as the probability distribution of reaching various nodes after t steps from v_i, thus it serves as the t-scale neighbor topology embedding of v_i.

According to Eq. (7), we can build neighbor topology embeddings from scale 0 to scale r. These scale neighbor topology embeddings are adaptively fused to obtain the multi-scale neighbor topology embedding θ_i for v_i,

$$\theta_i = \eta_0\theta_i(0) + \eta_1\theta_i(1) + \cdots + \eta_k\theta_i(k) + \cdots + \eta_t\theta_i(r), \tag{8}$$

where $\eta_k \in (0,1)$ are randomly initialized learnable parameters, and $\sum_{k=0}^{t}\eta_k = 1$.

After applying AMNE for each $v_i (0 \leqslant i < N_v)$, we can obtain the multi-scale neighbor embedding for all nodes. These embeddings are stacked vertically to form the embedding matrix $R \in \mathbb{R}^{N_v\times N_v}$,

$$R = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{N_v-1} \end{bmatrix}. \tag{9}$$

D. Node feature learning based on DMTT

Typically, multiple circRNAs and miRNAs form interactions and collaboratively participate in the processes of various diseases. Therefore, there are close relationships among the features of multiple circRNAs, miRNAs, and disease nodes, making it necessary to establish a self-attention mechanism to capture these relationships. Traditional transformers focus solely on the similarities between node features and do not fully exploit the topological structures formed between nodes, especially the multi-scale neighbor topological structures. Inspired by the transformer proposed by Vaswani *et al.* [57], we introduce a DMTT (dynamic multi-scale neighbor topology-guided transformer) mechanism (Figure 4) that utilizes the multi-scale neighbor topology embeddings established by AMNE to guide the learning of attention scores.

We use a multi-head attention mechanism to prevent single-head attention from getting stuck in local optima during training, reducing bias in the learning process. For the m-th attention head, we first establish the query matrix $Q_m^{(l)} \in \mathbb{R}^{N_v\times\frac{N_v}{h}}$, the key matrix $K_m^{(l)} \in \mathbb{R}^{N_v\times\frac{N_v}{h}}$, and the value matrix $V_m^{(l)} \in \mathbb{R}^{N_v\times\frac{N_v}{h}}$ as follows,

$$\begin{aligned} Q_m^{(l)} &= G^{(l-1)}W_m^{Q(l)} \\ K_m^{(l)} &= G^{(l-1)}W_m^{K(l)}, \\ V_m^{(l)} &= G^{(l-1)}W_m^{V(l)} \end{aligned} \tag{10}$$

where $G^{(l-1)} \in \mathbb{R}^{N_v\times N_v}$ is the feature matrix of the graph nodes that are input at layer l $(1 \leqslant l \leqslant L)$. When $l = 1$, $G^{(0)}$ represents the original feature matrix, and h is the number of attention heads. $Q_m^{(l)}$, $K_m^{(l)}$, and $V_m^{(l)}$ are obtained from $G^{(l-1)}$ through different linear projections, with $W_m^{Q(l)}$, $W_m^{K(l)}$, and $W_m^{V(l)} \in \mathbb{R}^{N_v\times\frac{N_v}{h}}$ being the corresponding weight matrices for the linear projections. Then, we perform a dot product operation on $Q_m^{(l)}$ and ${K_m^{(l)}}^T$ to obtain the attention score matrix $S_m^{(l)} \in \mathbb{R}^{N_v\times N_v}$,

$$S_m^{(l)} = \frac{Q_m^{(l)}{K_m^{(l)}}^T}{\sqrt{d}}, \tag{11}$$

where $d = \frac{N_v}{h}$ and $\sqrt{d}$ is a scaling factor used to adjust the magnitude of the attention scores to enhance numerical stability during the training process. The i-th row of $S_m^{(l)}$ records the attention scores from all circRNA, disease, and miRNA nodes to v_i.

After the $(l-1)$th layer of DMTT encoding, the topology between the nodes has changed. According to Eq. 7, we perform row normalization on $G^{(l-1)}$ to obtain the new $O^{(l)}$. A random walk is then conducted from each node as the starting point, i.e., Eq. 7 is executed again, resulting in new neighbor topology embeddings at different scales, which are adaptively fused into a new multi-scale neighbor topology embedding $\theta_i^{(l)}$ (Eq. 8). $\theta_i^{(l)}$ is vertically stacked to obtain the embedding matrix $R^{(l)}$ of the dynamically evolved layer (Eq. 9). The i-th row of $R^{(l)}$ records the neighbor topology of v_i with all other circRNA, disease, and miRNA nodes after the $(l-1)$th layer transformer encoding. We perform a Hadamard product operation between $R^{(l)}$ and $S_m^{(l)}$. This approach allows the multi-scale neighbor topology embeddings to guide the learning of attention scores. We establish the multi-scale neighbor topology-introduced attention score matrix $\widetilde{S}_m^{(l)} \in \mathbb{R}^{N_v\times N_v}$ as

follows,

$$\widetilde{S}_m^{(l)} = S_m^{(l)} \odot R^{(l)}, \tag{12}$$

where $\odot$ denotes the Hadamard product operation. Multiplying $\widetilde{S}_m^{(l)}$ with $V_m^{(l)}$ produces the node features $Z_m^{(l)} \in \mathbb{R}^{N_v \times \frac{N_v}{h}}$ learned by the m-th attention head,

$$Z_m^{(l)} = softmax(\widetilde{S}_m^{(l)})V_m^{(l)}. \tag{13}$$

Finally, by concatenating the node features learned by all h attention heads, we obtain the multi-scale neighbor topology-introduced node feature matrix $\hat{G}^{(l)} \in \mathbb{R}^{N_v \times N_v}$ for layer l,

$$\hat{G}^{(l)} = \Big\|_{m\in[1,h]} Z_m^{(l)}, \tag{14}$$

where $\|$ denotes the concatenation operation. The i-th row of $\hat{G}^{(l)}$ records the features of v_i learned at layer l.

E. Fusion of multiple types of features based on FGN

In DMTT, the feature matrix $G^{(l-1)}$ that is input at layer l contains more detailed information about each node, while the feature matrix $\hat{G}^{(l)}$ learned based on DMTT places greater emphasis on the information guided by multi-scale neighbor topology embedding. Therefore, it is necessary to incorporate $G^{(l-1)}$ into the feature learning process at layer l. To integrate the information contained in $\hat{G}^{(l)}$ and $G^{(l-1)}$, we establish an FGN (feature-gated network) after DMTT at layer l, with the weight matrix denoted as $\alpha^{(l)}$,

$$\alpha^{(l)} = \sigma(W^{gate(l)}(\hat{G}^{(l)} \| G^{(l-1)}) + b^{gate(l)}), \tag{15}$$

where $W^{gate(l)}$ and $b^{gate(l)}$ are learnable weight matrices and bias, and σ is the Sigmoid activation function. All parameters of the FGN are randomly initialized and are learnable during the training process, allowing it to discern the more significant features in $\hat{G}^{(l)}$ and $G^{(l-1)}$.

The FGN-enhanced node feature representation is denoted as $G^{(l)} \in \mathbb{R}^{N_v \times N_v}$,

$$G^{(l)} = \alpha^{(l)} \odot \hat{G}^{(l)} + (1 - \alpha^{(l)}) \odot G^{(l-1)}, \tag{16}$$

where $\odot$ denotes the Hadamard product operation. When $l \neq L$, $G^{(l)}$ will serve as the input for the next layer of DMTT, where $l = L$, $G^{(L)}$ represents the final node feature representation matrix.

F. Adaptive joint CNNs and KAN enhanced pairwise feature learning strategy

If the circRNA node v_i^{cir} and the disease node d_j^{dis} have similarities, associations, or interactions with the same circRNAs, diseases, and miRNAs, then the node pair $[v_i^{cir}, d_j^{dis}]$ is more likely to be associated. Based on this biological premise, we vertically stack the feature representations obtained from the FGN-enhanced DMTT for both circRNA and disease nodes, forming a node pair-level feature representation $Y_{ij}^{(0)} \in \mathbb{R}^{2\times N_v}$,

$$Y_{ij}^{(0)} = \begin{bmatrix} G_i^{(L)} \\ G_j^{(L)} \end{bmatrix}, \tag{17}$$

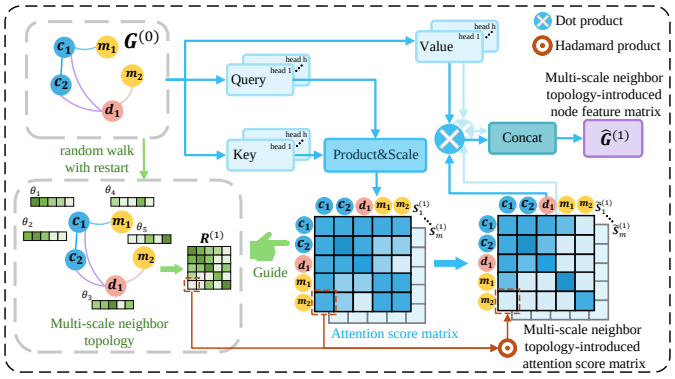

Fig. 4. Illustration of node feature learning process guided by multi-scale neighbor topology, using $\boldsymbol{G^{(0)}}$ as an example.

where the sets $\{G_i^{(L)} | 0 \leqslant i < N_{cir}\}$ and $\{G_j^{(L)} | N_{cir} \leqslant j < N_{cir} + N_{dis}\}$ represent the node feature representations of all circRNAs and diseases in $G^{(L)}$, respectively.

The ACK (adaptive joint CNNs and KAN learning strategy) we designed will further extract the information within the circRNA-disease node pairs. The KAN module we established learns the pairwise node representations from a global perspective, while the CNN module focuses more on extracting local information of the pairwise node representations.

1) Pairwise global feature learning based on KAN: Compared to the multi-layer perceptron (MLP), KAN replaces weight parameters and activation functions with learnable functions to adaptively learn the weights of the connections between neurons. This learnable function is typically composed of multiple stacked spline functions, enabling the KAN network to better capture the complex relationships among the features in $Y_{ij}^{(0)}$. Through multiple layers of the KAN network, a feature representation of the node pair $[v_i^{cir}, d_j^{dis}]$ is learned from a global perspective. The input feature representation of the node pair at the l-th layer $Y_{Glo}^{(l-1)}$ is transformed into the pairwise global feature representation $Y_{Glo}^{(l)}$ after passing through the l-th layer of the global feature learning module based on the KAN network,

$$Y_{Glo}^{(l)} = KAN^{(l)}(Y_{Glo}^{(l-1)}) = \Psi^{(l)}(Y_{Glo}^{(l-1)}), \tag{18}$$

where $KAN^{(l)}$ denotes the l-th KAN layer. When $l = 1$, $Y_{Glo}^{(0)}$ is obtain by flattening the original circRNA-disease node pair-level feature representation $Y_{ij}^{(0)}$. When $l = L$, $Y_{Glo}^{(L)}$ represents the final pairwise global feature representation $Y_{Glo} \in \mathbb{R}^{1\times f}$, where f is the feature dimension.

The matrix $\Psi^{(l)}$ is composed of learnable functions for the l-th KAN layer, containing a total of $n^{(l-1)} \times n^{(l)}$ learnable functions. Therefore, $\Psi^{(l)}$ is defined as,

$$\Psi^{(l)} = \begin{pmatrix} \psi_{1,1} & \psi_{1,2} & \cdots & \psi_{1,n^{(l-1)}} \\ \psi_{2,1} & \psi_{2,2} & \cdots & \psi_{2,n^{(l-1)}} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{n^{(l)},1} & \psi_{n^{(l)},2} & \cdots & \psi_{n^{(l)},n^{(l-1)}} \end{pmatrix}, \tag{19}$$

where $n^{(l-1)}$ and $n^{(l)}$ represent the number of neurons in the $(l-1)$th and l-th layers, respectively. The $\psi_{i,j}$ represents the learnable function associated with the edge connecting the j-th neuron $z_j^{(l-1)} (1 \leqslant j \leqslant n^{(l-1)})$ in the $(l-1)$th layer to the

i-th neuron $z_i^{(l)}(1 \leqslant i \leqslant n^{(l)})$ in the l-th layer. The output neuron $z_i^{(l)}$of the l-th layer is,

$$z_i^{(l)} = \sum_{j=1}^{n^{(l-1)}} \psi_{i,j}(z_j^{(l-1)}), \tag{20}$$

where the $\psi_{i,j}$ is composed of a basis function and a B-spline function,

$$\psi_{i,j}(z_j^{(l-1)}) = \omega_{i,j}^b b(z_j^{(l-1)}) + \omega_{i,j}^s spline_{i,j}(z_j^{(l-1)}), \tag{21}$$

$$spline_{i,j} = \sum_{k=1}^{n_{grid}} c_k B_k, \tag{22}$$

where b is the basis function $SiLU$, and $\omega_{i,j}^b$ and $\omega_{i,j}^s$ are learnable weight parameters. The B-spline function $spline_{i,j}$ is a linear combination of n_{grid} B-spline basis functions B_k, with learnable coefficients c_k.

2) Pairwise local feature learning based on CNNs: We have established CNNs to learn the local features of circRNA-disease node pairs. In the CNNs, each block consists of a convolution layer followed by a pooling layer. In the l-th block $(1 \leqslant d \leqslant D)$, given the feature representation of a circRNA-disease node pair $Y_{Loc}^{(d-1)}$, we use the convolution and pooling operations to extract its pairwise local features $Y_{Loc}^{(d)}$,

$$Y_{Loc}^{(d)} = max(\tau(W_{conv}^{(d)} * Y_{Loc}^{(d-1)} + d_{conv}^{(d)})), \tag{23}$$

where * denotes the convolution operation, $W_{conv}^{(d)}$ and $b_{conv}^{(d)}$ are the sets of convolution kernels and bias, respectively, τ is the Leaky ReLU activation function, and max represents the max pooling operation. $Y_{Loc}^{(0)}$ is the original feature representation of the circRNA-disease node pair $Y_{ij}^{(0)}$. When $d = D$, the dimension of $Y_{Loc}^{(D)}$ is reduced and flattened to obtain the final pairwise local feature representation $Y_{Loc} \in \mathbb{R}^{1\times f}$.

3) Adaptive fusion of pairwise local features and global features: The pairwise local feature representation Y_{Loc} and global feature representation Y_{Glo} hold varying degrees of importance for the feature representation learning of each circRNA (disease) node. We assign a randomly initialized learnable weight parameter s_β for Y_{Glo} and s_γ for Y_{Loc}, and after normalization, we obtain β and γ,

$$\beta = \frac{e^{s_\beta}}{e^{s_\beta} + e^{s_\gamma}}, \quad \gamma = \frac{e^{s_\gamma}}{e^{s_\beta} + e^{s_\gamma}}. \tag{24}$$

The final feature representation of the circRNA-disease node pair is defined as $Y_F \in \mathbb{R}^{1\times f}$,

$$Y_F = \beta \cdot Y_{Glo} + \gamma \cdot Y_{Loc}, \tag{25}$$

where $\cdot$ denotes scalar multiplication.

G. Association score evaluation and optimization

We utilize a fully connected layer to derive the association prediction score vector $p_{ij} \in \mathbb{R}^{1\times 2}$ for the node pair $[v_i^{cir}, d_j^{dis}]$,

$$p_{ij} = softmax(W_{Fin} Y_F + b_{Fin}), \tag{26}$$

where W_{Fin} and b_{Fin} are the weight matrix and bias of the fully connected layer, respectively. The vector $p_{ij} = [p_{pos}, p_{neg}]$ represents the probabilities that v_i^{cir} is associated with d_j^{dis} or not, denoted as p_{pos} and p_{neg}, respectively.

During the training process, we employ the AdamW algorithm and back propagation to optimize our model. We use the cross-entropy function to estimate the model's loss,

$$loss = -\sum_{(i,j)\in N} \left[y_{ij} \log(p_{pos}) + (1 - y_{ij}) \log(p_{neg}) \right], \tag{27}$$

where N denotes the sample set of all circRNA-disease node pairs. y_{ij} represents the true association label between the circRNA node v_i^{cir} and the disease node d_j^{dis}. When there is an association between v_i^{cir} and d_j^{dis}, $y_{ij} = 1$; otherwise, $y_{ij} = 0$.

III. Experimental Evaluations and Discussions

A. Parameter settings

In the AMNE module, we utilize the 0 to 2-scale neighbor topology to construct multi-scale neighbor topology embeddings, with the restart probability λ for random walk set to 0.7. For the DMTT module, the number of layers L is set to 2, and in each layer of the DMTT, the number of attention heads h is set to 4. In the pairwise local feature learning module, we employ three blocks, where the convolution kernel sizes of the first two blocks are 2×2, and the convolution kernel of the third block is 1×2. The number of channels in the blocks are taken from the set {32, 64, 128}. The pooling layer of the first block uses a window size of 2×2, while the pooling windows for the remaining two blocks are both set to 1×7. In the pairwise global feature learning module, we establish a 2-layer KAN network with the number of neurons set to 1024 and 256, respectively, and the number of n_{grid} for the B-spline function is set to 5. We train MKCD using an Nvidia GeForce RTX 4060, utilizing the PyTorch framework and optimizing using the AdamW algorithm. The training process consisted of 40 epochs, with a batch size of 32, a learning rate of 0.001, and a weight decay of 0.0001.

B. Evaluation metrics

We employ five-fold cross-validation to evaluate the predictive performance of MKCD and other comparative methods. All known circRNA-disease associations are treated as positive samples and randomly divided into five equal parts, while all unobserved circRNA-disease associations are considered as negative samples. In each fold, we use four parts of positive samples and an equal number of randomly selected negative samples as the training set, while the remaining positive samples and all unselected negative samples constitute the test set.

We select the area under the receiver operating characteristic curve (AUC) [58], the area under the precision-recall curve (AUPR) [59], F1 score and Precision as evaluation metrics. AUC and AUPR are calculated separately for each fold, and the averages of these five folds result in the final AUC and AUPR scores. F1 score and Precision are calculated based on the threshold Δ of the predicted association probability, where Δ is set to 0.5. When the model predicts a probability greater than 0.5 for the association between a pair of circRNA and

TABLE I
RESULTS OF ABLATION EXPERIMENTS OF MKCD.

AMNE	DMTT	FGN	ACK	Average AUC	Average AUPR
✗	✓	✓	✓	0.913	0.203
✓	✗	✗	✓	0.909	0.195
✓	✓	✗	✓	0.923	0.221
✓	✓	✓	✗	0.931	0.240
✓	✓	✓	✓	**0.946**	**0.267**

TABLE II
PREDICTION PERFORMANCE COMPARISON FOR VARYING MAXIMUM SCALES OF THE MULTI-SCALE NEIGHBOR TOPOLOGY.

r	**1**	**2**	**3**	**4**	**5**
AUC	0.944	**0.946**	0.934	0.930	0.924
AUPR	0.248	**0.267**	0.261	0.250	0.234

disease, the circRNA-disease pair is predicted as a positive instance; otherwise, it is predicted as a negative instance. Furthermore, considering that biologists typically choose candidates from the top of the ranked list for further validation, we calculate the recall rate of the top k disease-related circRNAs.

C. Ablation experiments

To validate the effectiveness of AMNE (adaptive multi-scale neighbor topology embedding construction strategy), DMTT (dynamic multi-scale neighbor topology-guided transformer), FGN (feature-gated network), and ACK (adaptive joint CNNs and KAN learning strategy), we conducted a series of ablation experiments (Table I). We sequentially remove the AMNE, FGN, and ACK modules from MKCD and calculate the corresponding AUC and AUPR. The DMTT learns the multi-scale neighbor topology-introduced node feature matrix $\hat{G}^{(l)}$, while FGN performs adaptive fusion on the original feature matrix $G^{(l-1)}$ and $\hat{G}^{(l)}$. When DMTT is removed, FGN loses its functionality as it relies on $\hat{G}^{(l)}$. Therefore, if DMTT is removed, FGN must also be removed simultaneously. We observed that when all modules were retained, the complete model MKCD achieved the best predictive performance, with AUC and AUPR values of 0.946 and 0.267, respectively. When the AMNE module was removed, the AUC and AUPR decreased by 3.3% and 6.4%, respectively, indicating that the introduction of multi-scale neighbor topology embedding plays a crucial role in enhancing the accuracy of circRNA and disease association predictions. The removal of DMTT and FGN resulted in a 3.7% drop in AUC and a 7.2% drop in AUPR, confirming the necessity of utilizing the multi-scale neighbor topology formed by circRNA, disease, and miRNA nodes to learn node feature representations. The complete model improved AUC and AUPR by 2.3% and 4.6%, respectively, compared to when FGN was ignored, suggesting that the incorporation of detailed features benefits the learning of node features. Finally, the removal of ACK led to a decrease of 1.5% in AUC and 2.7% in AUPR, demonstrating the effectiveness of the adaptive fusion of pairwise local features and global features in enhancing circRNA-disease association prediction performance.

TABLE III
PREDICTION PERFORMANCE COMPARISON FOR VARYING RESTART PROBABILITY OF THE RANDOM WALK.

λ	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
AUC	0.917	0.931	0.929	0.932	0.937	0.940	**0.946**	0.940	0.939
AUPR	0.221	0.235	0.232	0.240	0.241	0.254	**0.267**	0.250	0.242

The results of the ablation experiments highlight that the combination of DMTT with FGN has the greatest contribution. This is because FGN enhances the model by introducing more details from original features, while DMTT encodes the relationships among multiple features of circRNA, disease, and miRNA nodes. AMNE contributes the second largest to the prediction results, as AMNE effectively introduces multi-scale neighbor topology embedding into node feature learning. The ACK module enhances the model's ability to predict complex circRNA-disease associations by introducing the nonlinear feature learning capability of the KAN. However, since the convolutional network itself can effectively capture spatial features and local relationships, the gain from the ACK module is relatively limited. The additional features learned by the ACK module overlap to some extent with those captured by the convolutional network, leading to its relatively small contribution.

The maximum scale of the multi-scale neighbor topology is denoted as r. To evaluate its impact on circRNA-disease association prediction performance, we test values of r from $\{1, 2, 3, 4, 5\}$. As shown in Table II, our model achieves the optimal performance in terms of the AUC and AUPR metrics when $r = 2$. However, the model's prediction performance declines when $r = 3$, 4, and 5. This may be due to noise from including more neighbors. The restart probability of the random walk is denoted as λ. We conduct experiments with values selected from $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. According to Table III, when $\lambda = 0.7$, the model achieves the highest AUC and AUPR values. As the value of λ decreases, the exploration range increases. When λ is smaller than 0.7 the prediction performance of the model decreases. This may occur because a larger exploration range incorporates information from less relevant distant nodes, introducing noise that degrades performance. When λ increases above 0.7, the exploration range decreases, focusing more on the starting node, which limits the information from distant nodes.

The number of spline functions is denoted as n_{grid}. To evaluate the impact of n_{grid} on circRNA-disease association prediction performance, we select values for n_{grid} from $\{3, 4, 5, 6, 7\}$ for the experiment. The experimental results are shown in Supplementary Table 9. When $n_{grid} = 5$, our model achieves the optimal performance in both AUC and AUPR metrics. The variation in n_{grid} results in an AUC change range of 0.02, with a standard deviation of 0.0073, and an AUPR change range of 0.031, with a standard deviation of 0.0105. The order of the spline functions affects the smoothness of the function. The order is denoted as k, and the range of k is $\{1, 2, 3, 4\}$. When $k = 3$, the model achieves the highest values of AUC and AUPR (Supplementary Table 10). The variation in k results in an AUC change range of 0.019, with a standard deviation of 0.0078, and an AUPR change range of 0.055, with a standard deviation of 0.0201. The experimental results indicate that the order of the spline functions, k, is more sensitive to the final performance than n_{grid}.

We also conduct experiments on different numbers of

TABLE IV
PREDICTION PERFORMANCE COMPARISON FOR DIFFERENT SUBSTITUTION STRATEGIES.

Strategy	MKCD	MKCD_noKAN	MKCD_Att	MKCD_Con
AUC	**0.946**	0.932	0.930	0.926
AUPR	**0.267**	0.256	0.263	0.243

DMTT layers, attention heads, and the number of neurons in KAN. The analysis is added to the Supplementary File SF2.

To compare the performance differences between different neighbor topology generation strategies, we use Breadth-First Search (BFS) and Personalized PageRank as comparison methods, denoted as MKCD_BFS and MKCD_PR, respectively. The experimental results (Supplementary Table 5) show that the RWR-based AMNE module performs the best in terms of AUC and AUPR, with values of 0.946 and 0.267, respectively. Compared to MKCD_BFS, our model improves AUC by 2.9% and AUPR by 5.0%; compared to MKCD_PR, our model improves AUC by 1.0% and AUPR by 2.4%. The superior performance of RWR is attributed to its restart mechanism, which effectively balances local and global topology information, making it suitable for capturing the complex multi-scale relationships in the circRNA-disease association network.

To verify whether the multi-scale neighbor topology in MKCD outperforms single-order neighbor topologies in prediction performance, we test the model's prediction performance using only the i-th (i = 1, 2, 3) order neighbor topology. As shown in Supplementary Table 6, when using the multi-scale fused neighbor topology, our model achieves the optimal performance in both AUC and AUPR, with values of 0.946 and 0.267, respectively. Compared to our model, using only the 1-order neighbors results in a 0.4% and 1.7% decrease in AUC and AUPR, respectively; using only the 2-order neighbors leads to a 0.9% and 2.5% decrease, while using only the 3-order neighbors causes a 2.5% and 3.7% drop. The experimental results demonstrate that by adaptively weighted fusion of multi-order neighbor topologies, MKCD can effectively capture complementary information from different order neighbors, thereby improving the model's prediction performance.

To verify the superiority of KAN in capturing complex relationships between pairwise attributes, we compare the performance of the MLP-based and KAN-based models. As shown in Table IV, MKCD_noKAN represents the model where KAN is replaced by MLP in the global feature learning module. The model with KAN as the global feature learning module achieves an AUC and AUPR that are 1.4% and 1.1% higher than those of MKCD_noKAN, respectively. This indicates that KAN outperforms MLP in capturing the global feature relationships between node pairs.

The original node feature $G^{(l-1)}$ contains more detailed information, while DMTT learns the multi-scale neighbor topology-introduced node feature matrix $\hat{G}^l$. The FGN can adaptively learn the importance of both to better fuse these two features. In addition to fusing $G^{(l-1)}$ and $\hat{G}^l$ through FGN, other fusion methods are also available, such as attention-based mechanisms or concatenationg. The comparison experimental results of FGN, attention mechanisms, and concatenation are listed in Table IV. MKCD_Att represents the model using an attention mechanism as the fusion strategy, while MKCD_Con represents fusion strategy through concatenation and fully connected layers. The results show that when FGN is used as the fusion strategy, the model's prediction performance is optimal. compared to MKCD_Att, the AUC and AUPR are improved by 1.6% and 0.4%, respectively, and compared to MKCD_Con, they are improved by 2.0% and 2.4%, respectively. This demonstrates the superiority of FGN in feature fusion and integration.

The learnable parameters for scales 0 to 2 are denoted as η_0, η_1, and η_2, and they are randomly initialized. For each batch of data, we calculate the loss function for circRNA-disease association prediction and update the values of $\eta_0 \sim \eta_2$ using backpropagation. The training process consists of 40 epochs. As shown in Supplementary Figure 3, the values of η_0, η_1, and η_2 fluctuate significantly during the first 10 epochs, gradually converge between epochs 10 and 20, and stabilize after 20 epochs. The final values of η_0, η_1, and η_2 stabilize at 0.295, 0.487, and 0.218, respectively. The experimental results indicate that the weight proportion of the 1-scale neighbor topology embedding is the largest, which is likely because it includes neighbors that are directly connected to the target node, resulting in stronger associations. In contrast, the 2-scale neighbor topology has a smaller weight proportion, possibly due to its indirect, two-step connections to the target node.

To evaluate whether RWR (random walk with restart) significantly increases the training and inference time of the model, we compare the model with and without RWR, where the model without RWR is denoted as MKCD_noRWR. We also replace the KAN network in the pairwise global feature learning module with MLP, denoted as MKCD_noKAN, to analyze the time cost of KAN. The training time is defined as the time for one epoch, and inference time is the time predict whether a pair of circRNA-disease is associated. As shown in Supplementary Table 7, incorporating RWR adds 0.061 seconds to training time per epoch (an 8.63% increase) and 0.016 ms to inference time per circRNA-disease pair. The full MKCD model with KAN and RWR increases training time for 40 epochs by approximately 0.105 seconds compared to MKCD_noKAN (a 15.84% increase) and requires 0.028 ms more for inference per circRNA-disease pair. These results indicate that RWR and KAN only slightly increased the model's training and inference time.

In addition, we also analyze the performance of our model on the larger-scale dataset Pokec [60], and the corresponding description is added in Supplementary Table 5.

D. Comparison with other methods

We compared MKCD with six advanced methods for predicting circRNA-disease associations, including SGFCCD [43], MLNGCF [47], MDGF-MCEC [45], Bi-SGTAR [42], GraphCDA [49], and MPCLCDA [44]. Each method was trained using the optimal parameters provided in their original papers, and the same training and testing datasets were utilized in cross-validation to ensure fairness.

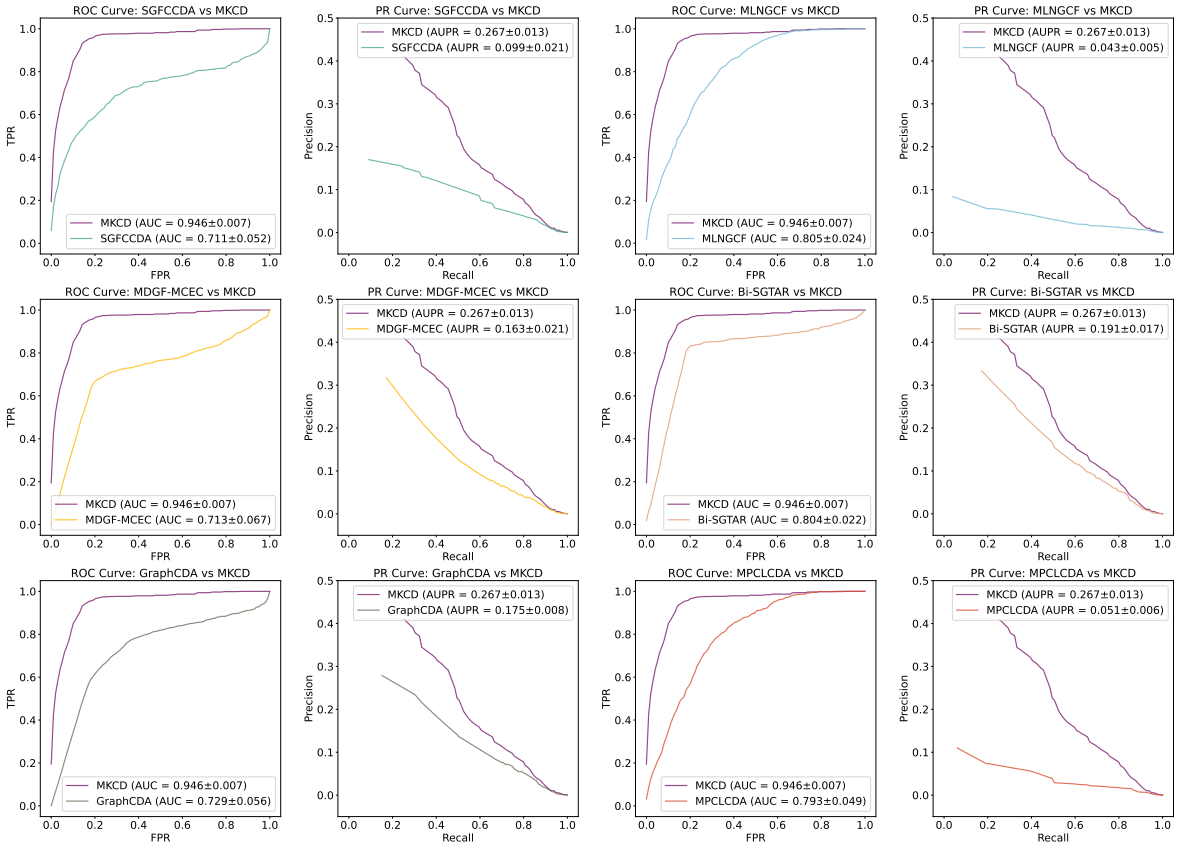

Fig. 5. ROC and PR curves of MKCD and other comparative methods.

TABLE V
COMPARISON OF F1 SCORE AND PRECISION OF MKCD WITH OTHER COMPARATIVE METHODS.

	MKCD	SGFCCDA	MLNGCF	MDGF-MCEC	Bi-SGTAR	GraphCDA	MPCLCDA
F1	**0.221**	0.122	0.041	0.158	0.184	0.171	0.089
Precision	**0.162**	0.092	0.042	0.116	0.141	0.148	0.069

TABLE VI
RESULTS OF THE PAIRED WILCOXON TEST COMPARING MKCD WITH ALL OTHER METHODS.

p-value	SGFCCDA	MLNGCF	MDGF-MCEC	Bi-SGTAR	GraphCDA	MPCLCDA
AUC	2.71e-111	7.20e-58	2.13e-109	3.27e-60	6.06e-108	1.22e-76
AUPR	6.02e-109	6.60e-115	1.18e-41	1.96e-09	2.86e-26	1.29e-113
F1 score	2.20e-56	4.38e-98	4.72e-30	2.57e-08	1.52e-04	3.43e-79
Precision	3.66e-83	4.22e-114	1.80e-47	2.47e-20	3.24e-33	2.52e-103

SGFCCDA: This model constructs a circRNA-disease heterogeneous graph and predicts potential circRNA-disease associations through scale graph convolutional networks and convolutional neural networks.

MLNGCF: In this model, various similarities between circRNAs and diseases are utilized to estimate the association scores based on a multilayer attention neural network.

MDGF-MCEC: This method establishes relationship graphs for circRNAs and diseases based on their respective similarities and learns node features through a multi-view dual attention graph convolution network.

Bi-SGTAR: The adjacency matrix of the circRNA-disease heterogeneous graph is decomposed into two views, and an encoder with sparse gating is employed to identify all circRNA-disease associations.

GraphCDA: This approach constructs separate similarity networks for circRNAs and diseases, utilizing a hybrid graph embedding model that combines graph convolutional networks and graph attention networks to simultaneously learn feature representations for circRNAs and diseases.

MPCLCDA: It automatically selects meta-paths for constructing meta-path graphs and employs graph convolutional networks and contrastive learning to learn node features for circRNAs and diseases.

Among them, SGFCCDA, MDGF-MCEC, and GraphCDA are models based on graph convolutional networks, with GraphCDA additionally incorporating an attention mechanism. MPCLCDA is based on contrastive learning methods, while Bi-SGTAR primarily utilizes sparse gating and graph decomposition techniques.

Figure 5 illustrates the ROC and PR curves for MKCD and the other methods. From the figure, it is evident that

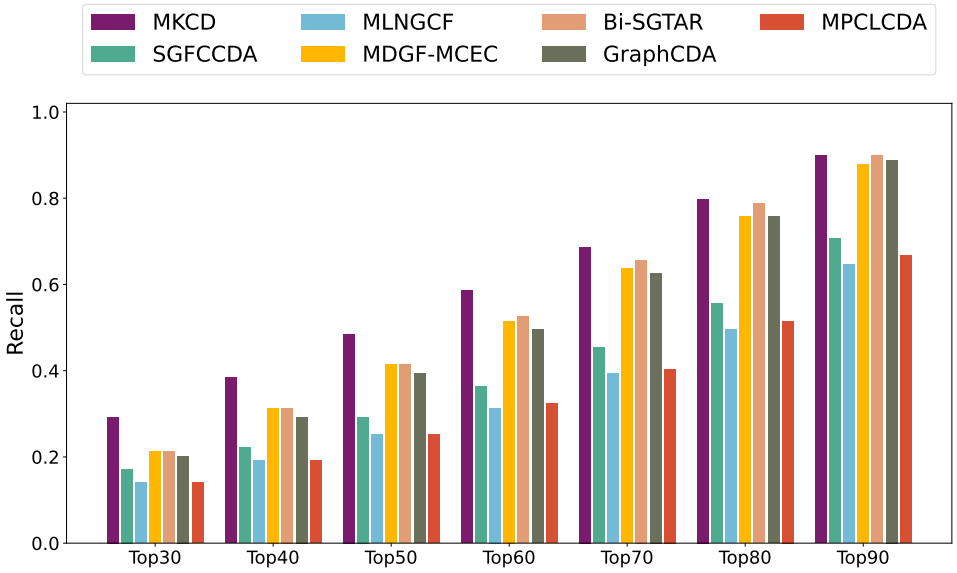

Fig. 6. Recall rates of diseases at multiple top k cutoffs.

MKCD achieved the highest average AUC of 0.946, surpassing SGFCCDA by 23.5%, MLNGCF by 14.1%, MDGF-MCEC by 23.3%, Bi-SGTAR by 14.2%, GraphCDA by 21.7%, and MPCLCDA by 15.3%. The average AUPR of MKCD was 0.267, which is higher than SGFCCDA, MLNGCF, MDGF-MCEC, Bi-SGTAR, GraphCDA, and MPCLCDA by 16.8%, 22.4%, 10.4%, 7.6%, 9.2%, and 21.6%, respectively. As shown in Table V, MKCD also achieves the highest F1 score (F1 = 0.221), which is 9.9% higher than SGFCCDA, 18.0% higher than MLNGCF, 6.3% higher than MDGF-MCEC, 3.7% higher than Bi-SGTAR, 5.0% higher than GraphCDA, and 13.2% higher than MPCLCDA. MKCD achieves the highest Precision of 0.162, outperforming the compared methods by 7.0%, 12.0%, 4.6%, 2.1%, 1.4% and 9.3%, respectively. Both MLNGCF and MPCLCDA employ graph neural networks, focusing solely on integrating the topology and node features of the circRNA-disease heterogeneous graph. In contrast to these two methods, SGFCCDA utilizes scale graph convolutional networks to address the issue of feature mixing between different channels caused by the linear layer structure of graph convolutional networks. MDGF-MCEC and GraphCDA primarily focus on learning from multiple similarity views, while Bi-SGTAR emphasizes learning from multiple heterogeneous graph views, which contributes to their superior performance. Our method outperforms these six methods mainly due to the embedding of multi-scale neighbor topology and the encoding of relationships among multiple features of circRNA, disease, and miRNA nodes.

To evaluate whether the proposed method significantly outperforms the six comparison methods in terms of AUC, AUPR, F1 score, and Precision metrics, we use the paired Wilcoxon test. In 5-fold cross-validation, all seven methods (including the proposed method and six comparison methods) calculate the AUC, AUPR, F1, and Precision scores for 138 diseases. The average score for each method is obtained by averaging the results across five folds, yielding the average score for each method on each disease. Then, for each disease, the score differences between the proposed method and each comparison method on these evaluation metrics are calculated. The absolute values of these differences are ranked, and the signed ranks are used to calculate the test statistic. Finally, the p-value is derived using the test statistic and its corresponding distribution to determine whether the proposed method significantly outperforms the comparison methods. According to the results in Table VI, all p-values are below 0.05, indicating that MKCD significantly outperforms SGFCCDA, MLNGCF, MDGF-MCEC, Bi-SGTAR, GraphCDA, and MPCLCDA with respect to AUC, AUPR, F1 score, and Precision.

For each disease, we calculated the recall rates of circRNA candidates at various top k values (Figure 6). When $k = 30$, MKCD achieved a recall rate of 29.3%, surpassing SGFCCDA by 12.1%, MLNGCF by 15.2%, MDGF-MCEC by 8.1%, Bi-SGTAR by 8.0%, GraphCDA by 9.1%, and MPCLCDA by 15.1%. When $k = 50$, 70, and 90, MKCD maintained its leading position with recall rates of 48.5%, 68.6%, and 90.1%, respectively. Bi-SGTAR (41.3%, 65.7%, 89.8%), MDGF-MCEC (41.3%, 63.6%, 87.9%), and GraphCDA (39.4%, 62.5%, 88.8%) also achieved decent performance. In contrast, SGFCCDA (29.3%, 45.5%, 70.6%), MLNGCF (25.2%, 39.4%, 64.6%), and MPCLCDA (25.3%, 40.5%, 66.7%) showed inferior performance.

To effectively evaluate the model's generalization capability, all known circRNA-disease associations and unknown circRNA-disease association pairs are randomly divided into 60% training set, 20% validation set, and 20% test set. This division ensures that the test set is independent of the training and loss evaluation process. The results (TABLE VII) indicate that MKCD performs best in terms of AUC, AUPR, F1 score, and Precision, with values of 0.956, 0.271, 0.269, and 0.193, respectively. For AUC, MKCD improves by 10.4% to 24.3% compared to other methods, while MKCD's AUPR exceeds other methods by 8.9% to 21.6%. MKCD's F1 score outperforms SGFCCDA, MLNGCF, MDGF-MCEC, Bi-SGTAR, GraphCDA, and MPCLCDA by 13.9%, 19.3%,

TABLE VII
COMPARISON OF AUC, AUPR, F1 SCORE AND PRECISION OF MKCD WITH OTHER COMPARATIVE METHODS.

	MKCD	SGFCCDA	MLNGCF	MDGF-MCEC	Bi-SGTAR	GraphCDA	MPCLCDA
AUC	**0.956**	0.852	0.789	0.728	0.813	0.713	0.801
AUPR	**0.271**	0.123	0.055	0.165	0.182	0.175	0.067
F1	**0.269**	0.130	0.076	0.143	0.187	0.179	0.086
Precision	**0.193**	0.129	0.082	0.120	0.155	0.119	0.099

12.6%, 8.2%, 9.0%, and 18.3%, respectively. MKCD's Precision improves by 6.4%, 11.1%, 7.3%, 3.8%, 7.4%, and 9.4% over other methods. These results demonstrate the excellent generalization ability of MKCD on the independent test set.

To comprehensively evaluate the time efficiency of MKCD and six comparison methods, we analyze their time complexities, where N is the total number of nodes, d is the feature dimension, and the number of circRNA and disease nodes are n and m, respectively. As shown in Supplementary Table 11, the time complexity of MKCD is $O(d \cdot N^2)$, while the time complexities of the comparison methods SGFCCDA, MLNGCF, MDGF-MCEC, GraphCDA, and MPCLCDA are all $O(d \cdot N^2)$, and Bi-SGTAR is $O(m \cdot n^2 + n \cdot m^2)$. We report the training and inference times of the models, defined as the time to achieve optimal performance and to predict a node pair association, respectively. The training and inference times for MKCD are 30.72 s and 0.352 ms, respectively. Bi-SGTAR, which is primarily based on a sparse gated network, exhibits the highest inference efficiency. GraphCDA, based on a graph convolutional network with shallow network depth, achieves the shortest training time. Overall, the training times for MKCD and the six comparison methods range from 6.67 s to 47.90 s, and the inference times range from 0.068ms to 0.568 ms, all showing high operational efficiency.

E. Case studies on three diseases

We conducted case studies on glioma, systemic lupus erythematosus, and glioblastoma to further validate MKCD's ability to mine potential circRNA candidates associated with these diseases. Glioma and glioblastoma are two types of primary brain tumors [61], [62], while systemic lupus erythematosus is a systemic autoimmune disease, primarily affecting women [63]. For each disease, the circRNA candidates were ranked in descending order based on their association scores, and the top 15 circRNAs were selected as candidates. Tables VIII, IX, and X show the top 15 circRNA candidates for glioma, systemic lupus erythematosus, and glioblastoma, respectively. The circRNADisease v2.0 database provides validated associations between circRNAs and various diseases, covering 4246 circRNAs and 330 diseases [64]. This database, along with relevant bioinformatics literature, was used to validate the predictions of circRNA-disease associations.

Using glioma as an example (Table VIII), all 15 circRNAs are validated in the literature, with 14 of them have been identified in circRNADisease v2.0. For instance, the top-ranked circ_002136 was found by He *et al.* [65] to inhibit the viability, migration, and tube formation of U87 glioma-exposed endothelial cells (GECs). Li *et al.* [66] found that the expression levels of hsa_circ_0061868 were upregulated in glioma cells.

All 15 circRNA candidates associated with systemic lupus erythematosus (SLE) (Table IX) are confirmed in the literature and have been included in circRNADisease v2.0. For example, Li *et al.* [13] identified hsa_circ_0046599 as a potential biomarker for systemic lupus erythematosus. Hsa_circ_0000479 was confirmed by Guo *et al.* [67] to be upregulated in peripheral blood mononuclear cells of patients with SLE.

For glioblastoma, 11 of the top 15 circRNA candidates (Table X) are validated in recent literature and documented in circRNADisease v2.0. For example, the study by Wang *et al.* [68] demonstrated that circNT5E exhibits tumor suppressor-like features in glioblastoma. Additionally, circMMP9 has been shown by Wang *et al.* [69] to promote the proliferation, migration, and invasion of glioblastoma multiforme cells. The unverified hsa_circ_101996 shares the same gene symbol as circSPECC1, which is shown to inhibit the proliferation, migration, invasion, and colony formation abilities of glioblastoma cells by encoding a new protein named SPECC1-415aa [70]. This suggests that hsa_circ_101996 is likely associated with glioblastoma. The unverified circPVT1, hsa_circ_100242, and hsa_circ_0003855 were ranked in the top 15 by Bi-SGTAR, and in the top 20 by MDGF-MCEC and GraphCDA, while SGFCCDA, MLNGCF, and MPCLCDA rank them among the top 30 disease-related circRNA candidates. This rankings indicates that both our method and the six comparative methods consider these unverified circRNA candidates as highly likely candidates for glioblastoma association.

F. Prediction of novel circRNA-disease associations

We utilized all known associations between circRNAs and diseases, and randomly selected an equal number of negative examples to train MKCD to predict circRNA candidates for 138 diseases. The top 15 circRNA candidates predicted by MKCD for each disease are listed in Supplementary File SF1.

IV. CONCLUSIONS

This paper presents a novel approach for encoding the relationships among circRNA, miRNA, and disease node features, while effectively learning and integrating both global and local features of node pairs to predict disease-related circRNAs. By adjusting the walking range of random walkers in the circRNA-disease-miRNA heterogeneous graph, multi-scale neighbor topologies are constructed, and the importance of each scale-specific neighbor topology is adaptively determined. The proposed multi-scale neighbor topology-guided transformer dynamically updates the neighbor topology and captures the evolving relationships between the features of circRNA, miRNA, and disease nodes. The FGN assigns higher

TABLE VIII
TOP 15 PREDICTED RESULTS OF GLIOMA-RELATED CIRCRNAS BASED ON MKCD.

Rank	CircRNA name	Evidence	Rank	CircRNA name	Evidence
1	circ_002136	L^a, PMID:30736838	9	hsa_circ_0079593	L^a, PMID:31148222
2	hsa_circ_0061868	PMID:30341906	10	circSMO742	L^a, PMID:31895689
3	circPTN	L^a, PMID:31511040	11	hsa_circ_0012129	L^a, PMID:29686222
4	circ-TTBK2	L^a, PMID:32196629	12	hsa_circ_0088732	L^a, PMID:32154171
5	circ-EZH2	L^a, PMID:31669648	13	circNFIX	L^a, PMID:30072869
6	hsa_circ_0000594	L^a, PMID:28219405	14	circ-PTPRZ1	L^a, PMID:31364003
7	hsa_circ_0005198	L^a, PMID:31038801	15	hsa_circ_0014359	L^a, PMID:30745107
8	hsa_circ_0000177	L^a, PMID:30010402			

L^a: circRNADisease v2.0.

TABLE IX
TOP 15 PREDICTED RESULTS OF SYSTEMIC LUPUS ERYTHEMATOSUS-RELATED CIRCRNAS BASED ON MKCD.

Rank	CircRNA name	Evidence	Rank	CircRNA name	Evidence
1	hsa_circ_0046599	L^a, PMID:29360436	9	hsa_circ_0008615	L^a, PMID:29360436
2	hsa_circ_0001866	L^a, PMID:29360436	10	hsa_circ_0021549	L^a, PMID:29360436
3	hsa_circ_0034398	L^a, PMID:29360436	11	has_circ_0049220	L^a, PMID:29606700
4	hsa_circ_0003146	L^a, PMID:29360436	12	hsa_circ_0092374	L^a, PMID:29360436
5	hsa_circ_0000479	L^a, PMID:31608065	13	hsa_circ_0040705	L^a, PMID:29360436
6	hsa_circ_0057762	L^a, PMID:30628013	14	hsa_circ_0012919	L^a, PMID:30237316
7	circPTPN22	L^a, PMID:30871426	15	hsa_circ_0049224	L^a, PMID:29606700
8	hsa_circ_0045272	L^a, PMID:29700819			

L^a: circRNADisease v2.0.

TABLE X
TOP 15 PREDICTED RESULTS OF GLIOBLASTOMA-RELATED CIRCRNAS BASED ON MKCD.

Rank	CircRNA name	Evidence	Rank	CircRNA name	Evidence
1	hsa_circ_0001801	L^a, PMID:31858556	9	circMTO1	L^a, PMID:31456594
2	circNT5E	L^a, PMID:29967262	10	circ-AKT3	L^a, PMID:31470874
3	circ-PITX1	L^a, PMID:31493405	11	circPVT1	unconfirmed
4	hsa_circ_0043949	L^a, PMID:31823158	12	circPTN	L^a, PMID:31511040
5	hsa_circ_0074027	L^a, PMID:30738578	13	hsa_circ_101996	unconfirmed
6	circMMP9	L^a, PMID:30470262	14	hsa_circ_100242	unconfirmed
7	circ-Foxo3	L^a, PMID:31802888	15	hsa_circ_0003855	unconfirmed
8	hsa_circ_0001946	L^a, PMID:31599076			

L^a: circRNADisease v2.0.

weights to topological and original features based on their importance. The ACK encodes the features of circRNA-disease node pairs, facilitating the identification of local and global dependencies among pairwise attributes. The comparison results show that our model outperforms other state-of-the-art methods in terms of AUC and AUPR. The recall rate of top-ranked circRNA candidates and case studies over three diseases further prove that MKCD is capable of providing reliable diseaserelated circRNA candidates. With the increasing availability of circRNA-diseases association data, our future work will focus on developing advanced methods for accurate prediction of circRNA-disease associations in large-scale datasets.

REFERENCES

[1] W.-Y. Zhou, Z.-R. Cai, J. Liu, D.-S. Wang, H.-Q. Ju, and R.-H. Xu, "Circular rna: metabolism, functions and interactions with proteins," *Molecular cancer*, vol. 19, pp. 1–19, 2020.

[2] K. Abdelmohsen, A. C. Panda, R. Munk, I. Grammatikakis, D. B. Dudekula, S. De, J. Kim, J. H. Noh, K. M. Kim, J. L. Martindale *et al.*, "Identification of hur target circular rnas uncovers suppression of pabpn1 translation by circpabpn1," *RNA biology*, vol. 14, no. 3, pp. 361–369, 2017.

[3] S. Gao, Y. Yu, L. Liu, J. Meng, and G. Li, "Circular rna hsa_circ_0007059 restrains proliferation and epithelial-mesenchymal transition in lung cancer cells via inhibiting microrna-378," *Life sciences*, vol. 233, p. 116692, 2019.

[4] Z. Li, Y. Ruan, H. Zhang, Y. Shen, T. Li, and B. Xiao, "Tumor-suppressive circular rnas: mechanisms underlying their suppression of tumor occurrence and use as therapeutic targets," *Cancer science*, vol. 110, no. 12, pp. 3630–3638, 2019.

[5] G. Liang, Y. Ling, M. Mehrpour, P. E. Saw, Z. Liu, W. Tan, Z. Tian, W. Zhong, W. Lin, Q. Luo *et al.*, "Autophagy-associated circrna circcdyl augments autophagy and promotes breast cancer progression," *Molecular cancer*, vol. 19, pp. 1–16, 2020.

[6] X. Wang, C. Zhang, Z. Wu, Y. Chen, and W. Shi, "Circibtk inhibits dna demethylation and activation of akt signaling pathway via mir-29b in peripheral blood mononuclear cells in systemic lupus erythematosus," *Arthritis Research & Therapy*, vol. 20, pp. 1–10, 2018.

[7] M. A. Khan, Y. J. Reckman, S. Aufiero, M. M. van den Hoogenhof, I. van der Made, A. Beqqali, D. R. Koolbergen, T. B. Rasmussen, J. Van Der Velden, E. E. Creemers *et al.*, "Rbm20 regulates circular rna production from the titin gene," *Circulation research*, vol. 119, no. 9, pp. 996–1003, 2016.

[8] D. Siede, K. Rapti, A. Gorska, H. Katus, J. Altmüller, J. Boeckel, B. Meder, C. Maack, M. Völkers, O. Müller *et al.*, "Identification of circular rnas with host gene-independent expression in human model systems for cardiac differentiation and disease," *Journal of molecular and cellular cardiology*, vol. 109, pp. 48–56, 2017.

[9] C. Jin, W. Zhao, Z. Zhang, and W. Liu, "Silencing circular rna circznf609 restrains growth, migration and invasion by up-regulating microrna-186-

5p in prostate cancer," *Artificial cells, nanomedicine, and biotechnology*, vol. 47, no. 1, pp. 3350–3358, 2019.
[10] F. J. Slack and A. M. Chinnaiyan, "The role of non-coding rnas in oncology," *Cell*, vol. 179, no. 5, pp. 1033–1055, 2019.
[11] Q. Shang, Z. Yang, R. Jia, and S. Ge, "The novel roles of circrnas in human cancer," *Molecular cancer*, vol. 18, pp. 1–10, 2019.
[12] X. Chen, J. Yin, J. Qu, and L. Huang, "Mdhgi: matrix decomposition and heterogeneous graph inference for mirna-disease association prediction," *PLoS computational biology*, vol. 14, no. 8, p. e1006418, 2018.
[13] H. Li, K. Li, W. Lai, X. Li, H. Wang, J. Yang, S. Chu, H. Wang, C. Kang, and Y. Qiu, "Comprehensive circular rna profiles in plasma reveals that circular rnas can be used as novel biomarkers for systemic lupus erythematosus," *Clinica Chimica Acta*, vol. 480, pp. 17–25, 2018.
[14] W. Huang, Y. Wu, M. Qiao, Z. Xie, X. Cen, X. Huang, and Z. Zhao, "Circrna–mirna networks in regulating bone disease," *Journal of Cellular Physiology*, vol. 237, no. 2, pp. 1225–1244, 2022.
[15] L. Shrestha and A. Leier, "Identification of a circrna-mirna-mrna interactome associated with parkinson's disease progression," *Journal of Parkinson's Disease*, p. 1877718X251331930, 2024.
[16] J. Yao, G. Xu, L. Zhu, and H. Zheng, "circgfra1 enhances nsclc progression by sponging mir-188-3p," *OncoTargets and therapy*, pp. 549–558, 2020.
[17] W. Peng, Z. He, W. Dai, and W. Lan, "Mhclmda: multihypergraph contrastive learning for mirna–disease association prediction," *Briefings in bioinformatics*, vol. 25, no. 1, p. bbad524, 2024.
[18] W. Peng, Z. Che, W. Dai, S. Wei, and W. Lan, "Predicting mirna-disease associations from mirna-gene-disease heterogeneous network with multi-relational graph convolutional network model," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 6, pp. 3363–3375, 2022.
[19] L. Huang, N. Sheng, L. Gao, L. Wang, W. Hou, J. Hong, and Y. Wang, "Self-supervised contrastive learning on attribute and topology graphs for predicting relationships among lncrnas, mirnas and diseases," *IEEE Journal of Biomedical and Health Informatics*, 2024.
[20] N. Sheng, Y. Wang, L. Huang, L. Gao, Y. Cao, X. Xie, and Y. Fu, "Multi-task prediction-based graph contrastive learning for inferring the relationship among lncrnas, mirnas and diseases," *Briefings in bioinformatics*, vol. 24, no. 5, p. bbad276, 2023.
[21] N. Sheng, L. Huang, L. Gao, Y. Cao, X. Xie, and Y. Wang, "A survey of computational methods and databases for lncrna-mirna interaction prediction," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 5, pp. 2810–2826, 2023.
[22] M.-M. Wei, L. Wang, B.-W. Zhao, X.-R. Su, Z.-H. You, and D.-S. Huang, "Integrating transformer and graph attention network for circrna-mirna interaction prediction," *IEEE Journal of Biomedical and Health Informatics*, 2025.
[23] M. Wei, L. Wang, Y. Li, Z. Li, B. Zhao, X. Su, Y. Wei, and Z. You, "Biokg-cmi: a multi-source feature fusion model based on biological knowledge graph for predicting circrna-mirna interactions," *Science China Information Sciences*, vol. 67, no. 8, p. 189104, 2024.
[24] W. Lan, Y. Dong, H. Zhang, C. Li, Q. Chen, J. Liu, J. Wang, and Y.-P. P. Chen, "Benchmarking of computational methods for predicting circrna-disease associations," *Briefings in Bioinformatics*, vol. 24, no. 1, p. bbac613, 2023.
[25] J. Yang and X. Lei, "Predicting circrna-disease associations based on autoencoder and graph embedding," *Information Sciences*, vol. 571, pp. 323–336, 2021.
[26] C. Fan, X. Lei, and F.-X. Wu, "Prediction of circrna-disease associations using katz model based on heterogeneous networks," *International journal of biological sciences*, vol. 14, no. 14, p. 1950, 2018.
[27] X. Lei, Z. Fang, L. Chen, and F.-X. Wu, "Pwcda: path weighted method for predicting circrna-disease associations," *International journal of molecular sciences*, vol. 19, no. 11, p. 3410, 2018.
[28] H. Wei and B. Liu, "icircda-mf: identification of circrna-disease associations based on matrix factorization," *Briefings in bioinformatics*, vol. 21, no. 4, pp. 1356–1367, 2020.
[29] M.-N. Wang, X.-J. Xie, Z.-H. You, L. Wong, L.-P. Li, and Z.-H. Chen, "Combining k nearest neighbor with nonnegative matrix factorization for predicting circrna-disease associations," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 5, pp. 2610–2618, 2022.
[30] X. Lei and C. Bian, "Integrating random walk with restart and k-nearest neighbor to identify novel circrna-disease association," *Scientific reports*, vol. 10, no. 1, p. 1943, 2020.
[31] C. Yan, J. Wang, and F.-X. Wu, "Dwnn-rls: regularized least squares method for predicting circrna-disease associations," *BMC bioinformatics*, vol. 19, pp. 73–81, 2018.
[32] L. Wang, L. Wong, Z. Li, Y. Huang, X. Su, B. Zhao, and Z. You, "A machine learning framework based on multi-source feature fusion for circrna-disease association prediction," *Briefings in Bioinformatics*, vol. 23, no. 5, p. bbac388, 2022.
[33] X. Lei and Z. Fang, "Gbdtcda: predicting circrna-disease associations based on gradient boosting decision tree with multiple biological data fusion," *International journal of biological sciences*, vol. 15, no. 13, p. 2911, 2019.
[34] K. Deepthi and A. Jereesh, "Inferring potential circrna–disease associations via deep autoencoder-based classification," *Molecular Diagnosis & Therapy*, vol. 25, pp. 87–97, 2021.
[35] W. Zhang, C. Yu, X. Wang, and F. Liu, "Predicting circrna-disease associations through linear neighborhood label propagation method," *Ieee Access*, vol. 7, pp. 83 474–83 483, 2019.
[36] L. Peng, C. Yang, L. Huang, X. Chen, X. Fu, and W. Liu, "Rnmflp: predicting circrna–disease associations based on robust nonnegative matrix factorization and label propagation," *Briefings in bioinformatics*, vol. 23, no. 5, p. bbac155, 2022.
[37] Y. Tian, Q. Zou, C. Wang, and C. Jia, "Mamlcda: A meta-learning model for predicting circrna-disease association based on maml combined with cnn," *IEEE Journal of Biomedical and Health Informatics*, 2024.
[38] L. Wang, Z.-H. You, Y.-A. Huang, D.-S. Huang, and K. C. Chan, "An efficient approach based on multi-sources information to predict circrna–disease associations using deep convolutional neural network," *Bioinformatics*, vol. 36, no. 13, pp. 4038–4046, 2020.
[39] C. Lu, M. Zeng, F.-X. Wu, M. Li, and J. Wang, "Improving circrna–disease association prediction by sequence and ontology representations with convolutional and recurrent neural networks," *Bioinformatics*, vol. 36, no. 24, pp. 5656–5664, 2020.
[40] S.-Z. Liang, L. Wang, Z.-H. You, C.-Q. Yu, M.-M. Wei, Y. Wei, T.-L. Shi, and C. Jiang, "Predicting circrna–disease associations through multisource domain-aware embeddings and feature projection networks," *Journal of Chemical Information and Modeling*, 2025.
[41] X. Zhang, Q. Zou, M. Niu, and C. Wang, "Predicting circrna-disease associations with shared units and multi-channel attention mechanisms," *Bioinformatics*, p. btaf088, 2025.
[42] S. Li, Q. Chen, Z. Liu, S. Pan, and S. Zhang, "Bi-sgtar: A simple yet efficient model for circrna-disease association prediction based on known association pair only," *Knowledge-Based Systems*, vol. 291, p. 111622, 2024.
[43] J. Shang, L. Zhao, X. He, X. Meng, L. Zhang, D. Ge, F. Li, and J.-X. Liu, "Sgfccda: Scale graph convolutional networks and feature convolution for circrna-disease association prediction," *IEEE Journal of Biomedical and Health Informatics*, 2024.
[44] W. Liu, T. Tang, X. Lu, X. Fu, Y. Yang, and L. Peng, "Mpclcda: predicting circrna–disease associations by using automatically selected meta-path and contrastive learning," *Briefings in Bioinformatics*, vol. 24, no. 4, p. bbad227, 2023.
[45] Q. Wu, Z. Deng, X. Pan, H.-B. Shen, K.-S. Choi, S. Wang, J. Wu, and D.-J. Yu, "Mdgf-mcec: a multi-view dual attention embedding model with cooperative ensemble learning for circrna-disease association prediction," *Briefings in Bioinformatics*, vol. 23, no. 5, p. bbac289, 2022.
[46] X. He, J. Shang, D. Ge, F. Li, and J.-X. Liu, "Rpmvcda: Random perturbation and multi-view graph convolutional networks for circrna-disease association prediction," *IEEE Transactions on Computational Biology and Bioinformatics*, 2025.
[47] Q. Wu, Z. Deng, W. Zhang, X. Pan, K.-S. Choi, Y. Zuo, H.-B. Shen, and D.-J. Yu, "Mlngcf: circrna–disease associations prediction with multilayer attention neural graph-based collaborative filtering," *Bioinformatics*, vol. 39, no. 8, p. btad499, 2023.
[48] L. Wang, L. Wong, Z.-H. You, and D.-S. Huang, "Amdecda: Attention mechanism combined with data ensemble strategy for predicting circrna-disease association," *IEEE Transactions on Big Data*, vol. 10, no. 4, pp. 320–329, 2023.
[49] Q. Dai, Z. Liu, Z. Wang, X. Duan, and M. Guo, "Graphcda: a hybrid graph representation learning framework based on gcn and gat for predicting disease-associated circrnas," *Briefings in Bioinformatics*, vol. 23, no. 5, p. bbac379, 2022.
[50] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, "Kan: Kolmogorov-arnold networks," *arXiv preprint arXiv:2404.19756*, 2024.
[51] W. Lan, Y. Dong, Q. Chen, R. Zheng, J. Liu, Y. Pan, and Y.-P. P. Chen, "Kgancda: predicting circrna-disease associations based on knowledge graph attention network," *Briefings in Bioinformatics*, vol. 23, no. 1, p. bbab494, 2022.

[52] W. Lan, M. Zhu, Q. Chen, B. Chen, J. Liu, M. Li, and Y.-P. P. Chen, "Circr2cancer: a manually curated database of associations between circrnas and cancers," *Database*, vol. 2020, p. baaa085, 2020.

[53] M. Rophina, D. Sharma, M. Poojary, and V. Scaria, "Circad: a comprehensive manually curated resource of circular rna associated with diseases," *Database*, vol. 2020, p. baaa019, 2020.

[54] Z. Zhao, K. Wang, F. Wu, W. Wang, K. Zhang, H. Hu, Y. Liu, and T. Jiang, "circrna disease: a manually curated database of experimentally supported circrna-disease associations," *Cell death & disease*, vol. 9, no. 5, p. 475, 2018.

[55] D. Wang, J. Wang, M. Lu, F. Song, and Q. Cui, "Inferring the human microrna functional similarity and functional network based on microrna-associated diseases," *Bioinformatics*, vol. 26, no. 13, pp. 1644–1650, 2010.

[56] X. Chen, C. Clarence Yan, C. Luo, W. Ji, Y. Zhang, and Q. Dai, "Constructing lncrna functional similarity network based on lncrna-disease associations and disease semantic similarity," *Scientific reports*, vol. 5, no. 1, p. 11338, 2015.

[57] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[58] K. Hajian-Tilaki, "Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation," *Caspian journal of internal medicine*, vol. 4, no. 2, p. 627, 2013.

[59] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," *PloS one*, vol. 10, no. 3, p. e0118432, 2015.

[60] L. Takac and M. Zabovsky, "Data analysis in public social networks," in *International scientific conference and international workshop present day trends of innovations*, vol. 1, no. 6, 2012.

[61] M. Weller, W. Wick, K. Aldape, M. Brada, M. Berger, S. M. Pfister, R. Nishikawa, M. Rosenthal, P. Y. Wen, R. Stupp *et al.*, "Glioma," *Nature reviews Disease primers*, vol. 1, no. 1, pp. 1–18, 2015.

[62] H.-G. Wirsching and M. Weller, "Glioblastoma," *Malignant Brain Tumors: State-of-the-Art Treatment*, pp. 265–288, 2017.

[63] A. Kaul, C. Gordon, M. Crow, Z. Touma, M. Urowitz, R. van Vollenhoven, G. Ruiz-Irastorza, and G. Hughes, "Systemic lupus erythematosus. nat reviews disease primers 2: 16039," 2016.

[64] C. Fan, X. Lei, J. Tie, Y. Zhang, F.-X. Wu, and Y. Pan, "Circr2disease v2. 0: an updated web server for experimentally validated circrna–disease associations and its application," *Genomics, Proteomics and Bioinformatics*, vol. 20, no. 3, pp. 435–445, 2022.

[65] Z. He, X. Ruan, X. Liu, J. Zheng, Y. Liu, L. Liu, J. Ma, L. Shao, D. Wang, S. Shen *et al.*, "Fus/circ_002136/mir-138-5p/sox13 feedback loop regulates angiogenesis in glioma," *Journal of Experimental & Clinical Cancer Research*, vol. 38, pp. 1–19, 2019.

[66] G. Li, M. Huang, Y. Cai, Y. Yang, X. Sun, and Y. Ke, "Circ-u2af1 promotes human glioma via derepressing neuro-oncological ventral antigen 2 by sponging hsa-mir-7-5p," *Journal of Cellular Physiology*, vol. 234, no. 6, pp. 9144–9155, 2019.

[67] G. Guo, H. Wang, L. Ye, X. Shi, K. Yan, K. Lin, Q. Huang, B. Li, Q. Lin, L. Zhu *et al.*, "Hsa_circ_0000479 as a novel diagnostic biomarker of systemic lupus erythematosus. front immunol 10: 2281," 2019.

[68] R. Wang, S. Zhang, X. Chen, N. Li, J. Li, R. Jia, Y. Pan, and H. Liang, "Circnt5e acts as a sponge of mir-422a to promote glioblastoma tumorigenesis," *Cancer research*, vol. 78, no. 17, pp. 4812–4825, 2018.

[69] R. Wang, S. Zhang, X. Chen, N. Li, J. Li, R. Jia, Y. Pan, and H. Liang, "Eif4a3-induced circular rna mmp9 (circmmp9) acts as a sponge of mir-124 and promotes glioblastoma multiforme cell tumorigenesis," *Molecular cancer*, vol. 17, pp. 1–12, 2018.

[70] C. Wei, D. Peng, B. Jing, B. Wang, Z. Li, R. Yu, S. Zhang, J. Cai, Z. Zhang, J. Zhang *et al.*, "A novel protein specc1-415aa encoded by n6-methyladenosine modified circspecc1 regulates the sensitivity of glioblastoma to tmz," *Cellular & molecular biology letters*, vol. 29, no. 1, p. 127, 2024.